Performance bounds on change detection with application to manoeuvre recognition for advanced driver assistance systems

Jan Erik Stellet[1], Jan Schumacher[1], Wolfgang Branz[1], J. Marius Zöllner[2]

Abstract— **Recognising the intended manoeuvres of other traffic participants is a crucial task for situation interpretation in driver assistance and autonomous driving. While many works propose algorithms for (computationally feasible) inference, much less attention is paid to finding analytic upper performance bounds for these problems.**

This work studies the statistical properties of the optimal detector in a binary change detection problem, i.e. the Generalised Likelihood Ratio test. With analytic models of the best attainable receiver operating characteristic, the influence of system design parameters can be investigated without the need for empirical evaluation. Moreover, these bounds can be used to derive objective performance metrics.

I. Introduction

A. Motivation

Advanced Driver assistance systems (ADAS) strive to support the driver with interventions in the vehicle controls (e.g. performing an automatic emergency brake before an imminent collision). Hence, these systems utilise predictions in order to be able to assess the future evolution of a traffic situation [1], [2]. To this end, kinematic motion models are used where the initial values of the state variables are estimated by a tracking filter.

However, vehicle trajectories result from driver intentions on a semantic level, i.e. performing a specific manoeuvre to safely reach a certain navigation goal. A viable approach to achieve accurate predictions for longer time spans is thus to recognise the discrete driver intention first and predict trajectories that correspond to the relevant manoeuvre only [3]. This concept is visualised in Fig. 1. Methods for driving situation estimation include e.g., algorithms for change detection [4], multiple model Kalman filter [5], and probabilistic inference in Bayesian Networks [6], Hidden Markov Models [7] or Dynamic Bayesian Networks [8].

In order to quantify the overall uncertainty in the predicted trajectory, one needs to model inaccuracies in kinematic motion models as well as uncertainty in the intention detection, i.e the receiver operating characteristic (ROC). Often, only upper bounds on the attainable performance can be analytically calculated. These can be useful as an absolute reference in order to objectively compare the performance of different algorithms. While the Cramér-Rao lower bound (CRLB) [9], [10] provides a lower bound on the error covariance for the estimation of continuous state variables, no similarly simple analytic result is available for the recognition of discrete intentions.

To study this problem, it is formulated as the detection of changes in a discrete-time linear non-Gaussian system [4]. Here, the Generalised Likelihood Ratio (GLR) test is known to be optimal in special cases [11] and asymptotically optimal in general [12]. The statistics of this test therefore give an idea of a general upper performance bound.

B. Background and previous results

A comprehensive overview of change detection in dynamic systems is given in [11]. The focus of this work is on additive changes in linear systems. One approach suitable in the presence of Gaussian process and measurement noise is to form residuals from the Kalman filter innovations [13]. This recursive algorithm is a GLR test and the test statistics follow from the Kalman filter covariance propagation.

However, many real applications are characterised by non-Gaussian errors. For example, [14] proposes a bimodal Gaussian distribution to model an automotive radar sensor. In these cases, a closed-form of the optimal detector usually does not exist.

In order to find an upper bound on the attainable performance, the test statistics of a GLR test over a sliding window of data is considered in [15], [16]. The two relevant factors which influence the detector performance are the length L of the data window and the noise parameters.

It is of interest to determine the minimum window length in order to achieve a desired probability of detection P_{D} at a maximum tolerable false alarm probability P_{FA}. Moreover, the potential improvement that can be expected from a detector which is based on the full noise information instead of a Gaussian approximation is to be analysed. This questions can be addressed using the notion of *intrinsic accuracy* (IA) [17]. The main advantage of a Gaussian approximation on the other hand is that closed-form solutions are available and thus computationally challenging methods, e.g. the particle filter, can be avoided.

C. Organisation of the paper

At first, preliminary background on the models considered in this work, estimation theory and the GLR test are reviewed in Sec. II. Subsequently, the GLR test statistic is derived. Previous results in block matrix notation are first introduced in Sec. III-A and then expanded on with the novel recursive model in Sec. III-B which is further discussed in Sec. III-C. These theoretical findings are applied to a simulation example in Sec. IV. The paper concludes with a summary in Sec. V.

[1]Jan Erik Stellet, Jan Schumacher and Wolfgang Branz are with Robert Bosch GmbH, Corporate Research, Vehicle Safety and Assistance Systems, 71272 Renningen, Germany

[2]J. Marius Zöllner is with Research Center for Information Technology (FZI), 76131 Karlsruhe, Germany

©2015 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works. DOI 10.1109/IVS.2015.7225833

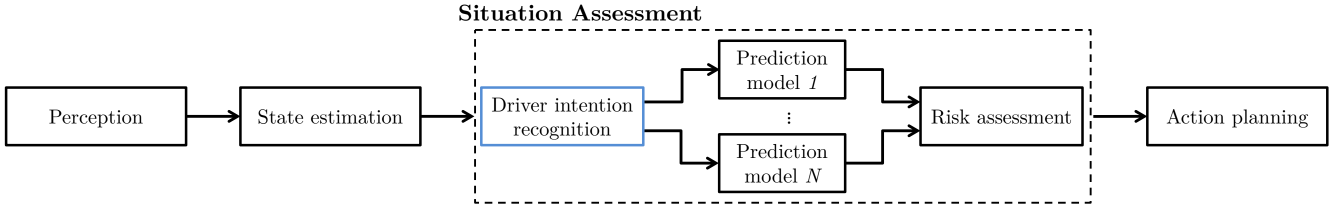


Fig. 1. Exemplary signal processing chain of a driver assistance system. From the perceptions of the environment sensor, the dynamic state as well as intentions of other traffic participants are estimated. Prediction, risk assessment and planning of future actions are then based on dynamic prediction models which correspond to the detected intentions.

II. Preliminaries

A. Model representations

Linear dynamic systems in state space form are considered in this work. The system state is denoted $\mathbf{x}_k \in \mathbb{R}^n$, $\mathbf{y}_k \in \mathbb{R}^m$ is the measurement output, $\mathbf{u}_k$ a deterministic known input and $\mathbf{f}_k$ a deterministic, but unknown additive input (fault)[1]:

$$\mathbf{x}_{k+1} = \mathbf{A}_k\mathbf{x}_k + \mathbf{B}_k^{\mathbf{u}}\mathbf{u}_k + \mathbf{B}_k^{\mathbf{f}}\mathbf{f}_k + \mathbf{B}_k^{\mathbf{w}}\mathbf{w}_k \tag{1a}$$

$$\mathbf{y}_k = \mathbf{C}_k\mathbf{x}_k + \mathbf{D}_k^{\mathbf{u}}\mathbf{u}_k + \mathbf{D}_k^{\mathbf{f}}\mathbf{f}_k + \mathbf{v}_k \; . \tag{1b}$$

The stochastic inputs $\mathbf{w}_k$ and $\mathbf{v}_k$ denote white, independent process and measurement noise, respectively.

In order to make the detection of unconstant $\mathbf{f}_k$ feasible, it is assumed that the time-dependence can be modelled as a linear parametrisation. The fault profile is thus given by known, time-dependent basis functions $\mathbf{F}_k$ and a ν-dimensional, time-invariant coefficient vector $\boldsymbol{\theta} \in \mathbb{R}^\nu$ as $\mathbf{f}_k = \mathbf{F}_k\boldsymbol{\theta}$ [15]. How to describe a fault signal in this way will be illustrated for an application example in Sec. IV. Hence, $\mathbf{B}_k^{\mathbf{f}}$ and $\mathbf{D}_k^{\mathbf{f}}$ in (1) are replaced by $\mathbf{B}_k^{\boldsymbol{\theta}} = \mathbf{B}_k^{\mathbf{f}}\mathbf{F}_k$ and $\mathbf{D}_k^{\boldsymbol{\theta}} = \mathbf{D}_k^{\mathbf{f}}\mathbf{F}_k$.

For notational simplicity, the known input $\mathbf{u}_k$ is neglected in the following. Moreover, a time-invariant system model is assumed and therefore time-dependence of system matrices is only taken into account for the fault profile $\mathbf{F}_k$.

Multiple measurements are integrated in the change detection approach. Given an initial state $\mathbf{x}_{k-L+1}$, the stacked measurements $\mathbf{Y}_L = \begin{bmatrix}\mathbf{y}_{k-L+1}^\top & \dots & \mathbf{y}_k^\top\end{bmatrix}^\top$ read [11]:

$$\mathbf{Y}_L = \boldsymbol{\mathcal{O}}_L\mathbf{x}_{k-L+1} + \mathbf{H}_L^{\boldsymbol{\theta}}\boldsymbol{\Theta}_L + \mathbf{H}_L^{\mathbf{w}}\mathbf{W}_L + \mathbf{V}_L \; . \tag{2}$$

In this notation, $\boldsymbol{\mathcal{O}}_L$ is the extended observability matrix

$$\boldsymbol{\mathcal{O}}_L = \begin{bmatrix}\mathbf{C}^\top & (\mathbf{C}\mathbf{A})^\top & \dots & (\mathbf{C}\mathbf{A}^{L-1})^\top\end{bmatrix}^\top , \tag{3}$$

and the influence of the inputs $i \in \{\boldsymbol{\theta}, \mathbf{w}\}$ is given by

$$\mathbf{H}_L^i = \begin{bmatrix} \mathbf{D}_{k-L+1}^i & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{C}\mathbf{B}_{k-L+1}^i & \mathbf{D}_{k-L+2}^i & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{C}\mathbf{A}^{L-2}\mathbf{B}_{k-L+1}^i & \mathbf{C}\mathbf{A}^{L-3}\mathbf{B}_{k-L+2}^i & \dots & \mathbf{D}_k^i \end{bmatrix} . \tag{4}$$

[1]For consistency with the notation in [15], $\mathbf{f}_k$ appears in both system (1a) and measurement (1b) equation. However, for the considered application of detecting changes in a target object's manoeuvre, the observation is not directly affected and hence $\mathbf{D}_k = \mathbf{0}$.

The stacked vectors $\boldsymbol{\Theta}_L$, $\mathbf{W}_L$ and $\mathbf{V}_L$ concatenate the respective signals $\boldsymbol{\theta}$, $\mathbf{w}$ and $\mathbf{v}$ from times $k-L+1, \dots, k$.

B. Information and Accuracy

Two concepts known from estimation theory will be briefly reviewed in the following. Firstly, the *Fisher information* matrix $\boldsymbol{\mathcal{I}}_{\mathbf{y}}(\boldsymbol{\theta})$ is used to quantify the information on a parameter $\boldsymbol{\theta}$ contained in the data $\mathbf{y}$. It is defined as

$$\boldsymbol{\mathcal{I}}_{\mathbf{y}}(\boldsymbol{\theta}) = -\mathbb{E}\left[\nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}^\top \ln p(\mathbf{y}\,|\boldsymbol{\theta})\right] \tag{5}$$

where the probability density of the data conditional on the parameter is denoted as $p(\mathbf{y}\,|\boldsymbol{\theta})$ [12]. One important consequence is that the covariance of any unbiased estimate $\hat{\boldsymbol{\theta}}$ calculated from $\mathbf{y}$ is bounded below by the inverse Fisher information matrix. This relation is known as the Cramér-Rao lower bound (CRLB):[2]

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \operatorname{cov}\left(\hat{\boldsymbol{\theta}}\right) \succeq \boldsymbol{\mathcal{I}}_{\mathbf{y}}^{-1}(\boldsymbol{\theta}) \; . \tag{6}$$

An estimator which reaches the CRLB is denoted *efficient*. If an efficient unbiased estimator exists for a given problem, this is the Maximum Likelihood estimator (MLE) [12]. Moreover, in a Gaussian setup, the MLE is a linear function of the data and a closed form exists. The best linear unbiased estimator (BLUE) and the MLE are identical then.

Secondly, the *intrinsic accuracy* (IA) characterises the probability density $p(\mathbf{e})$ of a zero mean noise process $\mathbf{e}$ [17]:

$$\boldsymbol{\mathcal{I}}_{\mathbf{e}} = -\mathbb{E}\left[\nabla_{\mathbf{e}}\nabla_{\mathbf{e}}^\top \ln p(\mathbf{e})\right] \; . \tag{7}$$

Here, we have the inequality

$$\boldsymbol{\Sigma}_{\mathbf{e}} = \operatorname{cov}(\mathbf{e}) \succeq \boldsymbol{\mathcal{I}}_{\mathbf{e}}^{-1} \tag{8}$$

which can be interpreted as the information about $\mathbf{e}$ contained in its covariance. It is easily verified that for Gaussian noise $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{e}})$, (7) holds with equality. Thus, approximating $\mathbf{e}$ with a Gaussian of the same covariance is the least informative choice.

The intuition here is that for a given estimation problem in a non-Gaussian context, reaching the CRLB is only possible with an MLE. However, there is potentially no closed form solution and numerically expensive methods are required. Employing a suboptimal BLUE for an approximately equivalent Gaussian problem might therefore be a viable option.

[2]The notation $\mathbf{A} \succeq \mathbf{B}$ for matrices $\mathbf{A}, \mathbf{B}$ denotes that the difference $\mathbf{A} - \mathbf{B}$ is positive semidefinite.

C. Generalised Likelihood Ratio test

The objective is to optimally detect the presence of a fault $\mathbf{f}_k$, parametrised by $\boldsymbol{\theta}$ in the system (1) given L measurements $\mathbf{Y}_L$. We assume that $\boldsymbol{\theta}_0 = \mathbf{0}$ corresponds to the nominal system behaviour (hypothesis $\mathcal{H}_0$) and $\boldsymbol{\theta}_1 \neq \mathbf{0}$ under $\mathcal{H}_1$ is to be detected.[3] This is a composite hypothesis testing problem because the precise value of $\boldsymbol{\theta}_1$ is not known beforehand. In contrast, testing between two possible $\boldsymbol{\theta}$ (simple hypotheses) has a different solution and test statistics.

An intuitive approach is to estimate the change parameter $\boldsymbol{\theta}$ from measured data and compare it to $\mathbf{0}$, which is known as the Wald test. Here, $L \gg \nu$ measurements $\mathbf{Y}_L$ are collected and (2) is rewritten as a linear regression problem:

$$\mathbf{Y}_L - \boldsymbol{\mathcal{O}}_L \mathbf{x}_{k-L+1} = \mathbf{H}_L^{\boldsymbol{\theta}} \boldsymbol{\Theta}_L + \mathbf{H}_L^{\mathbf{w}} \mathbf{W}_L + \mathbf{V}_L \ . \tag{9}$$

However, assuming precise knowledge of the initial value $\mathbf{x}_{k-L+1}$ is unrealistic in practice.[4] Thus, the estimate is written as $\hat{\mathbf{x}}_{k-L+1} = \mathbf{x}_{k-L+1} + \tilde{\mathbf{x}}_{k-L+1}$ where the intrinsic accuracy of the zero mean estimation error $\tilde{\mathbf{x}}_{k-L+1}$ is denoted as $\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{x}}_{k-L+1}}$. Then:

$$\underbrace{\mathbf{Y}_L - \boldsymbol{\mathcal{O}}_L \hat{\mathbf{x}}_{k-L+1}}_{=:\mathbf{R}_L} = \boldsymbol{\Phi}_L \boldsymbol{\theta} + \underbrace{\boldsymbol{\mathcal{O}}_L \tilde{\mathbf{x}}_{k-L+1} + \mathbf{H}_L^{\mathbf{w}} \mathbf{W}_L + \mathbf{V}_L}_{=:\mathbf{E}_L}$$

$$\text{where } \boldsymbol{\Phi}_L = \begin{bmatrix} \boldsymbol{\phi}_1 & \dots & \boldsymbol{\phi}_L \end{bmatrix}^\top ,$$
$$\boldsymbol{\phi}_l^\top = \mathbf{D}_{k-L+l}^{\boldsymbol{\theta}} + \sum_{j=1}^{l-1} \mathbf{C}\mathbf{A}^{l-j-1} \mathbf{B}_{k-L+j}^{\boldsymbol{\theta}} \ . \tag{10}$$

Now, a maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ can be calculated based on the measured residuals $\mathbf{R}_L$ and the distribution of the combined noise $\mathbf{E}_L$. It is asymptotically normal [12]:

$$\hat{\boldsymbol{\theta}} \overset{\text{asymp.}}{\sim} \mathcal{N}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}\right), \ \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \left(\boldsymbol{\Phi}_L^\top \boldsymbol{\mathcal{I}}_{\mathbf{E}_L} \boldsymbol{\Phi}_L\right)^{-1} \ . \tag{11}$$

In the special case of Gaussian noise, a closed form solution is given by the Generalised Least Squares (GLS) estimator:[5]

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{\Phi}_L^\top \boldsymbol{\Sigma}_{\mathbf{E}_L}^{-1} \boldsymbol{\Phi}_L\right)^{-1} \boldsymbol{\Phi}_L^\top \boldsymbol{\Sigma}_{\mathbf{E}_L}^{-1} \mathbf{R}_L \tag{12a}$$

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}} = \left(\boldsymbol{\Phi}_L^\top \boldsymbol{\Sigma}_{\mathbf{E}_L}^{-1} \boldsymbol{\Phi}_L\right)^{-1} \ . \tag{12b}$$

Alternatively, a recursive algorithm based on a Kalman filter is proposed by WILLSKY and JONES [13].

The decision rule of the Wald test is then based on the estimate's Mahalanobis norm, employing a threshold γ [15]:

$$\hat{\boldsymbol{\theta}}^\top \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1} \hat{\boldsymbol{\theta}} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{\gtrless}} \gamma \ . \tag{13}$$

[3]The case where $\boldsymbol{\theta}_0 \neq \mathbf{0}$ can be treated by attributing its influence to the known deterministic system input $\mathbf{u}_k$ and defining $\mathcal{H}_1$ as $\boldsymbol{\theta}_1 - \boldsymbol{\theta}_0$.

[4]One way to eliminate $\mathbf{x}_{k-L+1}$ is a *parity space* approach [15]. The residual is projected in a space orthogonal to $\mathbf{x}_{k-L+1}$. A comparison between this concept and the case where $\mathbf{x}_{k-L+1}$ is estimated, e.g. using a Kalman filter, can be found in [18].

[5]Comparing the MLE and GLS estimates exemplifies the theory from Sec. II-B. The GLS (12) solely considers the second moment $\boldsymbol{\Sigma}_{\mathbf{E}_L}$, which is a complete description only for Gaussian noise. Due to the inequality (8), i.e. $\boldsymbol{\Sigma}_{\mathbf{E}_L} \succeq \boldsymbol{\mathcal{I}}_{\mathbf{E}_L}^{-1}$ follows $\left(\boldsymbol{\Phi}_L \boldsymbol{\Sigma}_{\mathbf{E}_L}^{-1} \boldsymbol{\Phi}_L^\top\right)^{-1} \succeq \left(\boldsymbol{\Phi}_L \boldsymbol{\mathcal{I}}_{\mathbf{E}_L} \boldsymbol{\Phi}_L^\top\right)^{-1}$ and thus the GLS estimate has a larger uncertainty than the MLE (11).

The parameter γ determines the tradeoff between the probabilities of detection P_{D} and false alarm P_{FA}. In order to make an informed decision it is thus beneficial to know the statistical properties of $\hat{\boldsymbol{\theta}}$.

An important property of the Wald test is that it asymptotically possesses the same optimal statistics as the GLR test. Thus, the GLR test statistics in a non-Gaussian context can be derived using the previously reviewed concepts on estimator performance, which will be studied in the following.

III. GLR TEST STATISTIC

The main contributions are developed in this section. At first, the GLR test statistic is presented in Sec. III-A based on the regression problem (10). The central issue with this result which has been derived in [15] is that it allows no intuitive insight on the influence of the data window length L. Moreover, the (numerical) computation is demanding because depending on L, a large matrix has to be inverted. This motivates the derivation of a recursive solution in Sec. III-B, which has to the best of the authors' knowledge not been shown before. Properties of this result are then discussed in Sec. III-C.

A. Block matrix form

The Wald test is based on the unbiased, efficient MLE and therefore the estimation covariance equals the CRLB (6). Thus, the Mahalanobis norm of the estimate is asymptotically χ^2_ν-distributed with ν degrees of freedom

$$\hat{\boldsymbol{\theta}}^\top \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1} \hat{\boldsymbol{\theta}} \overset{\text{asymp.}}{\sim} \begin{cases} \chi^2_\nu\left(\lambda_L\right) & \text{under } \mathcal{H}_1 \\ \chi^2_\nu\left(0\right) & \text{under } \mathcal{H}_0 \end{cases} \tag{14a}$$

where the non-centrality parameter is

$$\lambda_L = \left(\boldsymbol{\Phi}_L \boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{E}_L} \left(\boldsymbol{\Phi}_L \boldsymbol{\theta}_1\right) \ . \tag{14b}$$

One can interpret (14b) as follows: λ_L is a measure on how different the distributions of $\hat{\boldsymbol{\theta}}^\top \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1} \hat{\boldsymbol{\theta}}$ under the two hypotheses are. For the detector, finding the correct distinction becomes easier for higher λ_L because this separates the non-central $\chi^2_\nu\left(\lambda_L\right)$-distribution further from $\chi^2_\nu\left(0\right)$.

It is remarkable that the probability of a false alarm P_{FA}, i.e. the detection threshold γ is exceeded under $\mathcal{H}_0$, remains independent of any system parameter. Therefore, a system improvement only affects the true positive detection performance, i.e. the non-centrality parameter of $\chi^2_\nu\left(\lambda_L\right)$. The false alarm probability is thus only indirectly reduced by choosing a higher threshold value γ which achieves the same probability of detection P_{D} but less false alarms.

Using (14), one can calculate a closed form for the detector ROC curve, i.e. P_{D} as function of P_{FA}. Let the cumulative distribution function (cdf) of the non-central $\chi^2_\nu\left(\lambda_L\right)$-distribution be denoted as $P_{\chi^2_\nu(\lambda_L)}\left(\cdot\right)$ and the quantile function (inverse cdf) of the central χ^2_ν-distribution as $Q_{\chi^2_\nu}\left(\cdot\right)$:

$$P_{\mathrm{D}} = 1 - P_{\chi^2_\nu(\lambda_L)}\left(Q_{\chi^2_\nu}\left(1 - P_{\mathrm{FA}}\right)\right) \ . \tag{15}$$

Although all relevant information on the test statistic is contained in the scalar λ_L, calculating this quantity requires the inversion of a matrix with dimension $L\cdot m \times L \cdot m$ [15]:

$$\boldsymbol{\mathcal{I}}_{\mathbf{E}_L} = \left(\boldsymbol{\mathcal{O}}_L \boldsymbol{\mathcal{I}}^{-1}_{\tilde{\mathbf{x}}_{k-L+1}} \boldsymbol{\mathcal{O}}_L^\top + \mathbf{H}^{\mathbf{w}}_L \boldsymbol{\mathcal{I}}^{-1}_{\mathbf{W}_L} \mathbf{H}^{\mathbf{w}\top}_L + \boldsymbol{\mathcal{I}}^{-1}_{\mathbf{V}_L}\right)^{-1} . \tag{16}$$

Due to the assumption of white noise processes it follows that $\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{W}_L}$ and $\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{V}_L}$ are block diagonal matrices with entries $\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{w}}$ and $\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{v}}$, respectively [15].

Still, the sum in (16) results in a dense matrix and finding the inverse is not straightforward. It will be analysed in the following section, how $\boldsymbol{\mathcal{I}}_{\mathbf{E}_{L+1}}$ for $L{+}1$ measurements relates to $\boldsymbol{\mathcal{I}}_{\mathbf{E}_L}$. This leads to a recursive expression for λ_{L+1} where the increment λ^{Δ}_{L+1} added by the $(L+1)$st measurement is identified.

B. Recursive form

At first, (16) is reformulated in terms of the predicted state $\hat{\mathbf{x}}_{k-l+1}$, $l = L, \dots 1$. Stacking these state predictions is denoted as $\hat{\mathbf{X}}_L = \begin{bmatrix}\hat{\mathbf{x}}^\top_{k-L+1} & \dots & \hat{\mathbf{x}}^\top_k\end{bmatrix}^\top$. Consequently, $\tilde{\mathbf{X}}_L$ is the stacked state prediction error. Moreover, define $\mathbf{H}_L = \mathbf{I}_L \otimes \mathbf{C}$ as a block diagonal matrix of L times $\mathbf{C}$ from the measurement model (1b). Hence:

$$\mathbf{H}_L \tilde{\mathbf{X}}_L = \boldsymbol{\mathcal{O}}_L \tilde{\mathbf{x}}_{k-L+1} + \mathbf{H}^{\mathbf{w}}_L \mathbf{W}_L \ . \tag{17}$$

Then, (16) is rewritten using the matrix inversion lemma[6]:

$$\begin{aligned}\boldsymbol{\mathcal{I}}_{\mathbf{E}_L} &= \left(\mathbf{H}_L \boldsymbol{\mathcal{I}}^{-1}_{\tilde{\mathbf{X}}_L} \mathbf{H}_L^\top + \boldsymbol{\mathcal{I}}^{-1}_{\mathbf{V}_L}\right)^{-1} \\ &= \boldsymbol{\mathcal{I}}_{\mathbf{V}_L} - \boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\mathbf{H}_L \left(\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_L} + \mathbf{H}_L^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\mathbf{H}_L\right)^{-1} \mathbf{H}_L^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_L} .\end{aligned} \tag{18}$$

At first glance, this has not alleviated the problem.[7] However, $\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{V}_L}$ is block-diagonal due to the white noise assumption so the inverse is readily available. Denote

$$\mathbf{M}_L := \left(\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_L} + \mathbf{H}_L^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\mathbf{H}_L\right)^{-1} \tag{19}$$

in (18) which comprises $\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_L}$ instead of its inverse. Thus, a result on the structure of this matrix can be employed which allows to find an expression for $\mathbf{M}_{L+1}$ as a function of $\mathbf{M}_L$. To this end, $\mathbf{M}_{L+1}$ is partitioned:

$$\mathbf{M}_{L+1} = \begin{array}{c c} & \begin{array}{cc} L\cdot n & n \end{array} \\ \begin{array}{c} L\cdot n \\ n \end{array} & \begin{bmatrix} \mathbf{M}_{L+1,11} & \mathbf{M}_{L+1,12} \\ \mathbf{M}_{L+1,21} & \mathbf{M}_{L+1,22}\end{bmatrix}\end{array} . \tag{20}$$

Lemma 1: A way to recursively calculate $\mathbf{M}_{L+1}$ is:

$$\mathbf{M}_{L+1,11} = \mathbf{M}_L - \tag{21a}$$

$$\begin{bmatrix}\mathbf{M}_{L,21}\\ \mathbf{M}_{L,22}\end{bmatrix} \cdot \boldsymbol{\Upsilon}\left(\boldsymbol{\Upsilon} + \mathbf{M}^{-1}_{L,22}\right)^{-1} \mathbf{M}^{-1}_{L,22} \cdot \begin{bmatrix}\mathbf{M}_{L,21} & \mathbf{M}_{L,22}\end{bmatrix}$$

$$\mathbf{M}_{L+1,22} = \tag{21b}$$

$$\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C} - \boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\mathcal{D}}_{11} + \mathbf{M}^{-1}_{L,22}\right)^{-1}\boldsymbol{\mathcal{D}}_{12}\right)^{-1}$$

$$\mathbf{M}_{L+1,21} = \mathbf{M}^\top_{L+1,12} = \boldsymbol{\Gamma}_{L+1} \cdot \begin{bmatrix}\mathbf{M}_{L,21} & \mathbf{M}_{L,22}\end{bmatrix} \tag{21c}$$

with the following abbreviations

$$\boldsymbol{\Upsilon} = \boldsymbol{\mathcal{D}}_{11} - \boldsymbol{\mathcal{D}}_{12}\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1}\boldsymbol{\mathcal{D}}_{21} \ , \tag{22a}$$

$$\boldsymbol{\Gamma}_{L+1} = \tag{22b}$$

$$-\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1}\boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\Upsilon} + \mathbf{M}^{-1}_{L,22}\right)^{-1}\mathbf{M}^{-1}_{L,22} \ ,$$

and[8]

$$\boldsymbol{\mathcal{D}}_{11} = \mathbf{A}^\top \left(\mathbf{B}^{\mathbf{w}} \boldsymbol{\mathcal{I}}^{-1}_{\mathbf{w}} \mathbf{B}^{\mathbf{w}\top}\right)^{-1} \mathbf{A}, \tag{22c}$$

$$\boldsymbol{\mathcal{D}}_{12} = \boldsymbol{\mathcal{D}}_{21}^\top = -\mathbf{A}^\top \left(\mathbf{B}^{\mathbf{w}} \boldsymbol{\mathcal{I}}^{-1}_{\mathbf{w}} \mathbf{B}^{\mathbf{w}\top}\right)^{-1} , \tag{22d}$$

$$\boldsymbol{\mathcal{D}}_{22} = \left(\mathbf{B}^{\mathbf{w}} \boldsymbol{\mathcal{I}}^{-1}_{\mathbf{w}} \mathbf{B}^{\mathbf{w}\top}\right)^{-1} . \tag{22e}$$

The initial value for $L = 1$ is

$$\mathbf{M}_1 = \mathbf{M}_{1,22} = \left(\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{x}}_{k-L+1}} + \mathbf{C}^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1} . \tag{23a}$$

Proof: See Sec. VI-A.

Remark: All expressions in (21) only require to invert matrices of size $n \times n$. Also note that $\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_L}$ has been derived for a *predicted* state $\hat{\mathbf{x}}_{k|k-L+1}$. Thus, $\boldsymbol{\mathcal{D}}_{22}$ in (22e) does not encompass a term related to measurement information $\boldsymbol{\mathcal{I}}_{\mathbf{v}}$. However, all occurrences of $\boldsymbol{\mathcal{D}}_{22}$ in (21)-(22) are in conjunction with $\mathbf{C}^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}$. Therefore, one could enhance (22e) with this term and recover the identical expression as for a *filtered* state estimate $\hat{\mathbf{x}}_{k|k}$ as shown in [9].

One remarkable aspect of (21) is that $\mathbf{M}_{L+1,11}$ consists of $\mathbf{M}_L$ and a second additive term. Still, matrices of growing dimension are used in the calculation which motivates further inspection of $\boldsymbol{\mathcal{I}}_{\mathbf{E}_L}$ in (18) and eventually λ_L in (14b).

A recursive formulation of λ_{L+1} is then achieved where all matrix operations are of dimension $n \times n$. This result, provided in theorem 1, is the main contribution of this work.

Theorem 1: The non-centrality parameter λ_L can be recursively calculated as

$$\begin{aligned}\lambda_1 &= \left(\boldsymbol{\phi}_1^\top \boldsymbol{\theta}_1\right)^\top \left(\mathbf{C}\boldsymbol{\mathcal{I}}^{-1}_{\tilde{\mathbf{x}}_{k-L+1}}\mathbf{C}^\top + \boldsymbol{\mathcal{I}}^{-1}_{\mathbf{v}}\right)^{-1}\left(\boldsymbol{\phi}_1^\top \boldsymbol{\theta}_1\right) \ , \\ \lambda_{L+1} &= \lambda_L + \lambda^{\Delta}_{L+1} \ , \quad L \geq 1\end{aligned} \tag{24}$$

where the increment λ^{Δ}_{L+1} is given by the quadratic form

$$\lambda^{\Delta}_{L+1} = (\mathbf{a}-\mathbf{b})^\top \left(\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{v}} - \mathbf{C}\mathbf{M}_{L+1,22}\mathbf{C}^\top\right)^{-1}(\mathbf{a}-\mathbf{b}) \ ,$$

with (25a)

$$\mathbf{a} = \mathbf{C}\boldsymbol{\Lambda}^\top_{L+1} \ , \tag{25b}$$

$$\mathbf{b} = \left(\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{v}} - \mathbf{C}\mathbf{M}_{L+1,22}\mathbf{C}^\top\right)\boldsymbol{\mathcal{I}}_{\mathbf{v}}\left(\boldsymbol{\phi}^\top_{L+1}\boldsymbol{\theta}_1\right) \ , \tag{25c}$$

$$\boldsymbol{\Lambda}_{L+1} = \left(\boldsymbol{\Lambda}_L + \left(\boldsymbol{\phi}^\top_L \boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\mathbf{M}_{L,22}\right)\boldsymbol{\Gamma}^\top_{L+1} \ . \tag{25d}$$

and $\boldsymbol{\Lambda}_1 = \mathbf{0}_{\nu\times n}$, $\boldsymbol{\phi}^\top_L$ as defined in (10), $\mathbf{M}_{L,22}$ according to (21b) and $\boldsymbol{\Gamma}_L$ from (22b). *Proof:* See Sec. VI-B.

Remark: The only submatrix of $\mathbf{M}_L$ which needs to be calculated recursively is $\mathbf{M}_{L,22}$, all other relevant components from (21) have been comprised in the recursion of $\boldsymbol{\Lambda}_L$.

[6] $\left(\mathbf{A}+\mathbf{C}\mathbf{B}\mathbf{C}^\top\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}\left(\mathbf{B}^{-1} + \mathbf{C}^\top \mathbf{A}^{-1}\mathbf{C}\right)^{-1}\mathbf{C}^\top \mathbf{A}^{-1}$

[7] In fact it is now required to invert a matrix of dimension $L\cdot n \times L\cdot n$ and typically we have more measurement than state variables, i.e. $n \geq m$.

[8] Here, it is assumed that the density $p\left(\mathbf{x}_{k+1} \,|\mathbf{x}_k\right)$ of the system (1a) is nonsingular and thus $\mathbf{B}^{\mathbf{w}}\boldsymbol{\mathcal{I}}^{-1}_{\mathbf{w}}\mathbf{B}^{\mathbf{w}\top}$ is regular. This case is mainly treated in [9], whereas further details on the singular case are provided in [10].

C. Discussion of the recursive form

The recursion for the non-centrality parameter from theorem 1 provides the scalar increments λ_{L+1}^{Δ} which increase λ with each additional measurement taken into account.

Eventually, the implications of λ_{L+1}^{Δ} on the detector performance, i.e. the ROC are of interest. Therefore, the evolution of the detection probability (15) with L is studied:

$$P_D^{\Delta} \approx \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(1 - P_{\chi_\nu^2(\lambda)}\left(Q_{\chi_\nu^2}\left(1 - P_{\mathrm{FA}}\right)\right)\right)\Big|_{\lambda_L} \cdot \lambda_{L+1}^{\Delta} \,. \quad (26)$$

However, as there is no closed-form expression to the cumulative distribution function $P_{\chi_\nu^2(\lambda_L)}(\cdot)$, we have to resort to approximate expressions. At first, P_D is therefore approximated in terms of elementary functions of λ_L. Secondly, the derivative thereof yields a first-order approximation of the amount of improvement P_D^{Δ} that can be expected from an additional measurement (i.e. λ_{L+1}^{Δ}).

There are numerous approximations of the non-central χ_ν^2-distribution. We employ a compact variant by PATNAIK [19]:

$$P_{\chi_\nu^2(\lambda)}(z) \approx P_{\mathcal{N}(\mu_\zeta(\lambda),1)}(\zeta(\lambda)) \quad (27)$$

$$\text{with } \zeta(\lambda) = \sqrt{2z\frac{\nu+\lambda}{\nu+2\lambda}},\ \mu_\zeta(\lambda) = \sqrt{2\frac{(\nu+\lambda)^2}{\nu+2\lambda} - 1}. \quad (28)$$

Thus, an approximate expression for (26) is obtained:

$$P_D^{\Delta} \approx \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\zeta(\lambda_L) - \mu_\zeta(\lambda_L)\right)^2\right) \cdot \left(\zeta'(\lambda_L) - \mu_\zeta'(\lambda_L)\right) \cdot \lambda_{L+1}^{\Delta} \quad (29)$$

where $\mu_\zeta'(\cdot)$ and $\zeta'(\cdot)$ denote the derivatives of (28).

Instead of studying the ROC in terms of P_{D} for a fixed P_{FA}, one could further analyse the area under the curve (AUC) metric, that is the integral of $\int_0^1 P_{\mathrm{D}}(P_{\mathrm{FA}})\,\mathrm{d}P_{\mathrm{FA}}$.

IV. SIMULATION EXAMPLE

The previously studied asymptotic performance bound is now compared to a GLR test in Monte-Carlo simulations. It is expected, that the simulated GLR test reaches the optimum ROC curve predicted by the model (15). While previous works [15], [16] have analysed the optimal MLE in contrast to an approximate BLUE in a non-Gaussian system, emphasis is put on the detector window length L here.

A. Setup and parameter values

An application from the driver assistance domain is considered. While the example is not explicitly related to a specific ADAS, the ambiguousness of the simulated situation makes it a typical candidate for driver intention recognition.

The initial conditions of the situation are illustrated in Fig. 2. Two vehicles are approaching a traffic light, e.g. at an intersection, which has just switched from green to yellow. Both vehicles are driving at $50\,\mathrm{km/h}$ with an initial distance of $37\,\mathrm{m}$ to the intersection and it is assumed that the duration for the yellow phase is $3\,\mathrm{s}$.[9] Therefore, the first vehicle may pass the traffic light (legally) by keeping its current velocity.

[9] According to German legislation for inner-city traffic lights.

TABLE I
SIMULATION PARAMETERS

Variable	**Value**
Sampling time	$T = 0.0675\,\mathrm{s}$
Initial state	$\mathbf{x}_{k-L+1} = \begin{bmatrix}10\,\mathrm{m} & 0\end{bmatrix}$
Initial uncertainty	$\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{x}}_{k-L+1}}^{-1} = \mathrm{diag}\begin{bmatrix}0.25\,\mathrm{m} & 0.2\,\mathrm{m/s}\end{bmatrix}^2$
Process noise	$\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{w}}}^{-1} = \left(0.2\,\mathrm{m/s^2}\right)^2\mathrm{s}^{-1}\begin{bmatrix}\frac{1}{3}T^3 & \frac{1}{2}T^2 \\ \frac{1}{2}T^2 & T\end{bmatrix}$
Measurement noise	$R = \mathcal{I}_v^{-1} = (0.2\,\mathrm{m})^2$
Brake ramp slope	$r_{\mathrm{brake}} = -8\,\mathrm{m/s^3}$
Brake ramp up time	$t_{\mathrm{ramp}} = 0.405\,\mathrm{s}$

Nevertheless, braking to stand-still is a plausible option for a more cautious driver.

From the perspective of a driver assistance system built into the second vehicle, early differentiation between the two driver intentions is crucial for reliable prediction and action planning. Thus, the problem is formulated as a change detection task in the framework from Sec. II-A. The time-discrete (sampling time T, $t_k = k \cdot T$) relative longitudinal dynamics are described by the state variables x (relative distance) and v_x (relative velocity) as:

$$\begin{bmatrix}x \\ v_x\end{bmatrix}_{k+1} = \begin{bmatrix}1 & T \\ 0 & 1\end{bmatrix}\begin{bmatrix}x \\ v_x\end{bmatrix}_k + \begin{bmatrix}0.5T^2 \\ T\end{bmatrix} f_k + \mathbf{w}_k \quad (30a)$$

$$y_k = \begin{bmatrix}1 & 0\end{bmatrix}\mathbf{x}_k + v_k\,. \quad (30b)$$

All noise processes are assumed Gaussian and the parameters of this system are detailed in Tab. I. The brake deceleration f_k is modelled using a simple ramp function with the slope parameter $\theta = r_{\mathrm{brake}}$ and fixed ramp up time t_{ramp}:

$$f_k = \left[t_k - (t_k - t_{\mathrm{ramp}})\,\sigma(t_k - t_{\mathrm{ramp}})\right] \cdot r_{\mathrm{brake}}\,. \quad (31)$$

The resulting deceleration and velocity are shown in Fig. 3.

It is assumed that a Gaussian estimate of the initial state with $\hat{\mathbf{x}}_{k-L+1} = \mathcal{N}\left(\mathbf{x}_{k-L+1}, \boldsymbol{\mathcal{I}}_{\tilde{\mathbf{x}}_{k-L+1}}^{-1}\right)$ is given. Then, the Generalised Least-Squares estimator (12) provides the maximum likelihood solution to the regression problem (10). The detection is then performed according to (13).

Monte-Carlo simulation results of this decision rule will be presented in the following section and compared to the asymptotic GLR test statistic (15).

B. Simulation results

Now, the previously described system is simulated in $N_{\mathrm{sim}} = 10^5$ independent iterations for different detection window lengths L. The initial time-step considered in the detection corresponds to the first occurrence of the deceleration.[10] In each iteration, an estimate $\hat{\theta}$ of the brake ramp slope

[10] In practice, a sliding window will encompass time-steps both before and after occurrence of the change. One could treat the time of occurrence as an additional unknown parameter which is to be estimated in the GLR test [13]. By considering only those measurements which contribute information to the detection problem, we thus analyse the optimal performance.

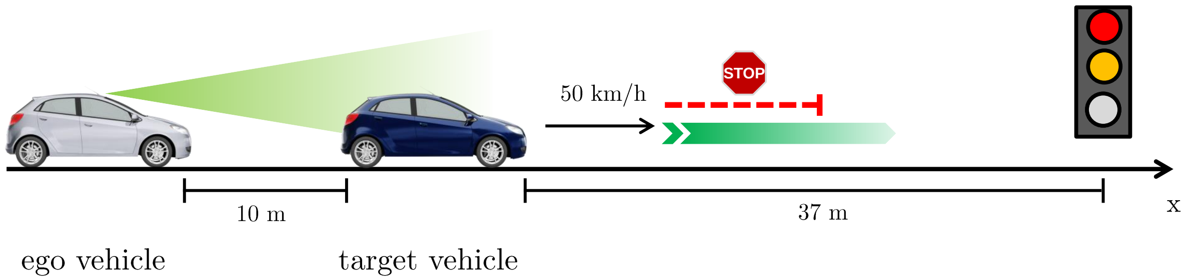
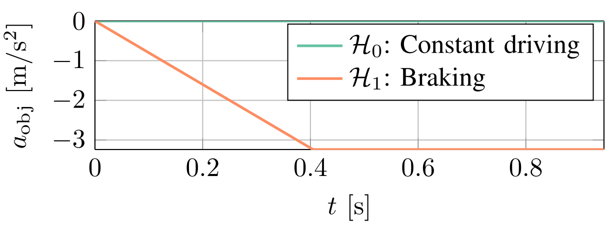
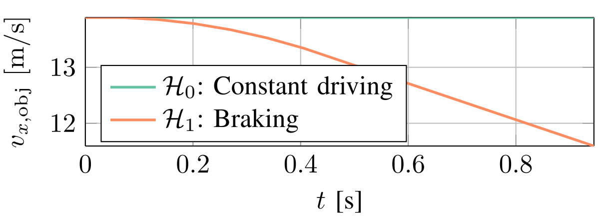


Fig. 2. Illustration of the application example. The ego vehicle (equipped with a front facing environment sensor) and a preceding target vehicle approach a traffic light at equal initial velocity. Given the remaining distance to the traffic light when the signal switches from green to yellow, it is ambiguous, whether the driver of the target vehicle intends to stop or pass. For driver assistance functions, early detection of this driver intention is a crucial task.



(a) Acceleration profile



(b) Relative velocity

Fig. 3. Acceleration and velocity profiles.

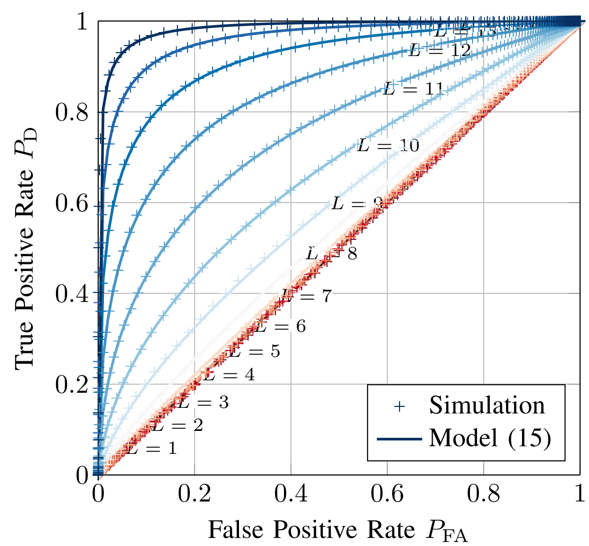


Fig. 4. Receiver Operating Characteristic curves.

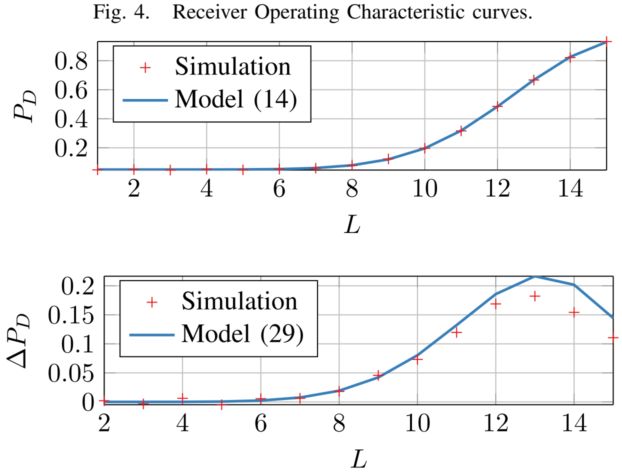


Fig. 5. Detection probability P_{D} and increments P_D^{Δ} for false positive probability $P_{\mathrm{FA}} = 0.05$.

parameter is calculated according to (12) and the detection rule (13) is evaluated for different threshold values γ:

- Firstly, γ is varied within the minimum and maximum observed values on the left hand side of (13). As the true realisations of the two hypotheses are known, the true and false positive rates can be calculated. This leads to the ROC curves displayed in Fig. 4. Each curve corresponds to one window length L.
- Secondly, the true positive rate P_{D} at a constant value $P_{\mathrm{FA}} = 0.05$ as part of the overall ROC curve is evaluated and visualised over L in Fig. 5(a).

In both cases, the predictions given by the asymptotic GLR test statistic (15) are shown for comparison. It can be observed that a a good correspondence to the simulated values is achieved. It is furthermore notable from Fig. 5(a) that P_{D} over L resembles a sigmoid curve. Therefore, even if only approximate models are available, these can be employed to estimate the interval of linear growth and thus a sensible initial value on the minimum required L.

V. Conclusion

Statistical models of upper performance bounds are important means for an efficient system design process. This work has studied such bounds in a binary decision problem, e.g. an intention recognition task for driver assistance functions. The theoretical contribution is a recursive form of the optimum test statistic. This allows to study how the number of measurements considered in the decision problem affects the attainable performance. A Monte-Carlo simulation has been used to verify and investigate the analytical result.

In future works, one could enhance the problem formulation to m-ary decision problems. Moreover, connecting the dynamic system description used here to other models, e.g. Hidden Markov Models [7], would be an interesting topic.

VI. Appendix

A. Proof of lemma 1

In order to find a recursion for $\mathbf{M}_{L+1}$, a partitioning for $\mathbf{M}_L$ is introduced:

$$\mathbf{M}_L = \left(\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_L} + \mathbf{H}_L^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\mathbf{H}_L\right)^{-1} \tag{32}$$
$$= \left(\begin{bmatrix}\boldsymbol{\mathcal{A}} & \boldsymbol{\mathcal{B}}\\ \boldsymbol{\mathcal{B}}^\top & \boldsymbol{\mathcal{C}}\end{bmatrix} + \begin{bmatrix}\mathbf{H}_{L-1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L-1}}\mathbf{H}_{L-1} & \mathbf{0}\\ \mathbf{0} & \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\end{bmatrix}\right)^{-1}$$

Calculating the inverse block-wise gives:

$$\mathbf{M}_{L,11} = \tag{33a}$$
$$\left(\boldsymbol{\mathcal{A}} + \mathbf{H}_{L-1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L-1}}\mathbf{H}_{L-1} - \boldsymbol{\mathcal{B}}\left(\boldsymbol{\mathcal{C}} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1}\boldsymbol{\mathcal{B}}^\top\right)^{-1}$$
$$\mathbf{M}_{L,22} = \tag{33b}$$
$$\left(\boldsymbol{\mathcal{C}} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C} - \boldsymbol{\mathcal{B}}^\top\left(\boldsymbol{\mathcal{A}} + \mathbf{H}_{L-1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L-1}}\mathbf{H}_{L-1}\right)^{-1}\boldsymbol{\mathcal{B}}\right)^{-1}$$
$$\mathbf{M}_{L,21} = -\mathbf{M}_{L,22}\boldsymbol{\mathcal{B}}^\top\left(\boldsymbol{\mathcal{A}} + \mathbf{H}_{L-1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L-1}}\mathbf{H}_{L-1}\right)^{-1} \tag{33c}$$
$$\mathbf{M}_{L,12} = \mathbf{M}_{L,21}^\top \tag{33d}$$

The central idea is then to leverage the Markov assumption on the system (1a). This yields that $\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_{L+1}}$ has a block-diagonal form [9]:[11]

$$\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_{L+1}} = \begin{bmatrix}\boldsymbol{\mathcal{A}} & \boldsymbol{\mathcal{B}} & \mathbf{0}\\ \boldsymbol{\mathcal{B}}^\top & \boldsymbol{\mathcal{C}}+\boldsymbol{\mathcal{D}}_{11} & \boldsymbol{\mathcal{D}}_{12}\\ \mathbf{0} & \boldsymbol{\mathcal{D}}_{21} & \boldsymbol{\mathcal{D}}_{22}\end{bmatrix} \tag{34}$$

and therefore

$$\mathbf{M}_{L+1} = \left(\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_{L+1}} + \mathbf{H}_{L+1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L+1}}\mathbf{H}_{L+1}\right)^{-1} \tag{35}$$

can be calculated block-wise. The exact expressions for $\boldsymbol{\mathcal{D}}_{11}$, $\boldsymbol{\mathcal{D}}_{21}$ and $\boldsymbol{\mathcal{D}}_{22}$ follow from inserting the system equation (1a) into the Fisher information matrix (5) and are given in (22).

Consider the inversion in (35) which defines $\mathbf{M}_{L+1,11}$:

$$\mathbf{M}_{L+1,11} = \left(\mathbf{M}_L^{-1} + \begin{bmatrix}\mathbf{0}\\ \mathbf{I}_n\end{bmatrix}\boldsymbol{\Upsilon}\begin{bmatrix}\mathbf{0} & \mathbf{I}_n\end{bmatrix}\right)^{-1}$$
$$= \mathbf{M}_L - \begin{bmatrix}\mathbf{M}_{L,12}\\ \mathbf{M}_{L,22}\end{bmatrix}\boldsymbol{\Upsilon}\left(\mathbf{I}_n + \mathbf{M}_{L,22}\boldsymbol{\Upsilon}\right)^{-1}\begin{bmatrix}\mathbf{M}_{L,21} & \mathbf{M}_{L,22}\end{bmatrix}$$
$$\text{where } \boldsymbol{\Upsilon} := \left(\boldsymbol{\mathcal{D}}_{11} - \boldsymbol{\mathcal{D}}_{12}\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1}\boldsymbol{\mathcal{D}}_{21}\right). \tag{36}$$

Here, we have used a version of the matrix inversion lemma suitable for singular $\boldsymbol{\Upsilon}$ [20].[12].

Next, the lower right element $\mathbf{M}_{L+1,22}$ is studied:

$$\mathbf{M}_{L+1,22} = \Bigg(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C} - \boldsymbol{\mathcal{D}}_{21}\Big(\boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{D}}_{11} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}$$
$$- \boldsymbol{\mathcal{B}}^\top\left(\boldsymbol{\mathcal{A}} + \mathbf{H}_{L-1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L-1}}\mathbf{H}_{L-1}\right)^{-1}\boldsymbol{\mathcal{B}}\Big)^{-1}\boldsymbol{\mathcal{D}}_{12}\Bigg)^{-1}$$
$$\overset{(33b)}{=} \left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C} - \boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\mathcal{D}}_{11} + \mathbf{M}_{L,22}^{-1}\right)^{-1}\boldsymbol{\mathcal{D}}_{12}\right)^{-1}. \tag{37}$$

Finally, $\mathbf{M}_{L+1,21}$ (which can be shown to equal $\mathbf{M}_{L+1,12}^\top$):

$$\mathbf{M}_{L+1,21} = -\mathbf{M}_{L+1,22}\cdot\begin{bmatrix}\mathbf{0} & \boldsymbol{\mathcal{D}}_{21}\end{bmatrix}$$
$$\cdot\begin{bmatrix}\boldsymbol{\mathcal{A}} + \mathbf{H}_{L-1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L-1}}\mathbf{H}_{L-1} & \boldsymbol{\mathcal{B}}\\ \boldsymbol{\mathcal{B}}^\top & \boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{D}}_{11} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\end{bmatrix}^{-1}$$
$$\overset{(33b)}{=} -\mathbf{M}_{L+1,22}\boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\mathcal{D}}_{11} + \mathbf{M}_{L,22}^{-1}\right)^{-1}$$
$$\cdot\begin{bmatrix}-\boldsymbol{\mathcal{B}}^\top\left(\boldsymbol{\mathcal{A}} + \mathbf{H}_{L-1}^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_{L-1}}\mathbf{H}_{L-1}\right)^{-1} & \mathbf{I}_n\end{bmatrix}$$
$$\overset{(33c)}{=} -\mathbf{M}_{L+1,22}\boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\mathcal{D}}_{11} + \mathbf{M}_{L,22}^{-1}\right)^{-1}\begin{bmatrix}\mathbf{M}_{L,22}^{-1}\mathbf{M}_{L,21} & \mathbf{I}_n\end{bmatrix}$$
$$\overset{(37)}{=} -\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C} - \boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\mathcal{D}}_{11} + \mathbf{M}_{L,22}^{-1}\right)^{-1}\boldsymbol{\mathcal{D}}_{12}\right)^{-1}$$
$$\cdot\boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\mathcal{D}}_{11} + \mathbf{M}_{L,22}^{-1}\right)^{-1}\mathbf{M}_{L,22}^{-1}\cdot\begin{bmatrix}\mathbf{M}_{L,21} & \mathbf{M}_{L,22}\end{bmatrix}$$
$$= -\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1}\boldsymbol{\mathcal{D}}_{21}$$
$$\cdot\left(\boldsymbol{\mathcal{D}}_{11} - \boldsymbol{\mathcal{D}}_{12}\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1}\boldsymbol{\mathcal{D}}_{21} + \mathbf{M}_{L,22}^{-1}\right)^{-1}$$
$$\cdot\mathbf{M}_{L,22}^{-1}\cdot\begin{bmatrix}\mathbf{M}_{L,21} & \mathbf{M}_{L,22}\end{bmatrix}$$
$$= \underbrace{-\left(\boldsymbol{\mathcal{D}}_{22} + \mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\right)^{-1}\boldsymbol{\mathcal{D}}_{21}\left(\boldsymbol{\Upsilon} + \mathbf{M}_{L,22}^{-1}\right)^{-1}\mathbf{M}_{L,22}^{-1}}_{=:\boldsymbol{\Gamma}_{L+1}}$$
$$\cdot\begin{bmatrix}\mathbf{M}_{L,21} & \mathbf{M}_{L,22}\end{bmatrix}$$
$$= \boldsymbol{\Gamma}_{L+1}\cdot\begin{bmatrix}\mathbf{M}_{L,21} & \mathbf{M}_{L,22}\end{bmatrix}. \tag{38}$$

Therefore, $\mathbf{M}_{L+1,21}$ consists of L submatrices $\mathbf{M}_{L+1,21}^{(l)}$ of size $n\times n$:

$$\mathbf{M}_{L+1,21} = \begin{bmatrix}\mathbf{M}_{L+1,21}^{(1)} & \dots & \mathbf{M}_{L+1,21}^{(L)}\end{bmatrix}. \tag{39}$$

With (38), one obtains an explicit expression:

$$\mathbf{M}_{L+1,21}^{(l)} = \prod_{j=l}^{L}\boldsymbol{\Gamma}_{j+1}\mathbf{M}_{l,22}. \tag{40}$$

B. Proof of theorem 1

First note, that $\boldsymbol{\Phi}_{L+1}$ as defined in (10) can be written as $\boldsymbol{\Phi}_{L+1} = \begin{bmatrix}\boldsymbol{\Phi}_L^\top & \boldsymbol{\phi}_{L+1}\end{bmatrix}^\top$. Then, inserting the block-matrices which constitute $\mathbf{M}_{L+1}$ from (21) into $\boldsymbol{\mathcal{I}}_{\mathbf{E}_{L+1}}$ in (18) and

[11] The notation of the entries of $\boldsymbol{\mathcal{I}}_{\tilde{\mathbf{X}}_{L+1}}$ is chosen in accordance with [9]. Though, in order to avoid confusion with the system matrices from (1), a calligraphic font ($\boldsymbol{\mathcal{A}}$) is used here.

[12] $\left(\mathbf{A}+\mathbf{C}\mathbf{B}\mathbf{C}^\top\right)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{C}\mathbf{B}\left(\mathbf{I}+\mathbf{C}^\top\mathbf{A}^{-1}\mathbf{C}\mathbf{B}\right)^{-1}\mathbf{C}^\top\mathbf{A}^{-1}$

applying the multiplications with $\mathbf{\Phi}_{L+1}\boldsymbol{\theta}_1$ from (14b) yields:

$$
\begin{aligned}
\lambda_{L+1} &= \begin{bmatrix} (\mathbf{\Phi}_L\boldsymbol{\theta}_1)^\top & \left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right)^\top \end{bmatrix} \boldsymbol{\mathcal{I}}_{\mathbf{E}_{L+1}} \begin{bmatrix} \mathbf{\Phi}_L\boldsymbol{\theta}_1 \\ \boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1 \end{bmatrix} \\
&= \underbrace{(\mathbf{\Phi}_L\boldsymbol{\theta}_1)^\top \left(\boldsymbol{\mathcal{I}}_{\mathbf{V}_L} - \boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\mathbf{H}_L\mathbf{M}_L\mathbf{H}_L^\top\boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\right)(\mathbf{\Phi}_L\boldsymbol{\theta}_1)}_{=\lambda_L} \\
&\quad + (\mathbf{\Phi}_L\boldsymbol{\theta}_1)^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\mathbf{H}_L \begin{bmatrix} \mathbf{M}_{L,21}^\top \\ \mathbf{M}_{L,22} \end{bmatrix} \mathbf{\Upsilon}\left(\mathbf{\Upsilon} + \mathbf{M}_{L,22}^{-1}\right)^{-1} \\
&\qquad \cdot \mathbf{M}_{L,22}^{-1} \begin{bmatrix} \mathbf{M}_{L,21} & \mathbf{M}_{L,22} \end{bmatrix} \mathbf{H}_L^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_L} (\mathbf{\Phi}_L\boldsymbol{\theta}_1)^\top \\
&\quad - 2(\mathbf{\Phi}_L\boldsymbol{\theta}_1)^\top \boldsymbol{\mathcal{I}}_{\mathbf{V}_L}\mathbf{H}_L\mathbf{M}_{L+1,21}^\top \mathbf{C}^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}} \left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right) \\
&\quad + \left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right)^\top \left(\boldsymbol{\mathcal{I}}_{\mathbf{v}} - \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\mathbf{M}_{L+1,22}\mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\right)\left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right) \\
&= \lambda_L + \lambda_{L+1}^{\Delta} \, . \qquad (41)
\end{aligned}
$$

The key to simplify λ_{L+1}^{Δ} is to study the product

$$
\underbrace{\mathbf{\Lambda}_{L+1}}_{\nu\times n} = \underbrace{(\mathbf{\Phi}_L\boldsymbol{\theta}_1)^\top}_{\nu\times L\cdot m} \cdot \underbrace{\boldsymbol{\mathcal{I}}_{\mathbf{V}_L}}_{L\cdot m\times L\cdot m} \cdot \underbrace{\mathbf{H}_L}_{L\cdot m\times L\cdot n} \cdot \underbrace{\mathbf{M}_{L+1,21}^\top}_{L\cdot n\times n} \, . \qquad (42)
$$

Inserting (40) into (42) yields

$$
\begin{aligned}
\mathbf{\Lambda}_{L+1} &= \sum_{l=1}^{L} \left(\boldsymbol{\phi}_l^\top\boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\mathbf{M}_{L+1,21}^{(l)}{}^\top \\
&= \sum_{l=1}^{L} \left(\boldsymbol{\phi}_l^\top\boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\left(\prod_{j=l}^{L}\mathbf{\Gamma}_{j+1}\mathbf{M}_{l,22}\right)^\top \\
&= \sum_{l=1}^{L-1} \left(\boldsymbol{\phi}_l^\top\boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\left(\prod_{j=l}^{L-1}\mathbf{\Gamma}_{j+1}\mathbf{M}_{l,22}\right)^\top \mathbf{\Gamma}_{L+1}^\top \\
&\quad + \left(\boldsymbol{\phi}_L^\top\boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\left(\mathbf{\Gamma}_{L+1}\mathbf{M}_{L,22}\right)^\top \\
&= \left(\mathbf{\Lambda}_L + \left(\boldsymbol{\phi}_L^\top\boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\mathbf{M}_{L,22}\right)\mathbf{\Gamma}_{L+1}^\top \, . \qquad (43)
\end{aligned}
$$

With this intermediate result, all occurrences of $\mathbf{M}_{L,21}$ and $\mathbf{M}_{L+1,21}$ in (41) can be replaced:[13]

$$
\begin{aligned}
\lambda_{L+1}^{\Delta} &= \underbrace{\left(\mathbf{\Lambda}_L + \left(\boldsymbol{\phi}_L^\top\boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\mathbf{M}_{L,22}\right)}_{\overset{(43)}{=}\mathbf{\Lambda}_{L+1}\left(\mathbf{\Gamma}_{L+1}^\top\right)^{-1}}\mathbf{\Upsilon} \qquad (44) \\
&\cdot \left(\mathbf{\Upsilon} + \mathbf{M}_{L,22}^{-1}\right)^{-1}\mathbf{M}_{L,22}^{-1}\left(\mathbf{\Lambda}_L + \left(\boldsymbol{\phi}_L^\top\boldsymbol{\theta}_1\right)^\top \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\mathbf{M}_{L,22}\right)^\top \\
&\quad - 2\mathbf{\Lambda}_{L+1}\mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right) \\
&\quad + \left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right)^\top \left(\boldsymbol{\mathcal{I}}_{\mathbf{v}} - \boldsymbol{\mathcal{I}}_{\mathbf{v}}\mathbf{C}\mathbf{M}_{L+1,22}\mathbf{C}^\top\boldsymbol{\mathcal{I}}_{\mathbf{v}}\right)\left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right) \\
&= (\mathbf{a} - \mathbf{b})^\top \left(\boldsymbol{\mathcal{I}}_{\mathbf{v}}^{-1} - \mathbf{C}\mathbf{M}_{L+1,22}\mathbf{C}^\top\right)^{-1}(\mathbf{a} - \mathbf{b}) \, ,
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{a} &= \mathbf{C}\mathbf{\Lambda}_{L+1}^\top , \\
\mathbf{b} &= \left(\boldsymbol{\mathcal{I}}_{\mathbf{v}}^{-1} - \mathbf{C}\mathbf{M}_{L+1,22}\mathbf{C}^\top\right)\boldsymbol{\mathcal{I}}_{\mathbf{v}}\left(\boldsymbol{\phi}_{L+1}^\top\boldsymbol{\theta}_1\right) \, . \qquad (45)
\end{aligned}
$$

[13] Showing the equality of (44) and (45) involves tedious but straightforward calculations. Due to lack of space, the details are omitted here.

References

[1] C. Laugier, I. Paromtchik, M. Perrollaz, M. Yong, J.-D. Yoder, C. Tay, K. Mekhnacha, and A. Negre, "Probabilistic analysis of dynamic scenes and collision risks assessment to improve driving safety," *Intelligent Transportation Systems Magazine, IEEE*, vol. 3, no. 4, pp. 4–19, 2011.

[2] M. Schreier, V. Willert, and J. Adamy, "Bayesian, maneuver-based, long-term trajectory prediction and criticality assessment for driver assistance systems," in *Intelligent Transportation Systems (ITSC), 17th International IEEE Conference on*, pp. 334–341, 2014.

[3] S. Lefèvre, D. Vasquez, and C. Laugier, "A survey on motion prediction and risk assessment for intelligent vehicles," *ROBOMECH Journal*, vol. 1, no. 1, 2014.

[4] J. Ru, V. Jilkov, X. Li, and A. Bashi, "Detection of target maneuver onset," *Aerospace and Electronic Systems, IEEE Transactions on*, vol. 45, no. 2, pp. 536–554, 2009.

[5] L. Yang, J. H. Yang, and E. Feron, "Multiple model estimation for improving conflict detection algorithms," in *Systems, Man and Cybernetics, IEEE International Conference on*, vol. 1, pp. 242–249, 2004.

[6] J. Schneider, A. Wilde, and K. Naab, "Probabilistic approach for modeling and identifying driving situations," in *Intelligent Vehicles Symposium, IEEE*, pp. 343–348, 2008.

[7] D. Meyer-Delius, C. Plagemann, and W. Burgard, "Probabilistic situation recognition for vehicular traffic scenarios," in *Robotics and Automation (ICRA), IEEE International Conference on*, pp. 459–464, 2009.

[8] T. Gindele, S. Brechtel, and R. Dillmann, "Learning context sensitive behavior models from observations for predicting traffic situations," in *Intelligent Transportation Systems (ITSC), 16th International IEEE Conference on*, pp. 1764–1771, 2013.

[9] P. Tichavsky, C. Muravchik, and A. Nehorai, "Posterior cramer-Rao bounds for discrete-time nonlinear filtering," *Signal Processing, IEEE Transactions on*, vol. 46, no. 5, pp. 1386–1396, 1998.

[10] N. Bergman, *Recursive Bayesian Estimation: Navigation and Tracking Applications*. Linköping studies in science and technology. thesis no 579, Linköping University, Department of Electrical Engineering, 1999.

[11] M. Basseville and I. V. Nikiforov, *Detection of Abrupt Changes: Theory and Application*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[12] B. C. Levy, *Principles of Signal Detection and Parameter Estimation*. Springer Publishing Company, Inc., 1 ed., 2008.

[13] A. Willsky and H. Jones, "A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems," *Automatic Control, IEEE Transactions on*, vol. 21, no. 1, pp. 108–112, 1976.

[14] R. Karlsson, J. Jansson, and F. Gustafsson, "Model-based statistical tracking and decision making for collision avoidance application," in *American Control Conference, Proceedings of the 2004*, vol. 4, pp. 3435–3440, 2004.

[15] G. Hendeby and F. Gustafsson, "Fundamental fault detection limitations in linear non-gaussian systems," in *Decision and Control, 44th IEEE Conference on and European Control Conference (CDC-ECC)*, pp. 338–343, 2005.

[16] G. Hendeby and F. Gustafsson, "Detection limits for linear non-gaussian state-space models," in *6th IFAC Symposium on Fault Detection, Supervision and Safety of Technical Processes*, pp. 282–287, 2006.

[17] D. Sengupta and S. Kay, "Parameter estimation and GLRT detection in colored non-gaussian autoregressive processes," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 38, no. 10, pp. 1661–1676, 1990.

[18] D. Törnqvist, F. Gustafsson, and I. Klein, "GLR tests for fault detection over sliding data windows," in *In Proceedings of 16th Triennial IFAC World Congress*, 2005.

[19] H. Sahai and M. M. Ojeda, "A comparison of approximations to percentiles of the noncentral χ^2-distribution," *Revista de Matemática: Teoría y Aplicaciones*, vol. 10, no. 1-2, pp. 57–76, 2003.

[20] H. V. Henderson and S. R. Searle, "On Deriving the Inverse of a Sum of Matrices," *SIAM Review*, vol. 23, no. 1, pp. 53–60, 1981.