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

Jan Erik Stellet\textsuperscript{1}, Jan Schumacher\textsuperscript{1}, Wolfgang Branz\textsuperscript{1}, J. Marius Zöllner\textsuperscript{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.
II. PRELIMINARIES

A. Model representations

Linear dynamic systems in state space form are considered in this work. The system state is denoted \( x_k \in \mathbb{R}^n \), \( y_k \in \mathbb{R}^n \) is the measurement output, \( u_k \) a deterministic known input and \( f_k \) a deterministic, but unknown additive input (fault):\(^1\)

\[
\begin{align*}
    x_{k+1} &= A_k x_k + B_k^u u_k + B_k^f f_k + B_k^w w_k \quad (1a) \\
    y_k &= C_k x_k + D_k^u u_k + D_k^f f_k + v_k. \quad (1b)
\end{align*}
\]

The stochastic inputs \( w_k \) and \( v_k \) denote white, independent process and measurement noise, respectively.

In order to make the detection of unconstant \( f_k \) 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 \( F_k \) and a \( n \)-dimensional, time-invariant coefficient vector \( \theta \in \mathbb{R}^n \) as \( f_k = F_k \theta \) [15]. How to describe a fault signal in this way will be illustrated for an application example in Sec. IV. Hence, \( B_k^f \) and \( D_k^f \) in (1) are replaced by \( B_k^f = F_k^T \theta \) and \( D_k^f = D_k^f F_k \).

For notational simplicity, the known input \( u_k \) 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 \( F_k \).

Multiple measurements are integrated in the change detection approach. Given an initial state \( x_{k-L+1} \), the stacked measurements \( Y_L = [y_{k-L+1}, \ldots, y_k] \) read [11]:

\[
    Y_L = \mathcal{O}_L x_k - L + 1 + H_L^\theta \Theta_L + H_L^\nu W_L + V_L.
\]

In this notation, \( \mathcal{O}_L \) is the extended observability matrix

\[
    \mathcal{O}_L = \begin{bmatrix} C^T & (CA)^T & \ldots & (CA^{L-1})^T \end{bmatrix}^T,
\]

and the influence of the inputs \( i \in \{\theta, w\} \) is given by

\[
    H_L^i = \begin{bmatrix} D_{k-L+1}^i & \ldots & 0 \\ C B_{k-L+1}^i & D_{k-L+2}^i & \ldots \\ \vdots & \vdots & \ddots & \ddots & \ddots \\ CA^{L-2} B_{k-L+1}^i & CA^{L-3} B_{k-L+2}^i & \ldots & 0 \\ \end{bmatrix},
\]

\(1\)For consistency with the notation in [15], \( f_k \) 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 \( D_k = 0 \).

The stacked vectors \( \Theta_L, W_L \) and \( V_L \) concatenate the respective signals \( \theta, w \) and \( v \) from times \( k-L+1, \ldots, k \).

B. Information and Accuracy

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

\[
    \mathcal{I}_y(\theta) = -E\left[ \nabla_{\theta} \nabla_{\theta}^T \ln p(y|\theta) \right]
\]

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

\[
    \Sigma_{\theta} = \text{cov}(\hat{\theta}) \succeq \mathcal{I}_y^{-1} (\theta).
\]

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 \( p(e) \) of a zero mean noise process \( e \) [17]:

\[
    \mathcal{I}_e = -E\left[ \nabla_e \nabla_e^T \ln p(e) \right].
\]

Here, we have the inequality

\[
    \Sigma_e = \text{cov}(e) \succeq \mathcal{I}_e^{-1}
\]

which can be interpreted as the information about \( e \) contained in its covariance. It is easily verified that for Gaussian noise \( e \sim \mathcal{N}(0, \Sigma_e) \), (7) holds with equality. Thus, approximating \( e \) 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 \( A \succeq B \) for matrices \( A, B \) denotes that the difference \( A - B \) is positive semidefinite.
The parameter $\gamma$ 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 $\theta$.

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 CRBL (6). Thus, the Mahalanobis norm of the estimate is asymptotically $\chi^2_{\nu}$-distributed with $\nu$ degrees of freedom

$$\hat{\theta} \sim N(\theta, \Sigma_0), \quad \Sigma_0 = \left(\Phi_L^T \mathcal{I}_{\Phi_L} \Phi_L\right)^{-1}. \tag{11}$$

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

$$\hat{\theta} = \left(\Phi_L^T \Sigma_{E_L}^{-1} \Phi_L\right)^{-1} \Phi_L^T \Sigma_{E_L}^{-1} \mathbf{R}_L, \tag{12a}$$

$$\Sigma_{\hat{\theta}} = \left(\Phi_L^T \Sigma_{E_L}^{-1} \Phi_L\right)^{-1}. \tag{12b}$$

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 $\gamma [15]$:

$$\hat{\theta} \sim \chi^2_{\nu} \frac{\gamma}{\nu}, \quad H_0 \tag{13}$$

3The case where $\theta_0 \neq 0$ can be treated by attributing its influence to the known deterministic system input $u_k$ and defining $H_1$ as $H_1 - H_0$.

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

5Comparing the MLE and GLS estimates exemplifies the theory from Sec. II-B. The GLS (12) solely considers the second moment $\Sigma_{E_L}$, which is a complete description only for Gaussian noise. Due to the inequality (8), i.e. $\Sigma_{E_L} \geq \mathcal{I}_{\Phi_E}$ follows $\left(\Phi_L^T \Sigma_{E_L}^{-1} \Phi_L\right)^{-1} \geq \left(\Phi_L^T \mathcal{I}_{\Phi_L} \Phi_L\right)^{-1}$ and thus the GLS estimate has a larger uncertainty than the MLE (11).
Although all relevant information on the test statistic is contained in the scalar $\lambda_L$, calculating this quantity requires the inversion of a matrix with dimension $L \cdot m \times L \cdot m$ [15]:

$$\mathcal{I}_{E_L} = \left( \mathcal{O}_L \mathcal{I}_{x_k-L+1}^{-1} \mathcal{O}_L \right)^{-1} + \mathcal{H}_w^T \mathcal{I}_{w_L} \mathcal{H}_w^{-1} + \mathcal{I}_{V_L}^{-1}.$$

Due to the assumption of white noise processes it follows that $\mathcal{I}_{w_L}$ and $\mathcal{I}_{V_L}$ are block diagonal matrices with entries $\mathcal{I}_{w}^{-1}$ and $\mathcal{I}_{V}^{-1}$, 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 $\mathcal{I}_{E_L}$ for $L+1$ measurements relates to $\mathcal{I}_{E_L}$. This leads to a recursive expression for $\lambda_{L+1}$ where the increment $\lambda_{L+1}^\Delta$ added by the $(L+1)$st measurement is identified.

B. Recursive form

At first, (16) is reformulated in terms of the predicted state $\hat{x}_{k-L+1}$. Stacking these state predictions is denoted as $\hat{X}_L = [\hat{x}_{k-L+1} \ldots \hat{x}_L]^T$. Consequently, $\hat{X}_L$ is the stacked state prediction error. Moreover, define $H_L = I_L \otimes C$ as a block diagonal matrix of $L$ times $C$ from the measurement model (1b). Hence:

$$H_L \hat{X}_L = \mathcal{O}_L \hat{x}_{k-L+1} + H^T w_L .$$

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

$$\mathcal{I}_{E_L} = \left( H_L \mathcal{I}_{\hat{X}_L} H_L^T + \mathcal{I}_{V_L}^{-1} \right)^{-1} \quad \text{(18)}$$

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

$$M_{L+1} = \begin{bmatrix} M_{L+1,11} & M_{L+1,12} \\ M_{L+1,21} & M_{L+1,22} \end{bmatrix} \quad \text{(20)}$$

and $\lambda_{L+1}^\Delta$ is given by the quadratic form

$$\lambda_{L+1}^\Delta = (a - b)^T \left( \mathcal{I}_{V}^{-1} - C M_{L+1,22} C^T \right)^{-1} (a - b),$$

with the following abbreviations

$$\mathbf{Y} = \mathcal{D}_{11} - \mathcal{D}_{12} \left( \mathcal{D}_{22} + C^T \mathcal{I}_v C \right)^{-1} \mathcal{D}_{21},$$

and

$$\mathbf{D}_{11} = \left( \mathcal{D}_{22} + C^T \mathcal{I}_v C \right)^{-1} \mathcal{D}_{21} \left( \mathbf{Y} + M_{L,22}^{-1} \right)^{-1} M_{L,22}^{-1},$$

and

$$\mathbf{D}_{12} = - \left( \mathcal{D}_{22} + C^T \mathcal{I}_v C \right)^{-1} \mathcal{D}_{21} \left( \mathbf{Y} + M_{L,22}^{-1} \right)^{-1} M_{L,22}^{-1} \mathbf{D}_{22}.$$
C. Discussion of the recursive form

The recursion for the non-centrality parameter from theorem 1 provides the scalar increments \( \lambda_{L+1}^{\Delta} \) which increase \( \lambda \) with each additional measurement taken into account.

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

\[
P_D^\Delta \approx \frac{d}{d\lambda} \left( 1 - P_{\Delta}^z(\lambda) \left( Q_{\lambda}^z (1 - P_{FA}) \right) \right) \lambda_{L+1}^{\Delta}.
\]

(26)

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

There are numerous approximations of the non-central \( \chi^2 \)-distribution. We employ a compact variant by Patnaik [19]:

\[
P_{\Delta}^z(\lambda)(z) \approx P_N(\mu_z(\lambda),1)(\zeta(\lambda))
\]

(27)

with \( \zeta(\lambda) = \sqrt{2\pi} \cdot \mu_z(\lambda) = \sqrt{2(\mu z + \lambda)} - 1 \).

(28)

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

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

(29)

where \( \mu_z'(\cdot) \) 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 \( P_D \) \( dP_{FA} \). in the GLR test statistic (15).

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 km/h with an initial distance of 37 m to the intersection and it is assumed that the duration for the yellow phase is 3 s.

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>
<thead>
<tr>
<th>Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sampling time</td>
<td>( T = 0.0675 ) s</td>
</tr>
<tr>
<td>Initial state</td>
<td>( x_{k-L+1} = [10 \text{ m}, 0] )</td>
</tr>
<tr>
<td>Initial uncertainty</td>
<td>( \mathcal{I}<em>{x</em>{k-L+1}}^{-1} = \text{diag } 0.25 \text{ m, } 0.2 \text{ m/s}^2 )</td>
</tr>
<tr>
<td>Process noise</td>
<td>( \mathcal{I}_{w}^{-1} = (0.2 \text{ m/s}^2)^2 s^{-1} )</td>
</tr>
<tr>
<td>Measurement noise</td>
<td>( R = \mathcal{I}_v^{-1} = (0.2 \text{ m})^2 )</td>
</tr>
<tr>
<td>Brake ramp slope</td>
<td>( \Delta r_{\text{brake}} = -8 \text{ m/s}^2 )</td>
</tr>
<tr>
<td>Brake ramp up time</td>
<td>( t_{\text{ramp}} = 0.405 ) s</td>
</tr>
</tbody>
</table>

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:

\[
\begin{bmatrix}
x \\
v_x\end{bmatrix}_{k+1} =
\begin{bmatrix}
1 & T \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
v_x\end{bmatrix}_k + \begin{bmatrix}
0.5T^2 \\
T
\end{bmatrix} f_k + \begin{bmatrix}
w_k
\end{bmatrix}
\]

(30a)

\[
y_k = \begin{bmatrix}
1 & 0
\end{bmatrix} x_k + v_k.
\]

(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 \( \theta = \Delta r_{\text{brake}} \) and fixed ramp up time \( t_{\text{ramp}} \):

\[
f_k = \left[ t_k - (t_k - t_{\text{ramp}}) \sigma \left( t_k - t_{\text{ramp}} \right) \right] \Delta r_{\text{brake}}.
\]

(31)

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

It is assumed that a Gaussian estimate of the initial state with \( x_{k-L+1} = \mathcal{N} (x_{k-L+1}, \mathcal{I}_{x_{k-L+1}}^{-1}) \) 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 \( N_{\text{sim}} = 10^5 \) 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.

![Image of diagram](image.png)

The ego vehicle and target vehicle are shown with a distance of 47 meters. The ego vehicle is at 10 meters from the traffic light, while the target vehicle is at 37 meters. The traffic light is shown in yellow with a 50 km/h speed limit. The diagram illustrates the application example described in the text.

Fig. 3. Acceleration and velocity profiles.

(a) Acceleration profile

(b) Relative velocity

Fig. 4. Receiver Operating Characteristic curves.

Fig. 5. Detection probability \( P_D \) and increments \( \Delta P_D \) for false positive probability \( P_{FA} = 0.05 \).

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 $M_{L+1}$, a partitioning for $M_L$ is introduced:

$$M_L = \left( I_{X_L} + H_L^T I_{V_{L-1}} H_L \right)^{-1}$$

(32)

$$= \left( \begin{bmatrix} A & B \\ B^T & C \end{bmatrix} + \begin{bmatrix} H_{L-1}^T I_{V_{L-1}} H_{L-1} & 0 \\ 0 & C^T I_{V_{C}} \end{bmatrix} \right)^{-1}$$

(32)

Calculating the inverse block-wise gives:

$$M_{L,11} = \left( A + H_{L-1}^T I_{V_{L-1}} H_{L-1} - B \left( C + C^T I_{V_{C}} \right)^{-1} B^T \right)^{-1}$$

(33a)

$$M_{L,22} = \left( C + C^T I_{V_{C}} - B^T \left( A + H_{L-1}^T I_{V_{L-1}} H_{L-1} \right)^{-1} B \right)^{-1}$$

(33b)

$$M_{L,21} = -M_{L,22} B^T \left( A + H_{L-1}^T I_{V_{L-1}} H_{L-1} \right)^{-1}$$

(33c)

$$M_{L,12} = M_{L,21}$$

(33d)

The central idea is then to leverage the Markov assumption on the system (1a). This yields that $I_{X_{L+1}}$ has a block-diagonal form [9]:

$$I_{X_{L+1}} = \begin{bmatrix} A & B & 0 \\ B^T & C & \mathcal{D}_{21} \\ 0 & \mathcal{D}_{21} & \mathcal{D}_{22} \end{bmatrix}$$

(34)

and therefore

$$M_{L+1} = \left( I_{X_{L+1}} + H_{L+1}^T I_{V_{L+1}} H_{L+1} \right)^{-1}$$

(35)

can be calculated block-wise. The exact expressions for $\mathcal{D}_{11}, \mathcal{D}_{21}$ and $\mathcal{D}_{22}$ 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 $M_{L+1,11}$:

$$M_{L+1,11} = \left( M_L^{-1} + \begin{bmatrix} 0 \\ I_n \end{bmatrix} \mathcal{Y} \begin{bmatrix} 0 & I_n \end{bmatrix} \right)^{-1}$$

$$= M_L - \frac{M_{L,12}}{M_{L,22}} \mathcal{Y} \left( I_n + M_{L,22} \mathcal{Y} \right)^{-1} \left[ M_{L,21} \ M_{L,22} \right]$$

(36)

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

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

Next, the lower right element $M_{L+1,22}$ is studied:

$$M_{L+1,22} = \left( \mathcal{D}_{22} + C^T I_{V_{C}} - \mathcal{D}_{21} \left( C + \mathcal{D}_{11} + C^T I_{V_{C}} \right)^{-1} - B^T \left( A + H_{L-1}^T I_{V_{L-1}} H_{L-1} \right)^{-1} \mathcal{D}_{12} \right)^{-1}$$

(37)

Finally, $M_{L+1,21}$ (which can be shown to equal $M_{L+1,12}$):

$$M_{L+1,21} = -M_{L+1,22} \cdot \left[ 0 \mathcal{D}_{21} \right]$$

$$\cdot \left[ A + H_{L-1}^T I_{V_{L-1}} H_{L-1} \right]^T C + \mathcal{D}_{11} + C^T I_{V_{C}} \right)$$

(33b)

$$\cdot \left[ B^T \left( A + H_{L-1}^T I_{V_{L-1}} H_{L-1} \right)^{-1} I_n \right]$$

(33c)

$$\cdot \left( \mathcal{D}_{22} + C^T I_{V_{C}} - \mathcal{D}_{21} \left( \mathcal{D}_{11} + M_{L,22}^{-1} \right)^{-1} \mathcal{D}_{12} \right)^{-1}$$

(37)

$$\cdot \left( \mathcal{D}_{22} + C^T I_{V_{C}} \right)^{-1} \mathcal{D}_{21}$$

(37)

$$\cdot \left[ M_{L,21} \ M_{L,22} \right]$$

$$= \left[ \mathcal{Y} \right]$$

(38)

Therefore, $M_{L+1,21}$ consists of $L$ submatrices $M_{L+1,21}^{(i)}$ of size $n \times n$:

$$M_{L+1,21} = \left[ M_{L+1,21}^{(1)} \cdots M_{L+1,21}^{(L)} \right]$$

(39)

With (38), one obtains an explicit expression:

$$M_{L+1,21}^{(i)} = \prod_{j=1}^{L} \Gamma_{j+1} M_{j,22}$$

(40)

B. Proof of theorem 1

First note, that $\Phi_{L+1}$ as defined in (10) can be written as

$$\Phi_{L+1} = \left[ \Phi_L \ \phi_{L+1} \right]^T$$

Then, inserting the block-matrices which constitute $M_{L+1}$ from (21) into $I_{E_{L+1}}$ in (18) and
applying the multiplications with $\Phi_{L+1}^T \theta_1$ from (14b) yields:

$$
\lambda_{L+1} = \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} \left(\Phi_{L+1}^T \theta_1\right)
= \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} - \mathbf{I}_{V_L} H_L^T M_{L+1}^T \mathbf{I}_{V_L} \left(\Phi_{L+1}^T \theta_1\right)
= \lambda_{L} + \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} - \mathbf{I}_{V_L} H_L^T M_{L+1}^T \mathbf{I}_{V_L} \left(\Phi_{L+1}^T \theta_1\right)
= \lambda_{L} + \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} \mathbf{H}_L \mathbf{M}_{L+1,22} \mathbf{Y} \left(\mathbf{Y} + \mathbf{M}_{L+1,22}^{-1}\right)^{-1}
- \mathbf{I}_{V_L} \mathbf{H}_L^T \mathbf{M}_{L+1,22}^T \mathbf{I}_{V_L} \left(\Phi_{L+1}^T \theta_1\right)^T
= \lambda_{L} + \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} \mathbf{H}_L \mathbf{M}_{L+1,22} \mathbf{Y} \left(\mathbf{Y} + \mathbf{M}_{L+1,22}^{-1}\right)^{-1}
\mathbf{I}_{V_L}^T \mathbf{H}_L \mathbf{M}_{L+1,22}^T \mathbf{I}_{V_L} \left(\Phi_{L+1}^T \theta_1\right)^T
$$

The key to simplify $\lambda_{L+1}^\Delta$ is to study the product

$$
\mathbf{A}_{L+1} = \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} \mathbf{H}_L \mathbf{M}_{L+1,22} \mathbf{Y} \left(\mathbf{Y} + \mathbf{M}_{L+1,22}^{-1}\right)^{-1}
\mathbf{I}_{V_L}^T \mathbf{H}_L \mathbf{M}_{L+1,22}^T \mathbf{I}_{V_L} \left(\Phi_{L+1}^T \theta_1\right)^T
$$

Inserting (40) into (42) yields

$$
\mathbf{A}_{L+1}^\Delta = \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} \mathbf{H}_L \mathbf{M}_{L+1,22} \mathbf{Y} \left(\mathbf{Y} + \mathbf{M}_{L+1,22}^{-1}\right)^{-1}
\mathbf{I}_{V_L}^T \mathbf{H}_L \mathbf{M}_{L+1,22}^T \mathbf{I}_{V_L} \left(\Phi_{L+1}^T \theta_1\right)^T
$$

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

$$
\lambda_{L+1}^\Delta = \left(\mathbf{A}_{L} + \left(\Phi_{L+1}^T \theta_1\right)^T \mathbf{I}_{V_L} \mathbf{H}_L \mathbf{M}_{L+1,22} \mathbf{Y}
$$

References


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