Robust Fault Estimation with Structured Uncertainty: Scalable Algorithms and Experimental Validation in Automated Vehicles

Chris van der Ploeg, Pedro Vieira Oliveira, Emilia Silvas, Peyman Mohajerin Esfahani and Nathan van de Wouw

***Abstract*—To increase system robustness and autonomy, in this article, we propose a non-linear fault estimation filter for a class of linear dynamical systems, subject to structured uncertainty, measurement noise and system delays, in the presence of additive and multiplicative faults. The proposed filter architecture combines tools from model-based control approaches, regression techniques, and convex optimization. The proposed method estimates the additive and multiplicative faults using a linear residual generator combined with non-linear regression. An offline simulator allows us to numerically characterize the mismatch between an assumed linear model and a range of alternative linear models that exhibit different levels of structured uncertainty. Moreover, we show how the performance bounds of the estimator, valid in the absence of uncertainty, can be used to determine appropriate countermeasures for measurement noise. In the scope of this work, we focus particularly on a fault estimation problem for SAE level 4 automated vehicles, which must remain operational in various cases and can not rely on the driver. The proposed approach is demonstrated in simulations and in an experimental setting, where it is shown that additive and multiplicative faults can be estimated in a real vehicle under the influence of model uncertainty, measurement noise, and delay.**

***Index Terms*—Fault estimation, model uncertainty, convex optimization, automated vehicles.**

I. Introduction

AUTOMATED vehicles are currently the subject of ongoing research aimed at achieving higher levels of automation and eventual autonomy. As we continue to strive for these higher levels of automation, it has become clear that this technology has the potential to have a positive impact on our society by increasing road capacity, reducing traffic congestion, improving safety, and reducing emissions [1]. Although these vehicles can positively impact society, proving their safety in operation remains tedious [2]. As vehicles reach these higher levels of automation, they should become self-aware of their state of health and limitations, tasks typically fulfilled by human drivers in non-automated vehicles. Specifically, for certain subsystems and functionalities, for example, the power steering system of an automated vehicle, this self-awareness is crucial, since it is considered to be safety-critical [3], i.e., a system whose malfunction may result in death or serious injury to people. Different types and magnitudes of faults may require different actions to mitigate them. This could involve using robust controllers for closed-loop mitigation or bringing the vehicle to a safe state if the combination of faults exceeds a certain threshold [4]. Therefore, it is crucial to have knowledge of current faults and their severity.

The problem of fault diagnosis can be divided into three parts. Firstly, the presence of a fault needs to be detected, which we define as the process of identifying anomalies or deviations from normal operation, which involves generating a residual signal. This is done through, e.g., parity-space-based approaches, banks of state observers, and parameter estimation [5], [6]. Second, the fault must be isolated, that is, the process of separating or distinguishing between multiple fault sources within a system. The detection and isolation of faults should be done while remaining insensitive to potential exogenous disturbances [7], uncertainties in the system [8], or other real-life phenomena, such as measurement or actuation delay [9] or noise [10]. This will help prevent false positives and/or misclassification of the fault. Third, the fault can be estimated to determine its severity and to be used in mitigating measures (e.g., closed-loop mitigation). This can be done through, e.g., proportional and integral observers [11], adaptive observers [12] or unknown input observers [13].

In the context of automated vehicle steering systems, the state-of-the-art specifically addresses the detection and estimation of additive and multiplicative faults [14]. The detection of additive faults in a steering system is covered in [15], followed by [16], [17] for its estimation. In [18] single multiplicative faults and sensor faults are detected and in [19], respectively, estimated. Simultaneous detection, isolation, and estimation of additive and multiplicative faults are covered in [4], although the approach does not cover its robustness in the presence of unmodeled perturbations (e.g., model uncertainty). Uncertainty in the automated driving application under study can be induced by model uncertainty, input/measurement delays, and measurement noise. In a general sense, uncertainty in linear systems is divided into two categories: structured uncertainty and uncertainty arising from non-linearities that cannot be accounted for in the system model. Structured uncertainty is a term frequently discussed in the field of robust control [20], which involves assuming a bounded uncertainty in certain parts of the linear model. A popular solution for fault esti-

C.J. van der Ploeg, E. Silvas and N. van de Wouw are with the Department of Mechanical Engineering, Eindhoven University of Technology, ({C.J.v.d.Ploeg, E.Silvas, N.v.d.Wouw}@tue.nl).
C.J. van der Ploeg, E. Silvas and P. Vieira Oliveira are with the Netherlands Organization for Applied Scientific Research, Integrated Vehicle Safety Group, 5700 AT Helmond, The Netherlands ({Chris.vanderPloeg, Emilia.Silvas, Pedro.VieiraOliveira}@tno.nl).
P. Mohajerin Esfahani is with the Delft Center for Systems and Control, Delft University of Technology (P.MohajerinEsfahani@tudelft.nl). Peyman Mohajerin Esfahani acknowledges the support of the ERC grant TRUST-949796.

mation, in the presence of structured uncertainty, is to use sliding mode observers [21], [22] (SMO). However, SMOs are potentially sensitive to measurement noise and chattering. Observer-based methods, e.g., Kalman filters, have the notion of modeling uncertainty and sensor noise embedded as process and measurement noise [23], which is, however, assumed to be Gaussian. Therefore, in the case of structured uncertainty, a Kalman-type filter may fail [24].

An alternative option is to use a data-driven approach to *learn* model uncertainty when non-linearities appear in the system difficult to capture via linear models, e.g., through neural networks [25], or a variety of other full-learning-based approaches [26]. Combining model knowledge and data could lead to a stronger combination than using either of the two in isolation. In [27], a linear detection and estimation filter is presented, which is trained (based on data) to reject the output mismatch between a non-linear high-fidelity simulator and an abstract linear model. Other mismatches in signals and dynamics in the high-fidelity non-linear simulator (e.g., states or disturbances) are difficult to interpret in a linear sense. This limits the performance of the algorithm when applied to systems where these mismatch signals could be characterized. The works [28], [29] examine non-linear systems by adding an additive non-linear term to an assumed linear system. [28] uses convex optimization to learn the uncertain behavior, originating from non-linearities and noise, using mismatch signatures of such effects. [29] employs an adaptive method to identify the dynamics of noise and uncertainties online. However, this method requires a certain degree of excitation of the signals inside the regressor term to identify the uncertainties.

Current fault detection, isolation, and estimation methodologies are evolving to address the challenges of structured uncertainties, measurement noise, and delays in various systems [5]. However, a significant gap persists in the literature concerning the robust simultaneous estimation of additive and multiplicative faults, where the faults act through the same input or output channel. Specifically in an experimental setting where uncertainty, delays, and noise may affect the estimation process. Although robust methods have been developed to estimate simultaneous faults under uncertain conditions [30]–[32], these works do not consider the simultaneous appearance of faults on the same channel. The challenge of estimating faults acting on the same channel has been covered in, e.g., [4], although their robustness against real-world phenomena remains untested in experimental settings. Consequently, this gap presents a critical area for research, as establishing the reliability of these methods in practical scenarios is essential to advance automated driving technologies.

Main Contributions: In view of the literature mentioned above, this study's contributions are summarized as follows:

(i) **Scalable design for robust multivariate fault estimation.** We propose a scalable algorithm to design robust fault estimation filters capable of simultaneously estimating both additive and multiplicative faults with similar dynamic effects. The particular features studied in this work are the presence of model uncertainties, input/output delays, and measurement noise. The scalability of the proposed approach is with respect to the states and input size of the dynamical systems. This study is the robust version of the nominal designs in [4], [16], considering the following additional key challenges:

- *Structured model uncertainty:* To address structured model uncertainty, our proposed method exploits both data and a model-based approach. We leverage prior knowledge about the source of the uncertainty and its effects. We propose a robust counterpart of existing convex optimization tools for estimation filters where the uncertain parameters belong to a known set.
- *Measurement noise and input delay:* To reduce the impact of noise on fault estimates, we borrow the performance bounds of the estimation error introduced in our previous study [4] to serve as an objective reference for training the filter parameters. Additionally, we address the potential input/output delay by augmenting the filter states with the average identified delays.

(ii) **Experimental validation in automated driving.** Another main contribution of this study is the validation of our theoretical results on a real vehicle. We demonstrate that the proposed robustified approach outperforms the state-of-the-practice in estimating faults in real-time operation.

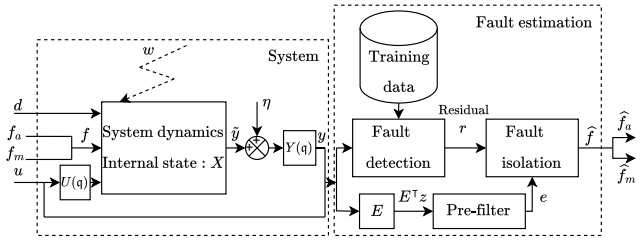


Fig. 1: Block diagram of the proposed robust diagnosis filter.

The remainder of this paper is organized as follows. Sec. II introduces the problem setting, the state-of-the-art on fault estimation of additive and multiplicative faults, its limitations given the real-life phenomena mentioned above, and an outline of the proposed approach. Sec. III introduces the robust approach toward fault estimation. Sec. IV then provides the results, both in a synthetic example and using real-life experimental data. Finally, Sec. V concludes the work.

Notation. *The symbol $\mathbb{R}$ represents the set of real numbers. The ones column vector with length n is denoted by $\mathbb{1}_n := [1, 1, \dots, 1]^\intercal$. The p-norm of a vector v is denoted by $\|v\|_p$ where $p \in [1, \infty]$. The $\mathcal{L}_2$-norm of a discrete-time signal $x(k)$ is defined as $\|x(k)\|_{\mathcal{L}_2} = (\sum_{n=-\infty}^{\infty} \|x(n)\|_2^2)^{\frac{1}{2}}$. Given a matrix $A \in \mathbb{R}^{n\times m}$, its transpose is denoted by $A^\intercal \in \mathbb{R}^{m\times n}$, and $A^\dagger := (A^\intercal A)^{-1} A^\intercal$ is the pseudoinverse. The operators $\mu_n[x]$ and $V_n[x]$ map $\mathbb{R}$-valued discrete-time signals to $\mathbb{R}$-valued discrete-time signals, and are defined as the first moment $\mu_n[x](k) := \frac{1}{n}\sum_{i=0}^{n-1} x(k-i)$ and the centered second moment $V_n^2[x](k) := \frac{1}{n}\sum_{i=0}^{n-1} x^2(k-i) - \mu_n^2[x](k)$ of the signal x over the last n time instants. Throughout this study, we reserve the bold sub-scripted by n, $\mathbf{x}_n$, as the concatenated version of the signal x over the last n time instants: $\mathbf{x}_n(k) := \big[x(k), x(k-1), \dots, x(k-n+1)\big]^\intercal$. The symbol $\mathfrak{q}$ represents the shift operator, i.e., $\mathfrak{q}[x(k)] = x(k+1)$.*

II. Problem description and outline of the proposed Approach

In this section, we present the class of systems considered throughout this work. Subsequently, we will formulate the high-level problem. We further elaborate on the challenges and shortcomings of the methods available in the current literature. Finally, an outline of the proposed solution that addresses the challenges is provided.

A. Model description and problem statement

Throughout this study, we examine nonlinear dynamical systems characterized by linear time-invariant (LTI) dynamics within a discrete-time framework. These systems are described using discrete-time differential algebraic equations (DAEs), similar to the model formulations presented in [4], [27], [28]. In contrast to the system description in previous work, we also incorporate the possible presence of model uncertainty in this description. Therefore, we employ the DAE description provided in [4, Eq. (1)], and reformulate the model as follows:

$$H(\mathfrak{q};w)[x]+L(\mathfrak{q};w)[z]+F(\mathfrak{q};w)\left[f_{\mathrm{a}}+E^{\intercal}zf_{\mathrm{m}}\right]=0. \quad (1)$$

Here, the variables x, z, f_{a}, and f_{m} represent discrete-time signals with values in $\mathbb{R}^{n_x}, \mathbb{R}^{n_z}$, and $\mathbb{R}^{n_f}$, respectively, and are indexed by the discrete time counter k. More specifically, the variable z is composed of all measurable signals, including control inputs u and measurements y. The variable x contains all unknown signals in the system, in this work defined as the true internal state X, unmeasurable disturbances d, and measurement noise η representing a set of uncorrelated Gaussian white noise sequences. The vector $E \in \mathbb{R}^{n_z}$ selects which signals in z will be affected by f_{m}. The matrices $H(\mathfrak{q};w)$, $L(\mathfrak{q};w)$, and $F(\mathfrak{q};w)$ are polynomial functions in the shift operator $\mathfrak{q}$, with n_r rows and n_x, n_z and n_f columns, respectively. These matrices depend on a set of parameters $w \in \mathcal{W} \subseteq \mathbb{R}^{n_w}$, where n_w represents the number of uncertain parameters. The symbol $\mathcal{W}$ represents a set that contains all the parametric uncertainties of the model. The exact value of the uncertainty is unknown a priori, but it is assumed that the parameters have a nominal value $w_0 \in \mathbb{R}^{n_w}$. To efficiently handle structured uncertainty without the need to account for every potential value in $\mathcal{W}$, we propose the notion of a *representative*. These representatives, denoted as $w_j \in \mathcal{W}, j \in \{0,\dots,v\}$, where v represents the total number of representatives, are particular points selected from the set of uncertainties to adequately capture the spectrum of uncertainties. They serve as practical stand-ins for the broader set of uncertainties $\mathcal{W}$, allowing for a more manageable analysis and optimization of the system under study. Finally, it is assumed that the uncertainty w comes from a probability distribution $\mathbb{P}$.

The last real-life phenomena that will appear in this work, but are not concretely reflected in (1) are the input and measurement delay, which are embedded in the true system matrices $H(\mathfrak{q};w)$, $L(\mathfrak{q};w)$, and $F(\mathfrak{q};w)$, and the measurement noise. Each input delay is characterized as $\tau_i^{(u)} \in \mathbb{R}$ time steps, where $i \in \{1\dots n_u\}$ and n_u represent the number of inputs. Similarly, the output delay is characterized as $\tau_j^{(y)} \in \mathbb{R}$ time steps, where $j \in \{1\dots n_y\}$ and n_y represents the number of inputs. The measurement noise, characterized by the variable $\eta \in \mathbb{R}^{n_y}$, consists of independent Gaussian white noise signals that affect the output measurements y embedded in the variable z. Let us now elaborate on how the mentioned *real-life phenomena* play a role within the DAE through the system dynamics depicted (and enclosed by *system boundaries*) in Fig. 1. One can define a set of causal linear time-invariant difference equations

$$\begin{cases} GX(k+1)=A(w)X(k)+B_u(w)u(k-\tau_u)+B_d(w)d(k) \\ +B_f(w)\begin{bmatrix} f_{\mathrm{a}}(k)+E^{\intercal}z(k-\tau_1^{(u)})f_{\mathrm{m}}(k) \\ \vdots \\ f_{\mathrm{a}}(k)+E^{\intercal}z(k-\tau_{n_u}^{(u)})f_{\mathrm{m}}(k) \end{bmatrix}, \\ \begin{bmatrix} y_1(k+\tau_1^{(y)}) \\ \vdots \\ y_{n_y}(k+\tau_{n_y}^{(y)}) \end{bmatrix} = C(w)X(k)+D_\eta\eta(k)+D_d(w)d(k) \end{cases} \quad (2)$$

where the matrices G, $A(w)$, $B_u(w)$, $B_d(w)$, $B_f(w)$, $C(w)$, and $D_d(w)$ are constant matrices with appropriate dimensions as a function of the (time-invariant) uncertainty w. The matrix D_η (in this work assumed to be diagonal) selects the noise signals η to be added to the output y. By defining $z := [y;u]$, $x := [X;d;\eta]$, and $E^{\intercal} = [0 \quad I]$, we can (2) as (1) with

$$H(\mathfrak{q};w)=\begin{bmatrix} I & 0 \\ 0 & Y(\mathfrak{q}) \end{bmatrix}\begin{bmatrix} -\mathfrak{q}G+A(w) & B_d(w) & 0 \\ C(w) & D_d(w) & D_\eta \end{bmatrix}, \quad (3a)$$

$$L(\mathfrak{q};w)=\begin{bmatrix} U(\mathfrak{q}) & 0 \\ 0 & I \end{bmatrix}\begin{bmatrix} 0 & B_u(w) \\ -I & 0 \end{bmatrix}, F(\mathfrak{q};w)=\begin{bmatrix} B_f(w) \\ 0 \end{bmatrix}. \quad (3b)$$

Moreover, the polynomial matrices $Y(\mathfrak{q})$, $U(\mathfrak{q})$ are diagonal polynomial matrices of size $n_y \times n_y$, $n_u \times n_u$, respectively, containing the delays of measurements and control inputs, i.e.,

$$Y(\mathfrak{q}) = \begin{bmatrix} \mathfrak{q}^{-\tau_1^{(y)}} & & \\ & \ddots & \\ & & \mathfrak{q}^{-\tau_{n_y}^{(y)}} \end{bmatrix}, U(\mathfrak{q}) = \begin{bmatrix} \mathfrak{q}^{-\tau_1^{(u)}} & & \\ & \ddots & \\ & & \mathfrak{q}^{-\tau_{n_u}^{(u)}} \end{bmatrix},$$

In the setting described above, the main objective of this study is to solve the following problem.

Problem. *Consider the DAE system* (1)*,* (3) *with the available measurement signal z under the influence of measurement noise η and the characterizations of the delay of the input and output $U(\mathfrak{q}), Y(\mathfrak{q})$, the uncertain parameters in $w \in \mathcal{W}$ and the multivariate signal $f = [f_{\mathrm{a}}^{\intercal}, f_{\mathrm{m}}^{\intercal}]^{\intercal}$ comprising both additive and multiplicative faults. We aim to design a diagnosis filter that turns the signal z to a signal $\widehat{f}$ (i.e., a causal dynamic mapping $z \mapsto \widehat{f}$), which is an accurate estimate of the fault signal f.*

In this work, we require the aggregated fault signal, now defined as $f_{agg} := f_{\mathrm{a}} + E^{\intercal}zf_{\mathrm{m}}$, where $f_{agg} \in \mathbb{R}$, to be detectable within the system. That is, we can detect and estimate the signal's presence or absence, irrespective of any other faults or disturbances acting on the system. This is formalized in the following assumption.

Assumption II.1 (Detectability)**.** *Given the system in* (1)*,* (3)*, in the absence of noise (that is, $\eta = 0$) and delay (i.e., $Y(\mathfrak{q}), U(\mathfrak{q}) = I$), and with uncertainty $w \in \mathcal{W}$. The polynomial matrices $H(\mathfrak{q}; w)$ and $F(\mathfrak{q}; w)$ in* (1)*,* (3) *satisfy the necessary and sufficient rank condition Rank $\{[H(\mathfrak{q}; w), F(\mathfrak{q}; w)]\}$ > Rank $\{H(\mathfrak{q}; w)\}$, $\forall w \in \mathcal{W}$. For simplicity of exposition, we further assume that $F(\mathfrak{q}; w)$ is a polynomial column vector, i.e., $n_{f_\mathrm{a}} = n_{f_\mathrm{m}} = 1$.*

Assumption II.1 enables us to design a filter that detects the aggregated fault signal f_{agg}. However, this process requires taking measurements of the uncertain parameters w while assuming that there is no delay nor noise.

B. State-of-the-art on estimation of multivariate faults

In the scope of the above problem description, some previous work has been carried out to estimate the two faults, $f_\mathrm{a}, f_\mathrm{m}$ (in the absence of uncertainty, delays, and noise, i.e., $w = w_0$, $Y(\mathfrak{q}) = U(\mathfrak{q}) = I$, and $\eta = 0$). Initially, we consider the LTI scenario, as described in [4]. Also, we assume that the system is free from noise and delay. The proposed approach has two steps. First, an estimation of the aggregated fault signal f_{agg} is performed. This is achieved by applying a suitable LTI estimation filter $N(\mathfrak{q}; w_0)$ to $L(\mathfrak{q}; w_0)[z]$ (which only requires the measurable input signals) as follows:

$$r \coloneqq a^{-1}(\mathfrak{q}) N(\mathfrak{q}; w_0) L(\mathfrak{q}; w_0)[z], \tag{4}$$

where the filter is generated using the linear program [Eq. (8)][4] and r represents the so-called residual. The denominator $a(\mathfrak{q})$ is intended to make the estimation filter proper. Using this residual generator, in view of the dynamical system (1), (3) and given assumption II.1, we can design a filter such that the following conditions hold:

$$N(\mathfrak{q}; w_0) H(\mathfrak{q}; w_0) = 0, \tag{5a}$$
$$a^{-1}(1) N(\mathfrak{q}; w_0) F(\mathfrak{q}; w_0) = 1. \tag{5b}$$

Here, (5a) ensures the rejection of unknown signals in the residual, and (5b) ensures that the residual generator (4) can estimate the aggregated fault f_{agg} in steady state. By combining the conditions (5) and applying them to (1), the residual (4) can equivalently be written as

$$r = -a^{-1}(\mathfrak{q}) N(\mathfrak{q}; w_0) F(\mathfrak{q}; w_0)[f_{agg}]. \tag{6}$$

This shows that such a residual generator results in a direct mapping between z and f_{agg}, and due to (5b) the residual estimates the fault in steady-state. The second step in the approach, towards the estimation of the individual faults in f (i.e., f_a and f_m), is to use the regression operator [4, Definition 3.2], which estimates the separate faults as

$$\Phi_n[e, r](k) \coloneqq \phi_n^\dagger[e](k)\, \mathbf{r}_n(k), \tag{7}$$
$$\text{where} \quad \phi_n[e](k) \coloneqq \left[\mathbf{e}_n(k),\ \mathbb{1}_n\right] \in \mathbb{R}^{n \times 2},$$

where the output $\Phi_n[e, r](k)$ represents the estimated faults $[\widehat{f_\mathrm{a}}, \widehat{f_\mathrm{m}}]^\intercal$ at time k. In (7), the regression horizon n determines how much past information from the residual (in $\mathbf{r}_n$) and the input is considered to estimate the fault signals. The signal e in (7) is obtained by filtering the signal $E^\intercal z$ through a pre-filter (see Fig. 1), as proposed in [4, Theorem 3.7], i.e.,

$$e = -a^{-1}(\mathfrak{q}) N(\mathfrak{q}) F(\mathfrak{q}; w_0)[E^\intercal z]. \tag{8}$$

The fault estimates obtained are paired with a performance bound that is demonstrated to be tight in [4], enabling users to gain an understanding of potential sources of error and ways to enhance filter performance. In addition, these performance bounds show that it is possible to estimate the individual fault components as long as the variance of the input signal e is non-zero and the signal is bounded, and the filter in Eq. (6) is stable [4, Proposition 3.3]. Effectively, this makes the approach bounded-input bounded-output (BIBO) stable. Given bounded signals $E^\intercal z, f_\mathrm{a}, f_\mathrm{m}$, the estimation error will remain bounded. This fundamental condition will remain valid throughout this study. In an ideal LTI setting, without uncertainty, delay, and noise, the approach is effective; see [4]. Each of the real-life phenomena, i.e., uncertainty w, measurement noise η, and the delayed versions of the variables x, z, f_{agg} will impact the residual r. In a healthy system, i.e., $f_{agg} = 0$, the residual (4) in the absence of real-life phenomena will also be zero, i.e., $r(w_0) = 0$, as can be derived from (6). Striving for such a property for a healthy system in the presence of real-life phenomena, combined with (5b), will allow the estimation of faults in a steady state. In the next section, we make explicit how these real-life phenomena arise in the context of automated vehicles. This motivates, first, the formulation of a generic problem setting in Sec. II-D and, secondly, supports the application of the general methodology in Sec. III to fault estimation in the automated driving context in Sec IV. Note that, driven by the experimental application, we only consider the estimation of faults acting on the input of the system in this work. However, the methodology can be applied to any system that allows its fault and dynamics to be described as in (1), as a result allowing the estimation of, e.g., input/output faults as well as faults that appear further down in the dynamics of the system.

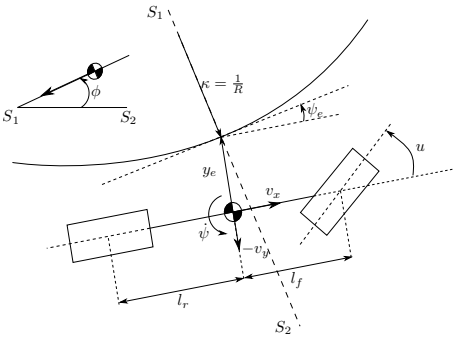


Fig. 2: Visual representation of the bicycle model for a vehicle.

C. Real-world challenges in automated vehicles

Automated vehicles are susceptible to faults in the lateral steering actuation, such as a bias of the actuator with respect to the desired setpoint (i.e., the occurrence of f_a), or a loss of effectiveness of the actuator that executes the setpoint of

the steering (i.e., the occurrence of f_{m}). If these faults are not compensated for, they can lead to potentially dangerous vehicle behavior, especially in the lane-keeping driver assistance system. Using the model description in Eq. (2), which can be rewritten as Eq. (1), we can analyze the lateral dynamics of the automated vehicle in the presence of uncertainties, measurement noise, and delay. This model is illustrated by the mechanical model in Fig. 2. The state of the vehicle is represented by a vector $X = [v_y, \dot{\psi}, y_e, \psi_e]^\intercal$, which includes the lateral velocity (v_y), the yaw rate ($\dot{\psi}$), lateral error from the lane center (y_e), and heading error from the center of the lane (ψ_e). The disturbance, d, is represented by a scalar variable, κ, which indicates the curvature of the road. The input, u, indicates the input to the vehicle, i.e., the steering angle of the front wheels. The remainder of the model can be described by (2) by using the following continous-time state-space matrices:

$$\begin{cases} \bar{A} = \begin{bmatrix} w^{(1)}\frac{C_f+C_r}{v_x m} & w^{(1)}\frac{l_f C_f - l_r C_r}{v_x m} - v_x & 0 & 0 \\ w^{(2)}\frac{l_f C_f - l_r C_r}{v_x I} & w^{(2)}\frac{l_f^2 C_f + l_r^2 C_r}{v_x I} & 0 & 0 \\ -1 & 0 & 0 & v_x \\ 0 & -1 & 0 & 0 \end{bmatrix}, \\ \bar{B}_u = \begin{bmatrix} -w^{(1)}\frac{C_f}{m} & -w^{(2)}\frac{l_f C_f}{I} & 0 & 0 \end{bmatrix}^\intercal, \bar{B}_f = \bar{B}_u, \\ \bar{B}_d = \begin{bmatrix} 0 & 0 & 0 & v_x \end{bmatrix}^\intercal, G = C = D_\eta = I, \\ D_d = 0, E = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^\intercal, \end{cases}$$

and their discrete-time equivalents:

$$\begin{cases} A = e^{\bar{A}h}, & B_u = \int_0^h e^{\bar{A}s}\bar{B}_u ds, \\ B_d = \int_0^h e^{\bar{A}s}\bar{B}_d ds, & B_f = \int_0^h e^{\bar{A}s}\bar{B}_f ds. \end{cases}$$

The uncertainty in this model is characterized by $w = [w^{(1)}, w^{(2)}]^\intercal$, where $w^{(1)}$ expresses the uncertainty that appears in the vehicle mass m and the stiffness of the front and rear corners C_f, C_r, respectively. The uncertainty value $w^{(2)}$ expresses the uncertainty that appears in the yaw moment of inertia I and the corner stiffness C_f, C_r, respectively. The mass and yaw moment of inertia may change according to the loading conditions of the vehicle. These uncertainties may lead to false positives in the detection of faults in the steering actuator. This can occur if incorrect parameters are used in the approach of Sec. II-B, such as when carrying multiple passengers or additional luggage. The cornering stiffness C_f, C_r may vary on the basis of tire conditions, their pressure, and the load on the axles or weather conditions.

D. Generic problem description

We will now take a closer look at how the real-life phenomena we propose affect the residual r, leading to an inaccurate estimate of the faults f. Each subsection will begin with a brief description of the objective to solve the problem description, which will be addressed in Sec. III.

1) Model uncertainty: The residual generator (4) estimates the aggregated fault under the assumption that $w_0 = w$, that is, the parameters in the dynamical system are known. Now, let us assume that the system parameters in (1) are defined by $w \neq w_0$. In that case, we can rewrite (4) as follows:

$$r(w_0, w, f_{agg}) = a^{-1}(\mathfrak{q})N(\mathfrak{q}; w_0)L(\mathfrak{q}; w_0)[z(w)]. \tag{10}$$

Here, we explicitly denote that the variable $z(w)$ (as well as the unknown variable $x(w)$) is driven by the system (1) with parameters $w \neq w_0$.

Remark II.2. *For the remainder of this work, it is important to note that the variables $x(w)$ and $z(w)$, along with the signals within them, are driven by the system* (1)*, which is subject to uncertainty from the variable w. However, to simplify the notation, we will not explicitly include the dependence of these variables on w.*

Now, let us characterize the mapping from the residual $r(w_0, w, f_{agg})$ (10) to the true aggregated fault f_{agg}. First, we rewrite the model (1) with model mismatch, given a nominal model with assumed parameter values w_0 and actual parameter values w, as follows:

$$\begin{aligned} &(H(\mathfrak{q}; w) - H(\mathfrak{q}; w_0))[x] + H(\mathfrak{q}; w_0)[x] + F(\mathfrak{q}; w)[f_{agg}] + \\ &(L(\mathfrak{q}; w) - L(\mathfrak{q}; w_0))[z] + L(\mathfrak{q}; w_0)[z] = 0, \end{aligned} \tag{11}$$

which characterizes the model mismatch between the actual system and the nominal dynamics. Substituting (11) into (10), results in the following residual:

$$\begin{aligned} r(w_0, w, f_{agg}) = &\underbrace{-a^{-1}(\mathfrak{q})\, N(\mathfrak{q}; w_0)\, \underbrace{(H(\mathfrak{q}; w) - H(\mathfrak{q}; w_0))}_{\Delta H(\mathfrak{q}; w, w_0)}[x]}_{(\mathbf{I})} \\ &\underbrace{+ a^{-1}(\mathfrak{q})\, N(\mathfrak{q}; w_0)\, \underbrace{(L(\mathfrak{q}; w) - L(\mathfrak{q}; w_0))}_{\Delta L(\mathfrak{q}; w, w_0)}[z]}_{(\mathbf{II})} \\ &+ a^{-1}(\mathfrak{q})N(\mathfrak{q}; w_0)F(\mathfrak{q}; w)[f_{agg}], \end{aligned} \tag{12}$$

which shows that the residual is a function of the aggregated fault f_{agg}, the unknown signals x, and the measured signals z. This may not be a good fault indicator when the terms (**I**), (**II**) in (12) are nonzero (i.e., in the presence of model uncertainty). Moreover, this effect of model mismatch propagates through the isolation filter, which assumes that the residual depends only on the faults f_{a}, f_{m} and $E^\intercal z$. This allows us to define the first objective towards solving the problem statement.

Objective 1. *Consider a system* (1)*,* (3) *in the presence of structured uncertainty $w \in \mathcal{W}$ with representatives $w_j \in \mathcal{W}$ and absence of delay and noise. Due to the structure of the uncertainty, employ the representatives $\{w_1 \dots w_v\}$ to minimize the model mismatch terms* (***I***)*,* (***II***) *in* (12) *of a healthy system (i.e., $f_{agg} = 0$), through a filter $N(\mathfrak{q}; w_0)$. In the scope of this objective we aim to find such a filter by minimizing the mismatch from an average point-of-view, i.e.,*

$$\min_{N(\mathfrak{q}; w_0)} \frac{1}{v}\sum_{j=1}^{v} \|r(w_0, w_j, 0)\|_{\mathcal{L}_2}^2 \tag{13}$$

$$\text{s.t.} \quad \text{(5a), (5b), at nominal } w_0.$$

Furthermore, a second aim in the scope of this objective is to perform such a minimization in a robust sense, i.e., minimizing the worst case, as follows:

$$\min_{N(\mathfrak{q};w_0)} \max_{j\leq v} \|r(w_0,w_j,0)\|_{\mathcal{L}_2}^2 \quad \text{s.t.} \quad \text{(5a), (5b), at nominal } w_0. \tag{14}$$

2) Input and measurement delay: The residual generator (4) originally assumes the absence of delay, uncertainty, and measurement noise in the system. We can evaluate how it would perform on the system with input and output delay, by inserting the system matrices, containing the effects of delay (3) (without the effects of noise, i.e., $D_\eta = 0$ and uncertainty, i.e., $w = w_0$), into the filter (4) as follows:

$$\begin{aligned} r_\tau(w_0,w_0,f_{agg}) =& a^{-1}(\mathfrak{q})N(\mathfrak{q};w_0)L(\mathfrak{q};w_0)[z], \\ =& -a^{-1}(\mathfrak{q})N(\mathfrak{q};w_0)U(\mathfrak{q})F(\mathfrak{q};w_0)[f_{agg}] \\ &+a^{-1}(\mathfrak{q})N(\mathfrak{q};w_0)\begin{bmatrix}0 & (I-U(\mathfrak{q}))B_u \\ 0 & 0\end{bmatrix}[z] \\ &+a^{-1}(\mathfrak{q})N(\mathfrak{q};w_0)\begin{bmatrix}0 & 0 \\ (I-Y(\mathfrak{q}))C & (I-Y(\mathfrak{q}))D_d\end{bmatrix}[x], \end{aligned}$$

which shows that r_τ, representing the residual of the delayed system with filter (4), is a function of the fault f_{agg} and delayed instances of the known signals in z and unknown signals in x, which could result in false positives for the fault detection. This allows us to define our second objective towards solving the problem statement.

Objective 2. *Consider a system* (1)*,* (3) *in the presence of input and measurement delay, and the absence of uncertainty w (i.e., $w = w_0$) and noise. Minimize the effect of the delay in the inputs and measurements in* (12)*, through a filter $N(\mathfrak{q};w_0)$, satisfying* (5)*. This is equivalent to designing $N(\mathfrak{q};w_0)$ for a healthy system, i.e., $f_{agg} = 0$, by solving the following optimization problem:*

$$\min_{N(\mathfrak{q};w_0)} \|r_\tau(w_0,w_0,0)\|_{\mathcal{L}_2} \quad \text{s.t.} \quad \text{(5a), (5b), at nominal } w_0. \tag{15}$$

3) Measurement noise: In this section, we evaluate the residual generator (4) in the absence of uncertainty (i.e., $w = w_0$) and delay (i.e., $U(\mathfrak{q}) = Y(\mathfrak{q}) = 1$). Using the detectability condition from Assumption II.1, the first intuition could be to handle measurement noise as an unwanted disturbance, as in (3a), by modeling the state/disturbance matrix as

$$H(\mathfrak{q};w_0) = \begin{bmatrix} -\mathfrak{q}G + A & B_d & 0 \\ C & D_d & D_\eta \end{bmatrix}. \tag{16}$$

The goal is then to find a filter that cancels out the effects of the noise η by finding a filter polynomial $N(\mathfrak{q};w_0)$ which belongs to the nullspace of (16). This approach might work for systems where not all measurements are affected by noise (i.e., having measurement redundancy) or through a possible linear independence between C and D_η. However, in practice, this is often not the case. For example, for the automated vehicle application (i.e., the system (1), (3) with matrices as in (9)), we consider $C = D_\eta = I$. This relates to the case where all states are measured and all are affected by measurement noise. Numerical analysis shows us that the rank condition in Assumption II.1 does not hold. This can be explained by the intuition of finding a filter that cancels the third block column in (16) which, given $D_\eta = I$, has an empty basis (i.e., the filter coefficients of $N(\mathfrak{q})$ that multiply with D_η are equal to 0). This implies that a residual generator, designed such that (5) hold, i.e.,

$$r(w_0,w_0,f_{agg}) = a^{-1}(\mathfrak{q})N(\mathfrak{q};w_0)\begin{bmatrix}0 & B_u \\ -I & 0\end{bmatrix}[z], \tag{17}$$

would cancel the effect of the first block column of $L(\mathfrak{q};w_0)$, i.e., the contribution of all available measurements. This restricts the residual generator to only use the input signal u to estimate the presence of f_{agg} in the system. In that case, it is not possible to estimate f_{agg} in the vehicle context, as $E^\intercal z = u$ is affected by the fault and this affected input cannot be measured (only the unaffected signal u is measured). Leaving out the measurement noise term in the matrix (16), i.e., considering a nominal design, one would be able to generate a residual generator. However, when finding a filter according to conditions (5) for (1) and (3a), the residual would be described as follows:

$$\begin{aligned} r(w_0,w_0,f_{agg}) = a^{-1}\big(\mathfrak{q}\big)\big(&N(\mathfrak{q};w_0)L(\mathfrak{q};w_0)[z] \\ &+ N(\mathfrak{q};w_0)[0\ \ D_\eta]^\intercal[\eta]\big), \end{aligned} \tag{18}$$

which results in a residual depending on the filtered aggregated fault and a filtered version of the measurement noise, which propagates further through the isolation filter and affects the quality of the fault estimates. Since our setting does not allow full decoupling of η, the treatment of noise in the residual (18) and therefore in the fault estimates $\widehat{f}_{\mathrm{a}}, \widehat{f}_{\mathrm{m}}$ requires a classical trade-off between estimation speed and accuracy. This brings us to the third and final objective.

Objective 3. *Consider a system* (1)*,* (3) *in the presence of measurement noise η and the absence of uncertainty w (i.e., $w = w_0$) and input/measurement delay (i.e., $U(\mathfrak{q}) = Y(\mathfrak{q}) = 1$). Characterize the trade-off in attenuating the effect of measurement noise η at the cost of the estimation speed and estimation accuracy of the faults $\widehat{f}_{agg}$, $\widehat{f}_{\mathrm{a}}, \widehat{f}_{\mathrm{m}}$.*

E. Outline of the proposed approach

As explained in Sec. II-A, detecting and estimating the aggregated fault f_{agg} and separating the additive fault f_{a} and the multiplicative fault f_{m} from it is the main challenge in fault isolation. It becomes even more challenging when their impacts on the dynamics are linearly dependent. Our study differs from the one conducted in [4] in that we take into account several real-life phenomena (which, e.g., arise in an experimental setting for automated driving) characterized by Objective 1, 2, and 3. The proposed approach involves incorporating prior knowledge of real-life phenomena such as the uncertainty representatives w_j, measurable delays $Y(\mathfrak{q}), U(\mathfrak{q})$, and characterization of measurement noise η. In the fault diagnosis part, we robustify the detection and isolation approach presented in [4, Theorem 3.7] while making use of the insights gained by the performance bounds developed in that work. The system, affected by uncertainty, delay, and noise, is

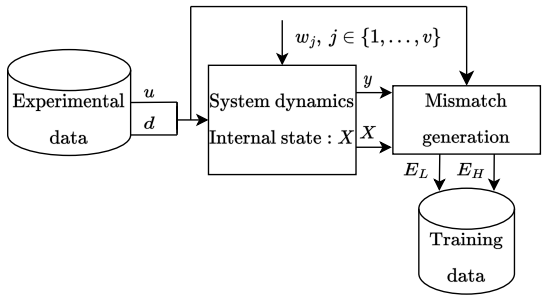


Fig. 3: Depiction of the uncertainty training procedure.

depicted on the left-hand side in Fig. 1. The proposed approach first detects and estimates the aggregated fault f_{agg} through a robustified filter, which is of the form in (4) and will be robustified in Sec. III for model uncertainty and delay. The individual contributions f_{a} and f_{m} are then estimated through a pre-filter and isolation filter, of the form in (7) and (8), respectively, and further elaborated on in Sec. II-D3 in the scope of noise attenuation. These blocks comprise the fault diagnosis algorithm, which is shown on the right-hand side in Fig. 1. We aim to achieve improved fault estimates compared to the baseline approach through robustification and prove its benefits by conducting real experimental field tests in the scope of automated driving.

III. Robustification procedure

A. Model uncertainty

In Section II-D1, we designed a nominal filter for an uncertain description of the system to detect the aggregated fault of interest. However, it is important to note that the residual r that we use to differentiate between faults f_{a} and f_{m} in that specific case depends on various factors, such as the mismatch between the assumed dynamics in the filter design versus the true dynamics of the system (12). Therefore, it is necessary to adjust the filter requirements in (5) to reflect the sensitivity to faults and the insensitivity to disturbances and minimize the effect of the model mismatch on the residual. Given (13) and (14) from Objective 1, the goal is to minimize the residual of a healthy system, thus minimizing the effect of model mismatch. Using (12) and setting $f = 0$ (i.e., a healthy system), this is equivalent to minimizing

$$\|a^{-1}(\mathfrak{q})N(\mathfrak{q};w_0)\big(\Delta H(\mathfrak{q};w,w_0)[x]+\Delta L(\mathfrak{q};w,w_0)[z]\big)\|_{\mathcal{L}_2}. \quad (19)$$

The approach for mitigating the effect of model uncertainty is an adaptation of the results presented in [27], where a similar detection filter (with the same objectives as in (5)) was trained to detect discrepancies between a high-fidelity simulator and an abstract linear model. The linear model represents a simplified representation of the high-fidelity simulator. In [27], the main source of the mismatch originates from nonlinearities. However, in that approach, the high-fidelity simulator does not have a representative internal state X, as such, [27, Eq. (6) (I)] can only be minimized by assuming linear state and disturbance dynamics. In our approach, depicted in Fig. 3, the main objective is to minimize the contribution of both (**I**) and (**II**) from (11) by using prior knowledge about the uncertain system and its behavior in an experimental setting. This will be described in more detail below.

Our approach utilizes inputs, gathered in experiments to, first, represent relevant healthy scenarios and, second, enable the characterization of relevant system behavior in these healthy scenarios. In the automated vehicle application, these inputs are the steering angle and curvature (9). These inputs, gathered from experimental data and used for simulation towards mismatch generation, are characterized as follows:

$$\mathbf{u} \coloneqq [u[1], u[2], \dots, u[T]],\ \mathbf{d} \coloneqq [d[1], d[2], \dots, d[T]], \quad (20)$$

where $\mathbf{u} \in \mathbb{R}^{n_u \times T}$, $\mathbf{d} \in \mathbb{R}^{n_d \times T}$, and where T is the total number of collected data samples. Note that it may not be possible to collect data on all disturbances d. In that case, it will be assumed that the specific disturbance signal is zero over the time horizon $[1, T]$. We then use these input matrices $\mathbf{u}$ and $\mathbf{d}$ to simulate the system output y and its internal state X, in scenarios represented by the inputs and disturbances in these matrices, on all representatives w_j of the uncertainty set. The resulting state and output evolutions are then characterized as follows.

$$\mathbf{y}_j \coloneqq [y_j[1], y_j[2], \dots, y_j[T]],\ \forall j \in [1 \dots v], \quad (21)$$
$$\mathbf{X}_j \coloneqq [X_j[1], X_j[2], \dots, X_j[T]],\ \forall j \in [1 \dots v]. \quad (22)$$

Having access to the synthetic training sets of time-series data of the output y, the state X, the input u, and disturbance d allows us to collect time-series data of the mismatch signatures (**I**) and (**II**) in (12). To write a finite-time version of (**I**) and (**II**) using (21), (22), which can be used to formulate a (optimization) program for finding an admissible filter $N(\mathfrak{q}; w_0)$ that meets the Objective 1, we use the results of [4, Lemma 3.1]. As an illustrative example, we decompose the matrix $H(\mathfrak{q}; w_0) = \sum_{i=0}^{d_H} H_i(w_0)\mathfrak{q}^i$, where d_H denotes the degree of $H(\mathfrak{q}; w_0)$, and $H_i(w_0) \in \mathbb{R}^{n_r \times n_x}$. Then, we can define $\bar{N}(w_0) \coloneqq \begin{bmatrix} N_0, N_1, \dots, N_{d_N} \end{bmatrix} \in \mathbb{R}^{1 \times d_N \cdot n_r}$, where d_N denotes the degree of $N(\mathfrak{q}; w_0)$, $N_i \in \mathbb{R}^{1 \times n_r}$, and

$$\bar{H}(w_0) \coloneqq \begin{bmatrix} H_0 & H_1 & \dots & H_{d_H} & 0 & \dots & 0 \\ 0 & H_0 & H_1 & \dots & H_{d_H} & 0 & \vdots \\ \vdots & & \ddots & \ddots & & \ddots & 0 \\ 0 & \dots & 0 & H_0 & H_1 & \dots & H_{d_H} \end{bmatrix}, \quad (23)$$

where we drop w_0 from the entries of $\bar{H}(w_0), \bar{N}(w_0)$ for compact notation, and $\bar{H}(w_0) \in \mathbb{R}^{(d_N \cdot n_r) \times ((d_N + d_H) \cdot n_x)}$. This allows us to rewrite (5a) as follows:

$$N(\mathfrak{q}; w_0)H(\mathfrak{q}; w_0) = \bar{N}(w_0)\bar{H}(w_0)[I, \mathfrak{q}I, \dots, \mathfrak{q}^{d_N + d_H}]^\intercal,$$

such that the linear formulation $\bar{N}\bar{H} = 0$ is equivalent to (5a). Similarly, we can decompose $F(\mathfrak{q}; w_0), L(\mathfrak{q}; w_0)$ (which are of degree d_F and d_L), by substituting (23) with F and L, respectively, thus retrieving $\bar{F}(w_0), \bar{L}(w_0)$. Finally, we can decompose $a(\mathfrak{q}) = \bar{a}[1, \mathfrak{q}, \dots, \mathfrak{q}^{d_a}]^\intercal$ where $\bar{a} = [a_0, a_1, \dots, a_{d_a}]$, where d_a denotes the degree of $a(\mathfrak{q})$. This allows us to rewrite (5b) as $\bar{N}\bar{F}(w_0)\mathbb{1}_{d_N \times d_F} = -\bar{a}\mathbb{1}_{d_a}$, as is also shown in [4, Lemma 3.1].

Now, by taking (**I**) and (**II**) from (12) for each representative $w_j \in \{w_1, \dots, w_v\}$, and replacing the variable x with a finite

time version $[\mathbf{X}_j^\intercal,\ \mathbf{d}^\intercal]^\intercal$, and replacing z with a finite time version $[\mathbf{y}_j^\intercal,\ \mathbf{u}^\intercal]^\intercal$, respectively, we find

$$N(\mathfrak{q};w_0)\Delta H(\mathfrak{q};w_j,w_0)\begin{bmatrix}\mathbf{X}_j\\ \mathbf{d}\end{bmatrix}= \underbrace{\bar{N}(w_0)\,(\bar{H}(w_j)-\bar{H}(w_0))[I,\mathfrak{q}I,\dots,\mathfrak{q}^{d_N+d_H}]^\intercal\begin{bmatrix}\mathbf{X}_j\\ \mathbf{d}\end{bmatrix}}_{E_{H,j}},$$

$$N(\mathfrak{q};w_0)\Delta L(\mathfrak{q};w_j,w_0)\begin{bmatrix}\mathbf{y}_j\\ \mathbf{u}\end{bmatrix}= \underbrace{\bar{N}(w_0)\,(\bar{L}(w_j)-\bar{L}(w_0))[I,\mathfrak{q}I,\dots,\mathfrak{q}^{d_N+d_L}I]^\intercal\begin{bmatrix}\mathbf{y}_j,\\ \mathbf{u}\end{bmatrix}}_{E_{L,j}},$$

The resulting model-mismatch terms $E_{H,j},\ E_{L,j}\in\mathbb{R}^{n_r\cdot d_N\times T}$ allow us to formulate (19) for a finite-time residual of length T, i.e., $\mathbf{r}_{T,j}$ where $j\in\{1\dots v\}$, as follows:

$$\|\mathbf{r}_{T,j}\|_2^2=\bar{N}\Big(E_{H,j}E_{H,j}^\intercal+E_{L,j}E_{L,j}^\intercal\Big)\bar{N}^\intercal. \tag{24}$$

As noted in [27, Remark 2], training a model for multiple model mismatch signatures resulting from a variety of uncertain systems can be approached in different ways. First, an *average-cost* approach can be taken, which weighs the effect of all residuals (24) equally, which is equivalent to minimizing

$$\frac{1}{v}\sum_{j=1}^{v}\|\mathbf{r}_{T,j}\|_2^2=\bar{N}\Big(\frac{1}{v}\sum_{j=1}^{v}\big(E_{H,j}E_{H,j}^\intercal+E_{L,j}E_{L,j}^\intercal\big)\Big)\bar{N}^\intercal. \tag{25}$$

The formulation in (25) allows us to formulate a first filter design according to (13). One of the representatives w_j may result in a much larger mismatch compared to the other representatives. In such a case, the focus should be on the representative that results in the most severe mismatch, which can be achieved by minimizing the worst-case mismatch. This is also known as a *worst-case* approach, which involves minimizing the worst-case mismatch from (24), i.e., minimizing

$$\max_{j\le v}\|\mathbf{r}_{T,j}\|_2^2=\max_{j\le v}\bar{N}\Big(E_{H,j}E_{H,j}^\intercal+E_{L,j}E_{L,j}^\intercal\Big)\bar{N}^\intercal. \tag{26}$$

The formulation in (26) allows us to formulate a first filter design according to (14). Using (25) and (26), two designs are introduced that allow us to satisfy the two optimality criteria (13) and (14), as such that we meet the Objective 1.

Design 1 (Average-cost robust fault estimator for structured uncertainty)**.** *Consider the uncertain system in* (11)*, where the parametric uncertainties are characterized by $\mathcal{W}$ with representatives $w_j, j\in\{1\dots v\}$, with a nominal known value w_0. An average-cost filter, according to* (13) *in Objective 1, can be found by minimizing* (25) *for all uncertainty representatives while satisfying* (5) *for nominal w_0. This is equivalent to solving the following quadratic program:*

$$\begin{aligned}\min_{\bar{N}}\quad & \bar{N}\Big(\tfrac{1}{v}\sum_{j=1}^{v}E_{H,j}E_{H,j}^\intercal+\tfrac{1}{v}\textstyle\sum_{j=1}^{v}E_{L,j}E_{L,j}^\intercal\Big)\bar{N}^\intercal\\ \text{s.t.}\quad & \bar{N}\bar{H}(w_0)=0\\ & \bar{N}\bar{F}(w_0)\mathbb{1}_{d_N\times d_F}=-\bar{a}\mathbb{1}_{d_a}.\end{aligned} \tag{27}$$

Similarly, thanks to (26), a second filter can be formulated according to (14) in Objective 1 can be formulated.

Design 2 (Worst-case robust fault estimator for structured uncertainty)**.** *Consider the uncertain system in* (11)*, where parametric uncertainties are characterized by $\mathcal{W}$ with representatives $w_j, j\in\{1\dots v\}$, with a nominal known value w_0. A worst-case filter, according to* (14) *in Objective 1, can be found by minimizing* (26) *for the worst-impact uncertainty representatives, while satisfying* (5) *for nominal w_0. This is equivalent to solving the following quadratic program:*

$$\begin{aligned}\min_{\bar{N}}\max_{i\le v}\quad & \bar{N}\Big(E_{H,i}E_{H,i}^\intercal+E_{L,i}E_{L,i}^\intercal\Big)\bar{N}^\intercal\\ \text{s.t.}\quad & \bar{N}\bar{H}(w_0)=0\\ & \bar{N}\bar{F}(w_0)\mathbb{1}_{d_N\times d_F}=-\bar{a}\mathbb{1}_{d_a}.\end{aligned} \tag{28}$$

Designs 1 and 2 allow us to find an improved solution for an uncertain system compared to a nominal filter, as will also be shown in Sec. IV.

Generalization to unseen scenarios: The design perspectives in Designs 1 and 2 rely on specific scenarios w_j that are available to us often through experimental data (cf. Fig. 3 for a pictorial illustration of such a process). However, it is important to ensure that these designs are also reliable when facing unseen scenarios (i.e., plausible scenarios that are not considered in Designs 1 and 2) in real-time operation. This subject is at the heart of learning problems and is often referred to as "*generalization error*". In general, it is not possible to draw conclusions (i.e., generalization bound) from seen (training) scenarios to unseen (test) ones. However, under certain regularity conditions and for a specific choice of probabilistic guarantee, one can provide a formal performance certificate. An example is when the training phase optimizes the worst-case cost evaluated in the seen (training) scenarios (that is, Design 2) and the resulting program is (effectively) convex optimization in the decision variables [28], [33]. Let us elaborate more on this. We define the constraint function

$$g(\bar{N},\gamma,w_j):=\bar{N}\Big(E_{H,j}E_{H,j}^\intercal+E_{L,j}E_{L,j}^\intercal\Big)\bar{N}^\intercal-\gamma\,. \tag{29}$$

Using the definition of function g in (29), the program (28) of Design 2 can be rewritten in an epigraph reformulation as

$$\begin{aligned}\min_{\bar{N},\gamma}\quad & \gamma\\ \text{s.t.}\quad & g(\bar{N},\gamma,w_j)\le 0,\quad \forall j\in\{1,\dots,v\}\\ & \bar{N}\bar{H}(w_0)=0\\ & \bar{N}\bar{F}(w_0)\mathbb{1}_{d_N\times d_F}=-\bar{a}\mathbb{1}_{d_a}\,.\end{aligned}$$

The above reformulation of the worst-case (28) falls into the category of the so-called scenario convex problem (SCP). As shown in [34, Theorem 1], the solution of the SCP, denoted by $(\bar{N}^*,\gamma^*)$, enjoys the probabilistic guarantee as a feasible solution to the so-called chance-constrained program (CCP)

$$\mathbb{P}\big[w\in\mathcal{W}\ :\ g(\bar{N}^*,\gamma^*,w)\le 0\big]\ge 1-\varepsilon, \tag{30}$$

where $\mathbb{P}$ is the distribution supported on the uncertainty set $\mathcal{W}$, governing the behavior of the possible uncertain parameter w, and $\varepsilon\in[0,1]$ is a prespecified level of constraints violation.

Note that the CCP constraint takes into consideration the unseen scenario $w \in \mathcal{W}$ and allows for constraint violation up to a probability of ε. In the context of fault detection, this probability of violation is often referred to as the "*false-alarm rate*". The CCP guarantees can also be extended to a class of nonconvex problems, which has a direct application for fault detection problems [28]. It is also interesting to note that the average cost (27) in Design 1 can also benefit from some probabilistic guarantees. However, this is beyond the scope of this study and we refer the interested readers to [28, Theorem 4.11] for further information.

False negative/missed detection rate: We have two types of errors in fault detection problems: (i) false positive (aka false alarm rate) and (ii) false negative (aka missed detection rate). Looking at the design optimization programs (28), the constraint $\bar{N}\bar{H}(w_0) = 0$ and the first term of the objective $\bar{N}(E_{H,j}E_{H,j}^{\intercal})\bar{N}^{\intercal}$ are concerned with the false positive error, while the constraint $\bar{N}\bar{F}(w_0)\mathbb{1}_{d_N \times d_F} = -\bar{a}\mathbb{1}_{d_a}$ and the second term of the objective $\bar{N}(E_{L,j}E_{L,j}^{\intercal})\bar{N}^{\intercal}$ relate to the false negative error. Providing a performance certificate for the false negative is typically more challenging as it requires additional conditions on the fault signal as well. Namely, in a practical setting where we have to tolerate non-zero threshold (i.e., variable γ in constraint function g in (29)), there are always sufficiently small faults whose contributions to the residual are suppressed under this threshold, and hence remain undetected. To determine a minimum value (in the $\mathcal{L}$ sense) for a detectable fault signal, we need to ensure that the fault aggregated contribution exceeds the term $\bar{N}(E_{H,j}E_{H,j}^{\intercal} + E_{L,j}E_{L,j})\bar{N}^{\intercal}$. In other words, solving the worst case program (28) using the training data set provides us with a worst-case value $\gamma^* = \bar{N}(E_{H,j^*}E_{H,j^*}^{\intercal} + E_{L,j^*}E_{L,j^*})\bar{N}^{\intercal}$ evaluated in a particular scenario w_{j^*} (or equivalently the optimal objective of (28)), which offers a similar probabilistic chance constraint guarantee for the false negative rate in unseen scenarios. Having said that, we wish to emphasize that this would represent a false negative rate for aggregated faults, not a guarantee for each additive and multiplicative fault separately. Breaking down the false negative rate requires a more comprehensive non-trivial analysis, which is beyond the scope of this study and could serve as a potential direction for future research. In Sec. IV, we demonstrate the effectiveness of this approach using both synthetic and real experimental data.

B. Input and measurement delay

When there are delays in the input (actuation) and output (measurements) of a system, the current state-of-the-art approach does not meet the conditions stated in (5), as explained in Sec. II-D2. That is, the residuals are affected not only by the fault but also by previous instances of unknown states and disturbances. The reason for this problem is that the presence and length of delays are not considered model knowledge during filter synthesis. Fortunately, there are several methods in the literature to estimate delays within a system, as reported in [35]. In the context of our application, i.e. a compact actuator and sensor network with wired connections, it is safe to assume that all delay lengths are known and of constant length. Given these assumptions, we can assume that the polynomial matrices $Y(\mathfrak{q})$, $U(\mathfrak{q})$ in (3) are known and can be incorporated into the synthesis of a residual generator by rewriting (3), and setting:

$$H(\mathfrak{q};w) \coloneqq \begin{bmatrix} -\mathfrak{q}G + A(w) & B_d(w) & 0 \\ Y(\mathfrak{q})C(w) & Y(\mathfrak{q})D_d(w) & Y(\mathfrak{q})D_\eta \end{bmatrix},$$
$$L(\mathfrak{q};w) \coloneqq \begin{bmatrix} 0 & U(\mathfrak{q})B_u(w) \\ -I & 0 \end{bmatrix}, F(\mathfrak{q};w) \coloneqq \begin{bmatrix} U(\mathfrak{q})B_f(w) \\ 0 \end{bmatrix}.$$

This allows us to incorporate delays as model knowledge and compensate for them in the filter synthesis problem in Design 1and 2, enabling a filter robust against model uncertainty and input/measurement delays and satisfying Objective 2.

C. Measurement noise

As noted in Sec. II-D3, generally the measurement noise cannot be decoupled by modeling it as a disturbance, as it would imply the rejection of all available measurements, which will violate Assumption II.1 for our system. Therefore, a different method must be found to attenuate the effect of noise on the residual, to reduce its effect on fault estimates $\widehat{f}_{\mathrm{a}}$, $\widehat{f}_{\mathrm{m}}$. In contrast to the previous section, we will not be discussing any methodology to completely decouple the impact of noise. Instead, we will be explaining various approaches to reduce noise by adjusting parameters in different filter components. In addition, we will discuss how these strategies can impact the accuracy of the estimated faults. Measurement noise affects the aggregated fault f_{agg} through the second term in (18). The first way to minimize the effect of noise is through the residual generator: by minimizing the energy of the noise contribution. Based on (18), this implies

$$\min_{N(\mathfrak{q};w_0),a(\mathfrak{q})} \|a^{-1}(\mathfrak{q})N(\mathfrak{q};w_0)\begin{bmatrix}0 & D_\eta\end{bmatrix}^{\intercal}[\eta]\|_2.$$

The numerator $N(\mathfrak{q};w_0)$ is mainly useful for rejecting the noise present at specific frequencies. This is effective for rejecting disturbances within a particular frequency band (by using, for example, a band-stop filter). However, it is not very effective for Gaussian white noise, which has a flat frequency spectrum for all frequencies. Furthermore, the design of $N(\mathfrak{q};w_0)$ is already determined through Design 1 and 2. As such, using the design of the denominator $a(\mathfrak{q})$ to attenuate the effect of noise is a better option. This can be done through a variety of different filter types, e.g., low-pass filters. This helps to attenuate the contribution of noise above certain frequency levels. It is important to note, however, that filtering the residual to reduce noise can have a downside. As explained in the state-of-the-art (Sec. II-B), a pre-filter is applied to the input signal before it enters the regression operator. This pre-filter, as shown in (8), is designed to compensate for the dynamic mismatch between the residual and the true aggregated fault. The pre-filter design uses the same filter denominator $a(\mathfrak{q})$. When observing the proposed performance bounds in [4, Theorem 3.7], and its application to constant faults in [4, Corollary 3.8], there is a linear relationship between the variance of the filtered signal e and the magnitude of the performance bound. If a denominator

$a(\mathfrak{q})$ is therefore designed to aggressively attenuate noise, even within the frequency range of the input signal e, the "*excitation*" of the signal e is actively reduced, and therefore the bound on the fault estimation error increases, which could lead to performance loss regardless of the benefits of noise attenuation.

The second component in which we can attempt to attenuate the effect of measurement noise is in the isolation filter (7) as depicted in Fig. 1. Using [4, Eq. (17)], one can observe that the mismatch between the aggregated fault and the residual can be bounded by:

$$\|\Phi_n[e, r-\delta](k)\| \leq \frac{\mathcal{C}_n(\mathbf{e}_n)}{\sqrt{n V_n[e]}} \|\mathbf{r}_n - \boldsymbol{\delta}_n\|_2, \text{with } \delta = f_{\mathrm{a}} + e f_{\mathrm{m}}, \tag{31}$$

where the constant $\mathcal{C}_n(\mathbf{e}_n)$ is defined in [4, Eq. (10b)]. Given the fact that the residual r is now affected by additive noise, the mismatch between the residual and the true aggregated fault δ can be reduced by increasing the filter horizon n, providing a second direction for noise attenuation. Intuitively, one would opt to increase n to large values. However, much like a moving average filter with a longer horizon, the convergence rate of a fault estimate will decrease proportionally with the horizon n. To achieve a good fault estimation performance, it is crucial to carefully balance two factors: the selection of an appropriate filter $a(\mathfrak{q})$ and the selection of a regression horizon n. Lowering the cutoff frequency of $a(\mathfrak{q})$ reduces the excitation of e to the regression problem. On the other hand, increasing the cut-off frequency allows more noise artifacts to enter the residual, which affects the estimation error. Similarly, reducing the horizon n minimizes the amount of information needed for the regression problem, resulting in less time required to reach the desired fault estimate. However, in a noisy setting, a reduction of n would compromise the quality of the estimation. This gives us a qualitative trade-off for designing a fault estimator in the presence of noise, according to Objective 3

In summary, the steps involved in the robustification process are as follows. First, in Sec. II-D1, we apply Designs 1 and 2 to achieve a robust design in the presence of model uncertainty (Objective 1). Secondly, in Sec. II-D2, we use a methodology to augment the input and measurement delay to the system matrices to compensate for its effect (Objective 2). Finally, in Sec. II-D3, we employ the denominator $a(\mathfrak{q})$ and the isolation filter horizon n to reduce the effect of measurement noise while balancing accuracy and time-based performance (Objective 3).

IV. Experimental results on an automated vehicle

In this section, we will be verifying the contributions made in Sec. III. To begin with, we will train Designs 1 and 2 by using experimental data. We will combine these data with the contributions from Sec. II-D2 and II-D3. After this, we will be using the same filter setup to detect and estimate faults in a real vehicle. The experimental data were collected using a real testing vehicle, which is shown in Fig. 4. The vehicle

Fig. 4: TNO Renault Grand Scenic (2018) testing platform.

TABLE I: Testing matrix for experimental data gathering.

Experiment	Road layout	f_{a} [rad]	f_{m} [-]
Training 1	Straight	0	0
1	Straight	0	0
2	Straight	0.02	0
3	Straight	0	0.3
4	Straight	0.02	0.3
Training 2	Corner	0	0
5	Corner	0	0
6	Corner	0.02	0
7	Corner	0	0.3
8	Corner	0.02	0.3

is a 2018 Renault Grand Scenic that was equipped with a variety of sensors and actuators to control inputs and measure outputs shown in Fig. 2 and (9). To measure lateral velocity v_y, a GNSS sensor was used, which communicated its data to the Axiomtek central computer through a controller area network (CAN) interface. The yaw rate $\dot{\psi}$ was measured by an inertial measurement unit (IMU), which communicated its data through the vehicle gateway to the data logging facility. The lateral error y_e and the heading error ψ_e were derived from the road markings observed by the road marking camera. The longitudinal velocity v_x is measured through the wheel speed sensors and communicated to the central computing unit. The built-in steering actuator, which can be accessed through a CAN interface, was used to actuate the steering angle u of the wheels. Longitudinal acceleration and braking were achieved through a retrofitted system that directly actuated the throttle valve and the position of the brake pedal. The logged data was then communicated to the logging platform. The Axiomtek central computer runs a variety of algorithms using the Robot Operating System (ROS). The platform has several controllers, including a lane-keeping controller based on [36], and an adaptive cruise controller based on [37]. These controllers help the vehicle maintain constant speed and stay in the lane. Additionally, custom software has been developed to inject the desired faults $f_{\mathrm{a}}, f_{\mathrm{m}}$ into the system by manipulating the steering wheel setpoint u. Although in practice faults are mainly caused by mechanical defects, it is considered unsafe to inject these mechanical failures while driving the vehicle. Therefore, software manipulation of the desired steering angle is considered the preferred option for testing purposes.

Experimental data have been collected at the RDW proving

ground in Lelystad, The Netherlands, which features an oval track that has an approximate straight section of 850m and a corner radius of around 160m. To represent typical urban or national road driving, all tests were carried out at a velocity of $50\text{km} \cdot \text{h}^{-1}$, which is equivalent to $13.88\text{m} \cdot \text{s}^{-1}$. According to [38], the validity of the linear bicycle model is guaranteed by constraining the lateral acceleration with $0.5g$, where g represents the gravitational constant. Calculating the lateral acceleration at the velocity of $13.88\text{m} \cdot \text{s}^{-1}$ through the corner shows that on the track we have a maximum lateral acceleration of $a_y = \frac{v_x^2}{R} \approx 1.2\text{m} \cdot \text{s}^{-2}$, which is well within the linear operating regime of the model in (9). The test variations carried out, including different fault scenarios, are shown in Table I. The tests labeled training are intended to design the two residual generator designs as proposed in Sec. III, as well as to gather knowledge to tune the filter parameters of the estimation filter. Using the trained filter and the set of parameters, the experimental results are evaluated. The vehicle parameters are as follows: $m = 1845$ kg, $I = 2372\ \text{kg} \cdot \text{m}^2$, $l_f = 1.219$ m, $l_r = 1.585$ m, $C_f = 138100\ \text{N} \cdot \text{rad}^{-1}$, and $C_r = 215300\ \text{N} \cdot \text{rad}^{-1}$. The actuation and measurement delays have been identified as $\tau_u = 0.15$s, $\tau_{v_y} = 0.06$s, $\tau_{\dot{\psi}} = 0.05$s, $\tau_{y_e} = 0.14$s, $\tau_{\psi_e} = 0.14$s. The uncertainty values, as introduced in (9), are chosen to be in the intervals $w^{(1)} \in [0.8, 1.2]$, $w^{(2)} \in [0.8, 1.2]$, therefore, assuming that the real system can have a 20% deviation in relation to the nominal parameters of the system. These limits are selected as representatives w_j, used in Designs 1, 2.

To strengthen this choice of uncertainty representatives, we performed the analysis of Sec. III-A concerning generalization to unseen scenarios. For each design, three variants of SCP are designed using $4, 40$, and 100 representatives, respectively, sampled from a uniform distribution $\mathbb{P} \sim U(0.8, 1.2)$. Furthermore one variant is designed using the uncertainty limits as representatives. Then, the constraint (30) of the resulting designs is tested on 4000 uncertainties sampled from $\mathbb{P}$. The results are shown in Fig. 5. For the average-cost filter, it is shown that choosing the corner points of the uncertainty set $\mathcal{W}$ results in the lowest violation of the constraint, that is, the lowest false alarm rate of a healthy system. In fact, a larger number of samples incorporated in the design results in an increase in false alarms (that is, $g(\bar{N}, w)$, and therefore a deterioration of performance, as can be observed in the average-cost results in Fig. 5. This shows that a tactical selection of representatives is required. For the worst-case design, it is shown that the selection of corner points greatly outperforms the option with four randomly chosen uncertainty representatives. However, the larger the number of samples, the less constraint violations occur. This shows that although the corner point representatives outperform random selection of the same number of representatives, the corner point representatives do not contain the absolute worst-case.

A. Preliminary simulation study

Section III-A outlines our proposed approach to robustifying against model uncertainty. We perform numerical verification to assess the effectiveness of our approach under uncertain

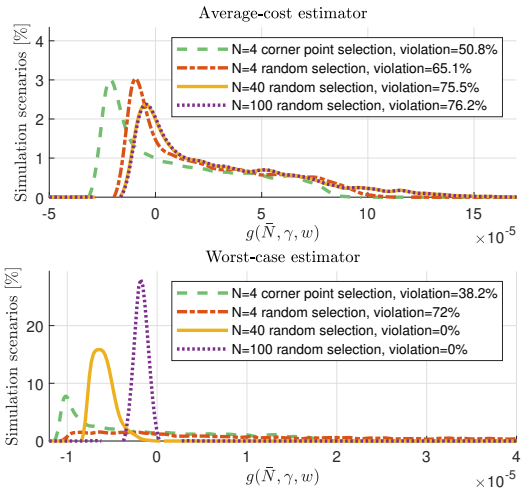


Fig. 5: Comparison of the performance of Designs 1 and 2 with corner point representatives and randomly samples representatives.

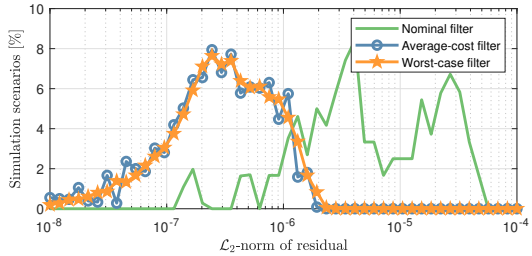


Fig. 6: Aggregated fault estimation error for simulating the synthetic system with residual generator over an equidistant 60×60 grid (sampled within the uncertainty set $\mathcal{W}$) of uncertainties.

conditions while in the absence of delay and noise that occur in the real vehicle. This enables us to demonstrate the performance of our approach in various uncertain scenarios that may differ from the condition of the real vehicle.

In this section, we aim to compare two residual generators - the "worst-case" (28) and "average-cost" (27) viewpoint - with the baseline nominal filter. We only rely on the input data u sent to the vehicle, since the outputs used in the residual generator are generated by a simulation model based on (9). This allows us to test the efficacy of our approach at different levels of uncertainty in vehicle model parameters. The synthetic model is free of measurement noise and delay. Two parameters are still to be designed, which are the filter polynomial $a(\mathfrak{q})$ and the regression horizon n. These parameters are primarily used to reduce noise. Since there is no noise present in this example, the selection will be explained in the next section. In this example, we fix the values of $a(\mathfrak{q})$ and n in $a(\mathfrak{q}) = (\mathfrak{q} - 0.75)^3$ and $n = 500$, respectively, to ensure comparability for the synthetic and experimental results.

Fig. 6 shows the performance of the three proposed residual generators. We tested the average-cost and worst-case residual

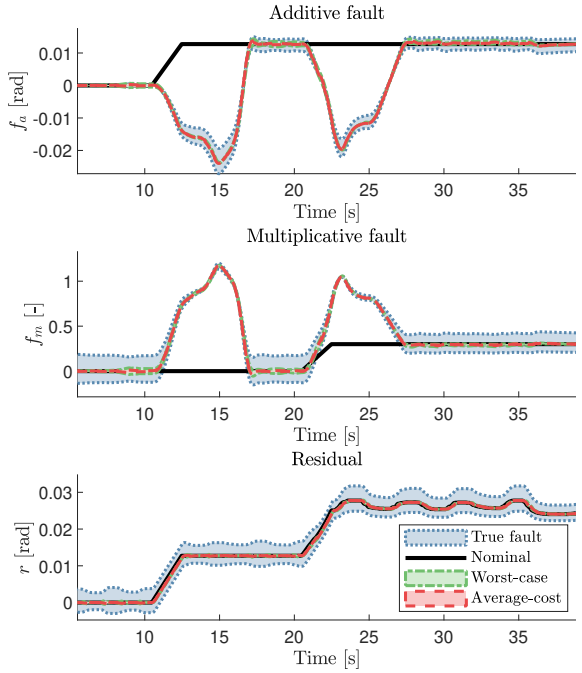


Fig. 7: Fault estimation results for the synthetic system, using the data from experiment 8 as system input, with our proposed estimation filter over a 60×60 grid of uncertainties.

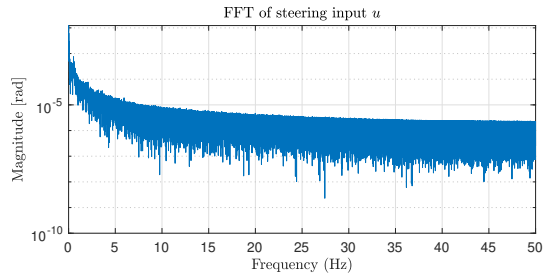
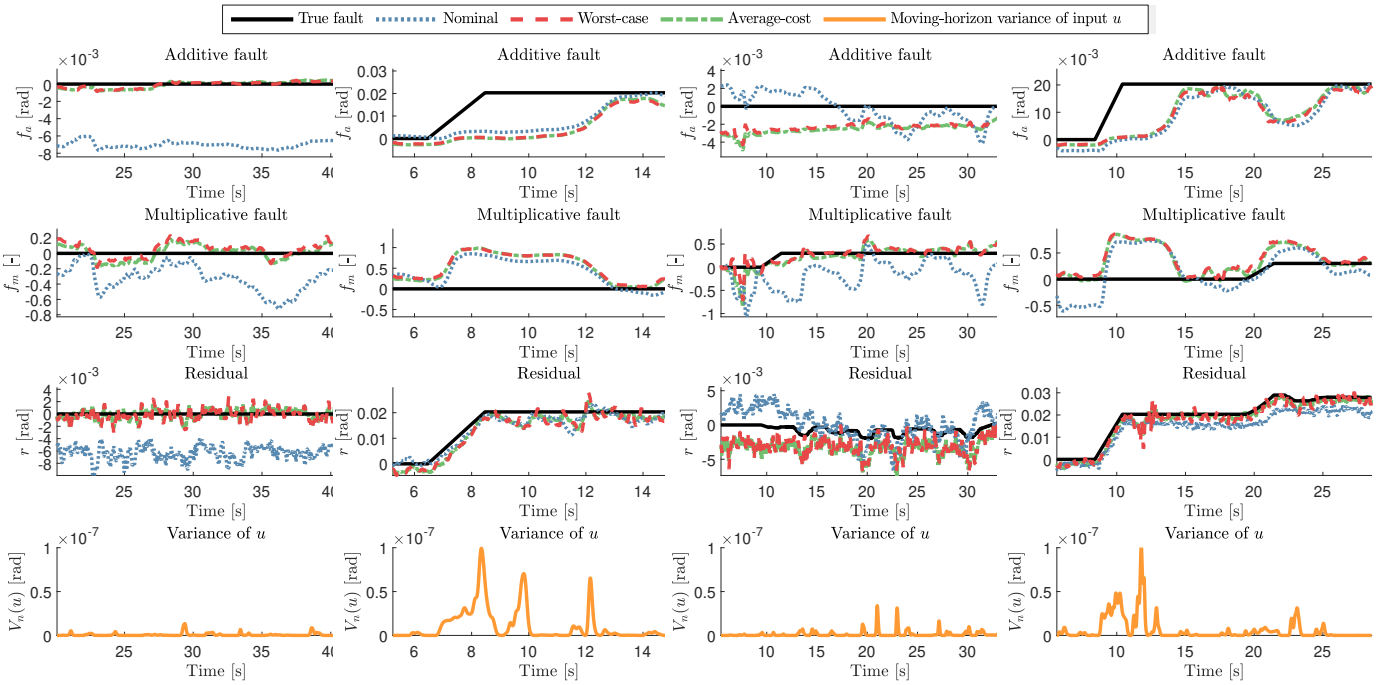


Fig. 8: Fast Fourier transform of input data u from the training experiments.

generator on the training data of Table I and evaluated their performance in the data of experiments 1 and 5, where there was no fault present. We varied the simulation model affected by the uncertainty using a 60×60 grid of uncertainties, bounded by $w^{(1)} \in [0.8, 1.2]$, $w^{(2)} \in [0.8, 1.2]$. The results indicate that both the average-cost and worst-case approaches outperform the nominal filter by several orders of magnitude. However, the average-cost filter performs slightly better than the worst-case filter. This is evident from the average $\mathcal{L}_2$-norm of the residual, which was $4.5 \cdot 10^{-7}$ rad for the average-cost filter, compared to $4.81 \cdot 10^{-7}$ rad for the worst-case filter. Furthermore, the maximum $\mathcal{L}_2$-norm for all experiments was $1.95 \cdot 10^{-6}$ rad for the average-cost filter, while it was $2.37 \cdot 10^{-6}$ rad for the worst-case filter. It is important to note that the worst-case filter is designed by finding a filter that minimizes the worst-case impact at one of the uncertainty representatives. It is unknown whether the representatives of the chosen model have the greatest impact on the performance of the residual generator. This question remains open to research.

Fig. 7 provides a closer look at the filter performance in the presence of faults and incorporates the estimation of faults $\widehat{f}_{\mathrm{a}}$, $\widehat{f}_{\mathrm{m}}$ from the residual r. In this example, we use the data from experiment 8 as input for the synthetic model. Using this input, the synthetic model is again simulated over the same grid in the uncertainty set as used for Fig. 6. The shaded areas in Fig. 7 depict the estimation performance for all filters considered. It has been observed that even in a noise-free and delay-free environment, there are inaccuracies in the nominal design, with errors in $\widehat{f}_{\mathrm{a}}$ of up to $8.9 \cdot 10^{-4}$ rad and in f_{m} of up to 0.18 in steady state. The worst-case filter has maximum errors of $1.9 \cdot 10^{-3}$rad in $\widehat{f}_{\mathrm{a}}$ and 0.02 in $\widehat{f}_{\mathrm{m}}$ in steady state, respectively. On the other hand, the estimation performed using the average-cost residual generator has maximum errors of $8.1 \cdot 10^{-4}$ rad in $\widehat{f}_{\mathrm{a}}$ and 0.015 in $\widehat{f}_{\mathrm{m}}$ in steady state. While the faults f_{a}, f_{m} are transient, the error increases and the difference between the different filter designs decreases. The primary cause of error in this scenario is that we assume that all the information in the regressor (7) is related to a constant fault f_{a}, f_{m}. However, in the case of a transient fault, this assumption is not accurate. After the fault stabilizes and remains constant for $n = 500$ time steps (i.e., the regressor horizon), the fault estimates gradually approach their true values. According to the findings in [4, Theorem 3.7], there is a source of error that could be reduced by any of the following methods: 1) increasing the horizon n to minimize the impact of the transient fault in the regressor, but this would lead to a slower estimation, 2) placing the poles of $a^{-1}(\mathfrak{q})$ toward the origin to reduce the dynamical mismatch between the residual and the regressor, but this would also increase the sensitivity to measurement noise, or 3) increasing the excitation on the steering input u. However, the last method is not within the scope of fault diagnosis in this work.

A note should be made on the time-based performance of the fault estimation. First, it should be noted that the residual converges to the true aggregated fault within approximately 0.15 s (as can be observed in the third column of Fig. 7); therefore, to detect and estimate the presence of a fault in f_{agg}, the residual generator outperforms a human response time of around 0.4s to a hazardous situation [39]. Estimating the faults $\widehat{f}_{\mathrm{a}}$ and $\widehat{f}_{\mathrm{m}}$ individually is a more time-consuming process compared to estimating the combined fault f_{agg} because it necessitates a historical record of the residual and input to distinguish between the faults. As previously discussed in Sec. III-C, selecting a larger horizon n reduces its susceptibility to the influences of measurement noise. In the scope of our experiments, due to the chosen values of n and $a(\mathfrak{q})$, the estimation time of f_{a} and f_{m} is around 5 seconds. Note, however, that the speed of the estimation of $\widehat{f}_{\mathrm{a}}$ and $\widehat{f}_{\mathrm{m}}$ is considered less urgent than the estimation speed of f_{agg}, since high-severity faults in f_{agg} would likely prompt the vehicle to move to a safe state. Consequently, less critical faults or incipient faults could be given some time before determining their precise nature.



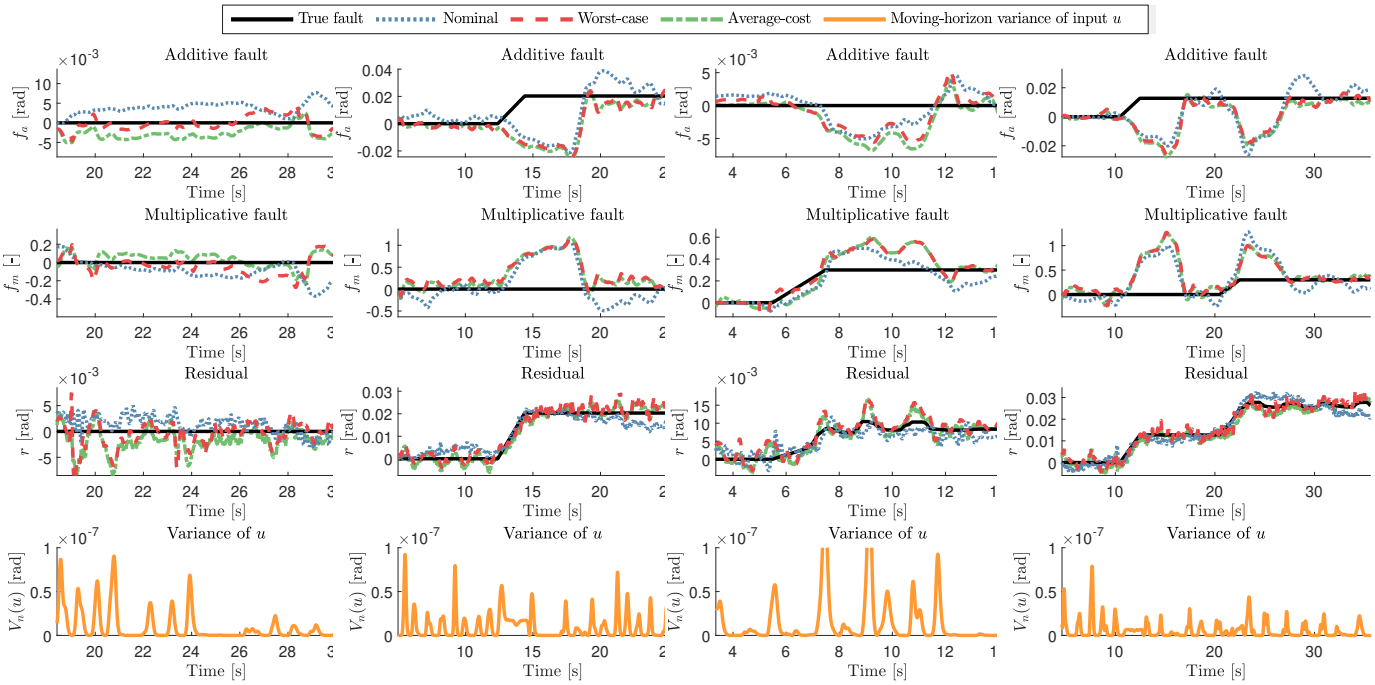
(b) Experiment 1 (c) Experiment 2 (d) Experiment 3 (e) Experiment 4

Fig. 9: Experimental results of experiments 1 to 4, depicting the fault estimation performance for single faults and simultaneous faults on a straight road.

B. Experimental results

In this section, we present the results the experiments from Table I. First, we will discuss the selection of filter parameters $a(\mathfrak{q})$ and n. We aim to find an appropriate filter $a(\mathfrak{q})$ that not only reduces noise, but also allows us to use the input signal u effectively. To analyze the frequency content of the input signal u, we perform a fast Fourier transform (FFT) of all the augmented training data, illustrated in Fig. 8. Upon observation, it is clear that around 30Hz, the roll-off of the magnitude stagnates, indicating that the actual steering dynamics diminishes at this frequency. This is the frequency range that contains the flat spectrum of noise in the signal. As such, to create a denominator that filters noise while preserving the frequency content of u, a low-pass filter with a cutoff frequency at 30Hz is selected. Combining this with the signal sampling time of 0.01s results in $a(\mathfrak{q}) = (\mathfrak{q} - e^{-0.01\cdot 30})^{d_a}$, where the degree $d_a = 3$ is chosen so that the residual generators resulting are causal. Due to the large effect of excitation on the performance bound (31), it must be preserved and not sacrificed by attenuating more noise through $a(\mathfrak{q})$. Hence, the horizon n can be used to attenuate noise from the relatively soft low-pass filter $a(\mathfrak{q})$, as well as to attenuate the coupling effect between the estimation of $\widehat{f}_{\mathrm{a}}$, $\widehat{f}_{\mathrm{m}}$, as was also observed in Sec. IV-A. The filter horizon is increased to $n = 500$ to attenuate the noise and coupling effects to a satisfactory level. As mentioned in the previous section, allowing around 5 s for the faults $\widehat{f}_{\mathrm{a}}, \widehat{f}_{\mathrm{m}}$ is acceptable as long as the residual has a satisfactory convergence time, which in all experiments is maximally around 0.4 s. Fig. 9 and 10 show the experimental results from Table I.

First, Fig. 9 shows the results of the experiments on the straight road. In these results, an additional graph has been added to show the moving-horizon variance of the signal u, i.e., $V_n[u]$, with $n = 500$, which indicates excitation in these experiments. Note that this quantity is an unfiltered version of $V_n[e]$ from (31), which means that its value over time will allow us to reason about the expected quality of the estimation. The main source of excitation is the injection of an additive fault in Fig. 9c and Fig. 9e. For all three variants of filters, as also explained in the preliminary simulation study, there is a coupling effect between $\widehat{f}_{\mathrm{a}}$ and $\widehat{f}_{\mathrm{m}}$. This effect is inevitable given the static relationship from which these faults are extracted. However, despite the lack of excitation and this coupling, the fault estimates do converge to their true constant values when employing either the average-cost or the worst-case filter. However, the nominal filter fails mainly to accurately estimate $\widehat{f}_{\mathrm{m}}$, which can be caused by the dynamic mismatch of the true behavior of the vehicle compared to the identified vehicle model. In general, the average cost and worst-case design are comparable and mostly identical in terms of performance, as had already been shown through simulation in Sec. IV-A. The results of the experiments while cornering (Fig. 10) show higher levels of excitation in the steering input u. Namely, in these scenarios, the vehicle must constantly regulate itself while taking the corner, which causes relatively higher excitation. As a result, the convergence of the fault estimates is more precise. The proposed average-cost and worst-case filter designs outperform the nominal design in terms of steady-state error in the faults and residual. However, during transient fault periods, all filter designs perform simi-



(b) Experiment 5 (c) Experiment 6 (d) Experiment 7 (e) Experiment 8

Fig. 10: Experimental results of experiments 5 to 8, depicting the fault estimation performance for single faults and simultaneous faults on a curved road.

larly in estimating $\widehat{f}_{\mathrm{a}}$, $\widehat{f}_{\mathrm{m}}$. The nominal design performs worse in estimating the additive fault $\widehat{f}_{\mathrm{a}}$ with respect to the results of the straight road. A mismatch in $\widehat{f}_{\mathrm{m}}$ implies that the residual r contains traces that are not part of the aggregated fault, but are still correlated with the input signal u. Therefore, a mismatch in f_{a} could be better explained by an uncorrelated or less correlated trace in r that still affects the residual. An example of such an uncorrelated, or less correlated, signal could be the curvature disturbance, which propagates through the residual generator through mismatch term $\mathbf{I}$ (12), in the nominal case, and less so in the average cost and worst-case design.

When using results for diagnostic purposes, it is crucial to consider their accuracy and reliability. The estimated faults can help detect a specific severity of faults using a set threshold or mitigate them through closed-loop control. However, fault estimates observed using the nominal filter on a straight road with a constant multiplicative fault, or in the corner with an additive and/or multiplicative fault, may not be accurate. This could lead to false positive detections or overcompensation in closed-loop mitigation. Therefore, it is important to be mindful when interpreting the results. Moreover, the inter-dependence between the estimation of the two faults $\widehat{f}_{\mathrm{a}}$, $\widehat{f}_{\mathrm{m}}$ can be problematic for all proposed filters, and depending on the application, different tuning parameters may be used to attenuate this phenomenon.

V. Conclusion

In this work, we focus on the estimation of additive and multiplicative faults that can cause errors in the steering system in the context of automated driving. In this experimental setting, several factors can introduce errors in fault estimation, such as model uncertainty, measurement noise, and system delays. We have proposed methodologies to mitigate these factors and improve the precision and accuracy of the fault estimation. Our approach has been tested using simulations and experiments in the field of automated driving, and we have discussed its effectiveness and limitations. The results show that incorporating the average-cost or worst-case fault estimation filter, compared to the baseline filter, improves the accuracy and precision of individual fault estimates. Future work includes the implementation of this methodology in closed-loop applications and exploring the possibility of active fault isolation within the automotive domain by introducing excitation to obtain more precise estimates of faults.

References

[1] D. Milakis, B. van Arem, and B. van Wee, "Policy and society related implications of automated driving: A review of literature and directions for future research," *J. of Int. Transp. Syst.*, vol. 21, no. 4, pp. 324–348, 2017.

[2] P. Koopman, U. Ferrell, F. Fratrik, and M. Wagner, "A Safety Standard Approach for Fully Autonomous Vehicles," in *Computer Safety, Reliability, and Security*, 2019, pp. 326–332.

[3] *ISO 26262-1 Road vehicles - Functional safety*, 2018.

[4] C. van der Ploeg, M. Alirezaei, N. van de Wouw, and P. M. Esfahani, "Multiple Faults Estimation in Dynamical Systems: Tractable Design and Performance Bounds," *IEEE Trans. on Autom. Control*, vol. 67, no. 9, pp. 4916–4923, 2022.

[5] Z. Gao, C. Cecati, and S. X. Ding, "A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part II: Fault Diagnosis With Knowledge-Based and Hybrid/Active Approaches," *IEEE Trans. on Ind. Electron.*, vol. 62, no. 6, pp. 3768–3774, 2015.

[6] R. Isermann, "Model-based fault-detection and diagnosis–status and applications," *Annual Reviews in control*, vol. 29, no. 1, pp. 71–85, 2005.

[7] R. Beard, *Failure Accommodation in Linear Systems Through Self-reorganization* (NASA CR). M.I.T. Man-Vehicle Laboratory, 1971.

[8] F. Boem, R. Ferrari, C. Keliris, T. Parisini, and M. Polycarpou, "A Distributed Networked Approach for Fault Detection of Large-Scale Systems," *IEEE Trans. Autom. Control*, vol. 62, no. 1, pp. 18–33, 2017.

[9] H. Fang, H. Ye, and M. Zhong, "Fault diagnosis of networked control systems," *Annual Reviews in Control*, vol. 31, no. 1, pp. 55–68, 2007.

[10] B. Svetozarevic, P. M. Esfahani, M. Kamgarpour, and J. Lygeros, "A robust fault detection and isolation filter for a horizontal axis variable speed wind turbine," in *American Control Conf.*, 2013, pp. 4453–4458.

[11] Z. Gao, T. Breikin, and H. Wang, "Discrete-time proportional-integral observer and observer-based controller for systems with unknown disturbances," in *Europ. Control Conf.*, 2007, pp. 5248–5253.

[12] M. Gholizadeh and F. R. Salmasi, "Estimation of State of Charge, Unknown Nonlinearities, and State of Health of a Lithium-Ion Battery Based on a Comprehensive Unobservable Model," *IEEE Trans. on Ind. Electron.*, vol. 61, no. 3, pp. 1335–1344, 2014.

[13] Z. Gao, X. Liu, and M. Chen, "Unknown input observer-based robust fault estimation for systems corrupted by partially decoupled disturbances," *IEEE Trans. on Ind. Electron.*, vol. 63, no. 4, pp. 2537–2547, 2016.

[14] T. Höfling and R. Isermann, "Fault detection based on adaptive parity equations and single-parameter tracking," *Control Eng. Practice*, vol. 4, no. 10, pp. 1361–1369, 1996.

[15] H. Zhang and J. Wang, "Active Steering Actuator Fault Detection for an Automatically-Steered Electric Ground Vehicle," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 3685–3702, 2017.

[16] C. van der Ploeg, E. Silvas, N. van de Wouw, and P. M. Esfahani, "Real-Time Fault Estimation for a Class of Discrete-Time Linear Parameter-Varying Systems," *IEEE Control Syst. Letters*, vol. 6, pp. 1988–1993, 2022.

[17] C. Huang, F. Naghdy, and H. Du, "Delta Operator-Based Fault Estimation and Fault-Tolerant Model Predictive Control for Steer-By-Wire Systems," *IEEE Trans. on Control Syst. Tech.*, vol. 26, no. 5, pp. 1810–1817, 2018.

[18] J. S. Im, F. Ozaki, T. K. Yeu, and S. Kawaji, "Model-based fault detection and isolation in steer-by-wire vehicle using sliding mode observer," *J. of Mech. Science and Technology*, vol. 23, no. 8, pp. 1991–1999, 2009.

[19] S. A. Arogeti, D. Wang, C. B. Low, and M. Yu, "Fault Detection Isolation and Estimation in a Vehicle Steering System," *IEEE Trans. on Ind. Electron.*, vol. 59, no. 12, pp. 4810–4820, 2012.

[20] S. Skogestad and I. Postlethwaite, *Multivariable Feedback Control: Analysis and Design*. John Wiley & Sons, Ltd, 2005.

[21] X.-G. Yan and C. Edwards, "Adaptive Sliding-Mode-Observer-Based Fault Reconstruction for Nonlinear Systems With Parametric Uncertainties," *IEEE Trans. on Ind. Electron.*, vol. 55, no. 11, pp. 4029–4036, 2008.

[22] R. Li and Y. Yang, "Sliding-Mode Observer-Based Fault Reconstruction for T-S Fuzzy Descriptor Systems," *IEEE Trans. on Syst., Man, and Cybern.: Syst.*, vol. 51, no. 8, pp. 5046–5055, 2021.

[23] R. Xiong, Q. Yu, W. Shen, C. Lin, and F. Sun, "A Sensor Fault Diagnosis Method for a Lithium-Ion Battery Pack in Electric Vehicles," *IEEE Trans. on Power Electron.*, vol. 34, no. 10, pp. 9709–9718, 2019.

[24] H. He, R. Xiong, X. Zhang, F. Sun, and J. Fan, "State-of-Charge Estimation of the Lithium-Ion Battery Using an Adaptive Extended Kalman Filter Based on an Improved Thevenin Model," *IEEE Trans. Veh. Technol.*, vol. 60, no. 4, pp. 1461–1469, 2011.

[25] H. Chen, L. Li, C. Shang, and B. Huang, "Fault Detection for Nonlinear Dynamic Systems With Consideration of Modeling Errors: A Data-Driven Approach," *IEEE Trans. on Cybern.*, vol. 53, no. 7, pp. 4259–4269, 2023.

[26] Z. Gao, C. Cecati, and S. Ding, "A survey of fault diagnosis and fault-tolerant techniques-part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. on Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, 2015.

[27] K. Pan, P. Palensky, and P. M. Esfahani, "Dynamic Anomaly Detection With High-Fidelity Simulators: A Convex Optimization Approach," *IEEE Trans. on Smart Grid*, vol. 13, no. 2, pp. 1500–1515, 2022.

[28] P. Mohajerin Esfahani and J. Lygeros, "A tractable fault detection and isolation approach for nonlinear systems with probabilistic performance," *IEEE Trans. on Autom. Control*, vol. 61, no. 3, pp. 633–647, 2016.

[29] V. Reppa, M. M. Polycarpou, and C. G. Panayiotou, "Adaptive Approximation for Multiple Sensor Fault Detection and Isolation of Nonlinear Uncertain Systems," *IEEE Trans. on Neural Networks and Learning Syst.*, vol. 25, no. 1, pp. 137–153, 2014.

[30] L. Chen, S. Fu, Y. Zhao, M. Liu, and J. Qiu, "State and Fault Observer Design for Switched Systems via an Adaptive Fuzzy Approach," *IEEE Trans. on Fuzzy Syst.*, vol. 28, no. 9, pp. 2107–2118, 2020.

[31] N. Mabrouk, A. Ben Brahim, and F. Ben Hmida, "Simultaneous Multiplicative and Additive Actuator Faults Estimation-Based Sliding Mode FTC for a Class of Uncertain Nonlinear System," en, *Mathematical Problems in Eng.*, vol. 2023, no. 1, p. 6 902 272, 2023.

[32] A. H. Tahoun, "Time-varying multiplicative/additive faults compensation in both actuators and sensors simul-

taneously for nonlinear systems via robust sliding mode control scheme," *J. of the Franklin Institute*, vol. 356, no. 1, pp. 103–128, 2019.

[33] P. Mohajerin Esfahani, T. Sutter, and J. Lygeros, "Performance Bounds for the Scenario Approach and an Extension to a Class of Non-Convex Programs," *IEEE Trans. on Autom. Control*, vol. 60, no. 1, pp. 46–58, 2015.

[34] M. C. Campi and S. Garatti, "The Exact Feasibility of Randomized Solutions of Uncertain Convex Programs," *SIAM J. on Optimization*, vol. 19, no. 3, pp. 1211–1230, 2008.

[35] T. Teräsvirta, "Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models," *J. of the American Statistical Association*, 1994.

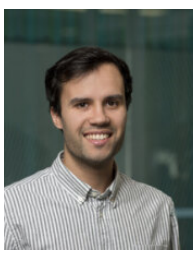
[36] A. Schmeitz, J. Zegers, J. Ploeg, and M. Alirezaei, "Towards a generic lateral control concept for cooperative automated driving theoretical and experimental evaluation," in *IEEE Int. Conf. on Models and Technologies for Intell. Transp. Syst.*, 2017, pp. 134–139.

[37] J. Ploeg, B. T. M. Scheepers, E. van Nunen, N. van de Wouw, and H. Nijmeijer, "Design and experimental evaluation of cooperative adaptive cruise control," in *IEEE Conf. on Int. Transp. Syst.*, 2011, pp. 260–265.

[38] J.-M. Park, D.-W. Kim, Y.-S. Yoon, H. J. Kim, and K.-S. Yi, "Obstacle avoidance of autonomous vehicles based on model predictive control," *Proc. of the Inst. of Mech. Engineers, Part D: J. of Automobile Eng.*, vol. 223, no. 12, pp. 1499–1516, 2009.

[39] B. Wolfe, B. Seppelt, B. Mehler, B. Reimer, and R. Rosenholtz, "Rapid holistic perception and evasion of road hazards," *J. Exp. Psychol.*, vol. 149, no. 3, pp. 490–500, 2020.

Chris van der Ploeg received the B.Sc.-degree in mechanical engineering and the M.Sc.-degree (*cum laude*) in systems and control from the Delft University of Technology, Delft, The Netherlands in 2016 and 2018, respectively. He received the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands in 2024. Since 2019, he has been a Research Scientist with the Integrated Vehicle Safety Department, Netherlands Organisation for Applied Scientific Research (TNO), Helmond, The Netherlands. His current research interests include fault diagnosis methods, fault mitigation strategies/methods, and risk-averse motion planning for connected and cooperative vehicles.



Pedro Vieira Oliveira received the B.Sc. in mechanical engineering at the University of Coimbra and M.Sc. degree in automotive technology at the Eindhoven University of Technology, in 2019 and 2022, respectively. Since 2022, he has been a Research Scientist with the Integrated Vehicle Safety Department, Netherlands Organisation for Applied Scientific Research (TNO), Helmond, The Netherlands.

Emilia Silvas received the B.Sc. degree in automatic control and computer science from the Politehnica University of Bucharest, Romania, in 2009, and the M.Sc. degree in systems and control and the Ph.D. degree from the Eindhoven University of Technology, Eindhoven, The Netherlands, in 2011 and 2015, respectively. Since 2016, she has been a Research Scientist with the Netherlands Organization for Applied Scientific Research (TNO), Helmond, The Netherlands, where she is working on the areas of cooperative vehicle systems and mobile robots and being responsible for the smart vehicles programs cluster. Her research interests include advanced control, system identification and modeling, machine learning techniques, and optimal system design. From 2020 to 2023 she has been the Chair of the Mobility, Transport and Logistics Working Group, Dutch AI Coalition.

Peyman Mohajerin Esfahani received the B.Sc. and M.Sc. degrees from Sharif University of Technology, Iran, and the Ph.D. degree from ETH Zurich, Switzerland. He is currently an associate professor in the Delft Center for Systems and Control at the Delft University of Technology, The Netherlands. Prior to joining TU Delft, he held several research appointments at EPFL, ETH Zurich, and MIT between 2014 and 2016. His research interests include theoretical and practical aspects of decision-making problems in uncertain and dynamic environments, with applications to control and security of large-scale and distributed systems.

He currently serves as an associate editor of Operations Research, Transactions on Automatic Control, and Open Journal of Mathematical Optimization. He was one of the three finalists for the Young Researcher Prize in Continuous Optimization awarded by the Mathematical Optimization Society in 2016, and a recipient of the 2016 George S. Axelby Outstanding Paper Award from the IEEE Control Systems Society. He received the ERC Starting Grant and the INFORMS Frederick W. Lanchester Prize in 2020. He is the recipient of the 2022 European Control Award.

Nathan van de Wouw obtained his M.Sc.-degree (with honours) and Ph.D.-degree in Mechanical Engineering from the Eindhoven University of Technology, the Netherlands, in 1994 and 1999, respectively. He currently holds a full professor position at the Mechanical Engineering Department of the Eindhoven University of Technology, the Netherlands. He has been working at Philips Applied Technologies, The Netherlands, in 2000 and at the Netherlands Organisation for Applied Scientific Research, The Netherlands, in 2001. He has been a visiting professor at the University of California Santa Barbara, U.S.A., in 2006/2007, at the University of Melbourne, Australia, in 2009/2010 and at the University of Minnesota, U.S.A., in 2012 and 2013. He has held a (part-time) full professor position the Delft University of Technology, the Netherlands, from 2015-2019. He has also held an adjunct full professor position at the University of Minnesota, U.S.A, from 2014-2021. He has published the books 'Uniform Output Regulation of Nonlinear Systems: A convergent Dynamics Approach' with A.V. Pavlov and H. Nijmeijer (Birkhauser, 2005) and 'Stability and Convergence of Mechanical Systems with Unilateral Constraints' with R.I. Leine (Springer-Verlag, 2008). In 2015, he received the IEEE Control Systems Technology Award "For the development and application of variable-gain control techniques for high-performance motion systems". He is an IEEE Fellow for his contributions to hybrid, data-based and networked control.