

2025-11-05

The self-archived postprint version of this journal article is available at Linköping University Institutional Repository (DiVA):

<https://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-219021>

Matrix-Valued Measures and Wishart Statistics for Target Tracking Applications

Robin Forsling, Simon J. Julier and Gustaf Hendeby

IEEE Transactions on Aerospace and Electronic Systems, (2025), 61(5), pp. 12234-12244.

Publisher: Institute of Electrical and Electronics Engineers

<https://doi.org/10.1109/taes.2025.3571685>

N.B.: When citing this work, cite the original publication.

© 2025 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.

Matrix-Valued Measures and Wishart Statistics for Target Tracking Applications

Robin Forsling, Simon J. Julier, and Gustaf Hendeby

Abstract—Ensuring sufficiently accurate models is crucial in target tracking systems. If the assumed models deviate too much from the truth, the tracking performance might be severely degraded. While the models are usually defined using multivariate conditions, the measures used to validate them are most often scalar-valued. In this paper, we propose matrix-valued measures for both offline and online assessment of target tracking systems. Recent results from Wishart statistics, and approximations thereof, are adapted and it is shown how these can be incorporated to infer statistical properties for the eigenvalues of the proposed measures. In addition, we relate these results to the statistics of the baseline measures. Finally, the applicability of the proposed measures are demonstrated using two important problems in target tracking: (i) distributed track fusion design; and (ii) filter model mismatch detection.

Index Terms—Target tracking, data fusion, evaluation measures, model imperfections, model validation, Wishart statistics.

I. INTRODUCTION

Target tracking involves estimating the state of a target of interest using noisy sensor measurements. The standard paradigm is model-based target tracking, where sensor models and motion models are combined for tracking the target state over time [1]. It is essential for tracking performance that the assumed models are sufficiently correct. If the assumed models deviate too far from the actual underlying models, there is often an unpredictable degradation in the tracking performance.

Developing methodologies and measures for accurate model assessment is still an open challenge

This work was performed within the Competence Center SED-DIT, Sensor Informatics and Decision making for the Digital Transformation, supported by Swedens Innovation Agency within the research and innovation program Advanced digitalization.

Robin Forsling is with Advanced Programs, Saab AB, Linköping, Sweden, and also with the Department of Electrical Engineering, Linköping University, Linköping, Sweden (e-mail: robin.forsling@liu.se). Simon J. Julier is with the Department of Computer Science, University College London, London, UK. Gustaf Hendeby is with the Department of Electrical Engineering, Linköping University, Linköping, Sweden.

[2–4]. Model imperfection in target tracking is often evaluated using the *normalized estimation error squared* (NEES) and the *normalized innovation squared* (NIS) [5]. Both measures penalize the *mean squared error* (MSE) weighted by the computed covariance matrix. Hence, they are scale-invariant in contrast to the MSE. NEES requires the ground truth to be known and is therefore suitable for offline analyses. Since NIS can be computed both online and offline it is typically the preferred choice. However, while the models in general are multivariate, both NEES and NIS are scalar-valued. Hence, despite their widely spread usage in application areas such as navigation and target tracking, NEES and NIS cannot sufficiently address the multivariate relations. Moreover, as pointed out in [4], NEES and NIS often fail to be useful even for evaluating scalar relations.

In this paper we propose matrix generalizations of the NEES and NIS. In particular, by using the eigenvalues of these matrices, different multivariate properties and model imperfections can be examined. We further utilize recent results from Wishart statistics to facilitate the analysis of target tracking systems based on eigenvalue statistics. A few applications¹ are used to demonstrate the usage of the proposed matrix-valued measures and the implied statistics.

II. RELATED SCALAR-VALUED MEASURES

We start with the notation and mathematical preliminaries. Related measures are then reviewed.

A. Notation

Let \mathbb{R}^n be the set of all n -dimensional real-valued vectors. By $\mathbf{A} \succeq \mathbf{B}$ and $\mathbf{A} \succ \mathbf{B}$ we denote that the difference $\mathbf{A} - \mathbf{B}$ is positive semidefinite and positive definite, respectively. The identity matrix of applicable size is given by \mathbf{I} . The expected value and covariance of random vector \mathbf{z} are denoted $E(\mathbf{z})$ and $\text{cov}(\mathbf{z})$, respectively.

¹MATLAB® code for all developments and applications of this paper is available at: <https://github.com/robinforsling/dtt/>.

Let $\mathbf{x}_k \in \mathbb{R}^{n_x}$ be an n_x -dimensional state at time k to be estimated. An *estimate* of \mathbf{x}_k is given by the pair $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$, where $\hat{\mathbf{x}}_k$ is the state estimate and $\mathbf{P}_k \succ \mathbf{0}$ the covariance computed by the estimator for $\hat{\mathbf{x}}_k$. Similarly, $(\hat{\mathbf{x}}_k^i, \mathbf{P}_k^i)$ is the estimate computed in the i th sample or realization, e.g., in a *Monte Carlo* (MC) simulation. The estimation error is defined as $\tilde{\mathbf{x}}_k = \hat{\mathbf{x}}_k - \mathbf{x}_k$ and $\tilde{\mathbf{x}}_k^i = \hat{\mathbf{x}}_k^i - \mathbf{x}_k$ is the estimation error in the i th sample. It is assumed that $\hat{\mathbf{x}}_k$ is unbiased, i.e., $E(\tilde{\mathbf{x}}_k) = \mathbf{0}$. The matrix $\Sigma_k = \text{cov}(\tilde{\mathbf{x}}_k) = E(\tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k^T)$ is referred to as the MSE matrix or the true covariance of the estimation error. Note, we use the same notation for a random variable and a realization of it.

If $\mathbf{z} \sim \mathcal{N}_m(\boldsymbol{\mu}, \Sigma)$, then \mathbf{z} is a Gaussian distributed m -dimensional random vector, where $\boldsymbol{\mu} = E(\mathbf{z})$ and $\Sigma = \text{cov}(\mathbf{z})$. Moreover, if $\mathbf{z} \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I})$, then $\mathbf{z}^T \mathbf{z} \sim \chi_m^2$, where χ_m^2 denotes the central chi-squared distribution with m degrees of freedom. Let $\mathbf{Z} = [\mathbf{z}_1 \ \dots \ \mathbf{z}_n]$ be a $m \times n$ real-valued random matrix, where each column $\mathbf{z}_i \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I})$ is independent and identically distributed (i.i.d.). Then $\mathbf{Z}\mathbf{Z}^T \sim \mathcal{W}_m(n, \mathbf{I})$ is an $m \times m$ positive semidefinite matrix, where $\mathcal{W}_m(n, \mathbf{I})$ is the real *Wishart distribution* [6] with n degrees of freedom and covariance parameter \mathbf{I} . The Wishart distribution is the sampling distribution of covariance matrices where the underlying samples are i.i.d. Gaussian random vectors. It is hence relevant when computing sampled covariance matrices from a Gaussian distributed error.

B. Preliminaries

Two central concepts are *credibility* and *conservativeness*. An estimator of \mathbf{x}_k that computes $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ is *credible* at time k if $E(\tilde{\mathbf{x}}_k) = \mathbf{0}$ and²

$$\mathbf{P}_k = \Sigma_k. \quad (1)$$

An estimator of \mathbf{x}_k that computes $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ is *conservative* at time k if

$$\mathbf{P}_k \succeq \Sigma_k. \quad (2)$$

In, e.g., [7], the conservativeness criterion is relaxed using the trace operator. An estimator of \mathbf{x}_k that computes $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ is *trace-conservative* at time k if

$$\text{tr}(\mathbf{P}_k) \geq \text{tr}(\Sigma_k), \quad (3)$$

which is a weaker property than conservative [8].

Since $\mathbf{P}_k \succ \mathbf{0}$, it has a unique Cholesky factorization $\mathbf{P}_k = \mathbf{L}_k \mathbf{L}_k^T$, where \mathbf{L}_k is lower-triangular

²It should be pointed out that the probability that a sampled approximation, e.g., an MC estimator, of \mathbf{P}_k equals to Σ_k is zero.

and invertible. Moreover, an eigendecomposition of an $n \times n$ symmetric positive semidefinite matrix \mathbf{S} is given by

$$\mathbf{S} = \sum_{i=1}^n \lambda_i \mathbf{u}_i \mathbf{u}_i^T, \quad (4)$$

where $\lambda_i = \lambda_i(\mathbf{S}) \geq 0$ is the i th eigenvalue of \mathbf{S} and \mathbf{u}_i is the associated eigenvalue. Note, if unambiguous, for simplicity we use λ_i instead of $\lambda_i(\mathbf{S})$ for the i th eigenvalue of \mathbf{S} . It is assumed that

$$\lambda_{\max} = \lambda_1 \geq \dots \geq \lambda_n = \lambda_{\min}. \quad (5)$$

The condition in (1) is equivalent to

$$\mathbf{P}_k = \Sigma_k \iff \mathbf{I} = \mathbf{L}_k^{-1} \Sigma_k \mathbf{L}_k^{-T}. \quad (6)$$

Similarly, the condition in (2) is equivalent to

$$\begin{aligned} \mathbf{P}_k \succeq \Sigma_k &\iff \mathbf{I} \succeq \mathbf{L}_k^{-1} \Sigma_k \mathbf{L}_k^{-T} \\ &\iff 1 \geq \lambda_i(\mathbf{L}_k^{-1} \Sigma_k \mathbf{L}_k^{-T}), \forall i \\ &\iff 1 \geq \lambda_{\max}(\mathbf{L}_k^{-1} \Sigma_k \mathbf{L}_k^{-T}). \end{aligned} \quad (7)$$

C. Related Work

The NEES is introduced in [9] as a measure for the uncertainty assessment in target tracking algorithms. The NEES is computed as

$$\text{NEES}_k = \frac{1}{M} \sum_{i=1}^M (\tilde{\mathbf{x}}_k^i)^T (\mathbf{P}_k^i)^{-1} \tilde{\mathbf{x}}_k^i, \quad (8)$$

where M is the number of MC runs. The NIS is computed similarly. Let $\tilde{\mathbf{y}}_k$ be the innovation at time k , where $\tilde{\mathbf{y}}_k = \mathbf{y}_k - \hat{\mathbf{y}}_k$ is the difference between measurement \mathbf{y}_k and predicted measurement $\hat{\mathbf{y}}_k$ at time k , cf. a Kalman filter (KF, [10]). Let $\mathbf{S}_k = \mathbf{B}_k \mathbf{B}_k^T$ be the covariance computed for $\tilde{\mathbf{y}}_k$. Then the NIS at time k is computed as

$$\text{NIS}_k = \frac{1}{K} \sum_{l=k-K+1}^K \tilde{\mathbf{y}}_l^T \mathbf{S}_l^{-1} \tilde{\mathbf{y}}_l, \quad (9)$$

where K is the number of time steps used.

As pointed out in [11–13], NEES exhibits a few drawbacks: (i) it penalizes optimism and pessimism asymmetrically; and (ii) it is inconvenient for comparing different estimators' credibility. To overcome these drawbacks the same authors propose the *non-credibility index* (NCI) defined as

$$\text{NCI}_k = \frac{10}{M} \sum_{i=1}^M \log_{10} \left(\frac{(\tilde{\mathbf{x}}_k^i)^T (\mathbf{P}_k^i)^{-1} \tilde{\mathbf{x}}_k^i}{(\tilde{\mathbf{x}}_k^i)^T (\Sigma_k^i)^{-1} \tilde{\mathbf{x}}_k^i} \right), \quad (10)$$

where Σ_k^i is the MSE matrix of i th MC run. If unknown, Σ_k^i is approximated by

$$\hat{\Sigma}_k = \frac{1}{M} \sum_{i=1}^M \tilde{\mathbf{x}}_k^i (\tilde{\mathbf{x}}_k^i)^\top. \quad (11)$$

In [5] NEES and NIS are used to evaluate *filter consistency*. A consistent filter ensures two important properties [14]: (i) the error statistics computed by the filter is the same as the true error statistics; and (ii) the filter mixes information obtained from the process with measurement information in an optimal way. To this end, a recent paper [4] suggests extending NEES and NIS by including terms for the second-order moments. This works remarkably well for filter tuning and it has been shown that the extended measures can be integrated into an automatic filter tuning framework [4].

The measures mentioned so far—including NEES, NIS, and NCI—are all scalar measures. However, for an estimator to be credible or conservative, certain matrix conditions must be fulfilled. Satisfying the scalar conditions is not sufficient to ensure that the matrix conditions are satisfied. For instance, for a strict evaluation of conservativeness it is necessary to consider semidefinite conditions, cf. (2). This aspect is briefly addressed in [12], where the *credibility interval* is defined as³

$$[\lambda_{\min}(\Xi_k), \lambda_{\max}(\Xi_k)], \quad (12)$$

with

$$\Xi_k = \mathbf{L}_k^{-1} \Sigma_k \mathbf{L}_k^{-\top}. \quad (13)$$

To compute Ξ_k , both Σ_k and $\mathbf{P}_k = \mathbf{L}_k \mathbf{L}_k^\top$ must be known. A workaround is to approximate these covariances. An unknown covariance $\Sigma_k = \text{cov}(\tilde{\mathbf{x}}_k)$ can be approximated by $\hat{\Sigma}_k$ as defined in (11). Similarly, \mathbf{P}_k can be approximated by the mean

$$\hat{\mathbf{P}}_k = \frac{1}{M} \sum_{i=1}^M \mathbf{P}_k^i. \quad (14)$$

If the system is linear, then $\mathbf{P}_k^i = \mathbf{P}_k$ for all i and hence $\hat{\mathbf{P}}_k = \mathbf{P}_k$.

In [15], the *conservativeness index* (COIN) is defined by

$$\text{COIN}_k = \lambda_{\max}(\mathbf{L}_k^{-1} \hat{\Sigma}_k \mathbf{L}_k^{-\top}). \quad (15)$$

The next proposition⁴ is a direct consequence of (7).

³For a perfectly credible estimator the credibility interval would reduce to the single value 1.

⁴If $\hat{\Sigma} \neq \Sigma$, then this is only an approximation. In [15] it is assumed that $\mathbf{P}_k^i = \mathbf{P}_k$. If not, \mathbf{P}_k can be approximated by $\hat{\mathbf{P}}_k$.

Proposition 1. If $\hat{\Sigma}_k = \Sigma_k$, then $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$ is conservative if and only if $\text{COIN}_k \leq 1$.

Proof. An estimate $(\hat{\mathbf{x}}_k, \mathbf{P}_k)$, where $\mathbf{P}_k = \mathbf{L}_k \mathbf{L}_k^\top$, is conservative if $\mathbf{P}_k \succeq \Sigma_k$. Hence, if $\hat{\Sigma}_k = \Sigma_k$, it follows from (7) and by definition of COIN_k , that

$$\mathbf{P}_k \succeq \hat{\Sigma}_k \iff 1 \geq \lambda_{\max}(\mathbf{L}_k^{-1} \hat{\Sigma}_k \mathbf{L}_k^{-\top}) = \text{COIN}_k. \quad \square$$

III. DEVELOPING STATISTICAL MEASURES FOR EVALUATING MATRIX-VALUED PROPERTIES

We start with a motivating example to illustrate that scalar-valued measures such as NEES in general fail to evaluate matrix-valued conditions.

A. Motivating Example

It is now illustrated how merely looking at NEES might lead to the conclusion that an estimator is credible or conservative when it in fact is neither. Let

$$\Sigma = \begin{bmatrix} 8 & 1 \\ 1 & 2 \end{bmatrix}, \quad \mathbf{P} = \begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}.$$

Clearly $\Sigma \neq \mathbf{P}$, i.e., the credibility condition is violated. Moreover,

$$\Xi = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix},$$

with eigenvalues $\lambda_{\min} = 0.75$ and $\lambda_{\max} = 1.25$. Hence, neither the conservativeness condition holds.

Consider now a stationary setting, where the estimation error $\tilde{\mathbf{x}}_k \sim \mathcal{N}_2(\mathbf{0}, \Sigma)$ and $\mathbf{P} = \mathbf{L}\mathbf{L}^\top$ is the covariance computed for $\tilde{\mathbf{x}}_k$ at each k . If we sample $\tilde{\mathbf{x}}_k \sim \mathcal{N}_2(\mathbf{0}, \Sigma)$ over independent MC runs, the NEES statistics can be computed using (8). For a filter consistent estimator we should have $\text{NEES}_k = n_x = 2$ and that NEES is χ^2 distributed. The NEES statistics is plotted in Fig. 1 and we see that $\text{NEES}_k/2$ is very close to 1. In addition, the sampled probability density function (PDF) of $\|\mathbf{L}^{-1}\tilde{\mathbf{x}}\|^2$ is computed as a histogram and plotted against the theoretical χ^2 PDF. By pure inspection, the NEES statistics is what we would expect for a filter consistent estimator. However, since $\lambda_{\min}(\Xi) = 0.75$ and $\lambda_{\max}(\Xi) = 1.25$ we know that this estimator is not credible nor even conservative.

B. Problem Formulation

In this paper, the objective is to develop measures that can be used to evaluate matrix-valued conditions, e.g., credibility and conservativeness. This means that we do not only need new measures, but also statistical properties related to these measures.

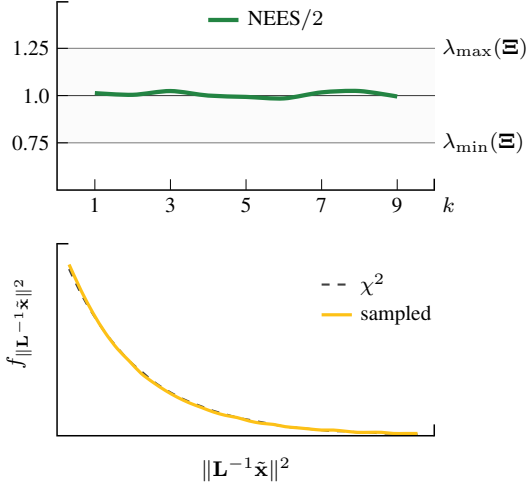


Fig. 1. Motivating example. The estimation error is sampled from $\mathcal{N}_2(\mathbf{0}, \Sigma)$ and $\mathbf{P} \neq \Sigma$ is the covariance computed by the estimator. However, from the NEES it might be concluded that $\mathbf{P} = \Sigma$ despite that $\lambda_{\min}(\Xi) = 0.75$ and $\lambda_{\max}(\Xi) = 1.25$.

IV. PROPOSED MATRIX-VALUED STATISTICS

In this section the proposed matrix-valued statistics are presented. We start with a matrix generalization of NEES which is suitable for offline evaluation of target tracking and data fusion systems. Then a similar generalization of NIS is proposed that can be used in online applications.

A. The Normalized Estimation Error Squared Matrix

The motivating example in the preceding section illustrates that NEES is not sufficient to evaluate credibility and conservativeness. However, still, the normalized error and innovation are useful tools which we want to generalize to the multivariate case. The *NEES matrix* is proposed below in Definition 1. It is defined for MC based simulations and interpreted as the sampled Ξ , i.e., the sampled covariance of the normalized estimation error $\mathbf{L}^{-1}\tilde{\mathbf{x}}$.

Definition 1 (The NEES Matrix). Let $\tilde{\mathbf{x}}_k^i$ be the estimation error of the i th sample at time k . Let $\mathbf{P}_k^i = \mathbf{L}_k^i(\mathbf{L}_k^i)^\top \succ \mathbf{0}$ be the covariance computed by the estimator. The *NEES matrix* is defined as

$$\hat{\Xi}_k = \frac{1}{M} \sum_{i=1}^M (\mathbf{L}_k^i)^{-1} \tilde{\mathbf{x}}_k^i (\tilde{\mathbf{x}}_k^i)^\top (\mathbf{L}_k^i)^{-\top}. \quad (16)$$

Note that

$$\begin{aligned} \text{tr}(\hat{\Xi}_k) &= \text{tr} \left(\frac{1}{M} \sum_{i=1}^M (\mathbf{L}_k^i)^{-1} \tilde{\mathbf{x}}_k^i (\tilde{\mathbf{x}}_k^i)^\top (\mathbf{L}_k^i)^{-\top} \right) \\ &= \frac{1}{M} \sum_{i=1}^M \text{tr} \left((\tilde{\mathbf{x}}_k^i)^\top (\mathbf{L}_k^i)^{-\top} (\mathbf{L}_k^i)^{-1} \tilde{\mathbf{x}}_k^i \right) \\ &= \frac{1}{M} \sum_{i=1}^M (\tilde{\mathbf{x}}_k^i)^\top (\mathbf{P}_k^i)^{-1} \tilde{\mathbf{x}}_k^i = \text{NEES}_k, \end{aligned}$$

where it is utilized that $(\mathbf{L}_k^i)^{-\top} (\mathbf{L}_k^i)^{-1} = (\mathbf{P}_k^i)^{-1}$. Let $\bar{\lambda}(\mathbf{A})$ denote the average of the eigenvalues of \mathbf{A} . If we normalize NEES using n_x

$$\frac{1}{n_x} \text{tr}(\hat{\Xi}_k) = \frac{1}{n_x} \sum_{i=1}^{n_x} \lambda_i(\hat{\Xi}_k) = \bar{\lambda}(\hat{\Xi}_k), \quad (17)$$

which is also referred to as the *average NEES* [5].

B. The Normalized Innovation Squared Matrix

The NEES statistic requires knowledge about the true error and is hence used in offline applications where a large number of independent MC runs are simulated for a particular problem. A statistic that can be computed online, and for single runs, is the NIS statistics defined in (9). In a single run evaluation we average over subsequent time steps instead of MC runs. The NIS matrix is proposed in Definition 2. It is interpreted as the sampled covariance of the normalized innovation $\mathbf{B}^{-1}\tilde{\mathbf{y}}$.

Definition 2 (The NIS Matrix—Single Run Statistics). Let $\tilde{\mathbf{y}}_k \in \mathbb{R}^{n_y}$ be the innovation at time k . Let $\mathbf{S}_k = \mathbf{B}_k \mathbf{B}_k^\top \succ \mathbf{0}$ be the covariance computed for $\tilde{\mathbf{y}}_k$. The single run *NIS matrix* is defined as

$$\hat{\Pi}_k = \frac{1}{K} \sum_{l=k-K+1}^k \mathbf{B}_l^{-1} \tilde{\mathbf{y}}_l \tilde{\mathbf{y}}_l^\top \mathbf{B}_l^{-\top}. \quad (18)$$

Analogously to the NEES case, we have that

$$\begin{aligned} \text{tr}(\hat{\Pi}_k) &= \text{tr} \left(\frac{1}{K} \sum_{l=k-K+1}^k \mathbf{B}_l^{-1} \tilde{\mathbf{y}}_l \tilde{\mathbf{y}}_l^\top \mathbf{B}_l^{-\top} \right) \\ &= \frac{1}{K} \sum_{l=k-K+1}^k \text{tr} \left(\tilde{\mathbf{y}}_l^\top \mathbf{B}_l^{-\top} \mathbf{B}_l^{-1} \tilde{\mathbf{y}}_l \right) \\ &= \frac{1}{K} \sum_{l=k-K+1}^k \tilde{\mathbf{y}}_l^\top \mathbf{S}_l^{-1} \tilde{\mathbf{y}}_l = \text{NIS}_k, \end{aligned}$$

where it is utilized that $\mathbf{S}_l^{-1} = \mathbf{B}_l^{-\top} \mathbf{B}_l^{-1}$. If we divide NIS by n_y

$$\frac{1}{n_y} \text{tr}(\hat{\Pi}_k) = \frac{1}{n_y} \sum_{i=1}^{n_y} \lambda_i(\hat{\Pi}_k) = \bar{\lambda}(\hat{\Pi}_k). \quad (19)$$

In various applications, e.g., filter tuning, it is relevant to compute also the NIS in an MC setup. In this case the NIS matrix is defined as follows.

Definition 3 (The NIS Matrix—MC Statistics). Let $\tilde{\mathbf{y}}_k^i$ be the innovation of the i th sample at time k . Let $\mathbf{S}_k^i = \mathbf{B}_k^i (\mathbf{B}_k^i)^\top \succ \mathbf{0}$ be the covariance computed for $\tilde{\mathbf{y}}_k^i$. The MC based NIS matrix is defined as

$$\hat{\mathbf{\Pi}}'_k = \frac{1}{M} \sum_{i=1}^M (\mathbf{B}_k^i)^{-1} \tilde{\mathbf{y}}_k^i (\tilde{\mathbf{y}}_k^i)^\top (\mathbf{B}_k^i)^{-\top}. \quad (20)$$

C. Test Statistics

The NEES matrix and the NIS matrix are statistics suitable for testing matrix relationships such as credibility, cf. (1), and conservativeness, cf. (2). In particular, we are interested in the probability distributions of the smallest and largest eigenvalues of $\hat{\mathbf{\Xi}}$ and $\hat{\mathbf{\Pi}}$. With such distributions it is, for instance, possible to evaluate if the innovations are samples from a white process with $\mathbf{B}_k^{-1} \tilde{\mathbf{y}}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. The *null hypothesis*, in this case, is formulated as

$$\mathcal{H}_0: \mathbf{\Pi}_k = \mathbf{I}, \quad (21)$$

which is accepted or rejected using $\lambda_{\min}(\hat{\mathbf{\Pi}}_k)$ and $\lambda_{\max}(\hat{\mathbf{\Pi}}_k)$ as test statistics. In the next section we study the marginal distributions of the smallest and largest eigenvalues of Wishart distributed matrices.

Remark 1. In this paper we mainly focus on the problem of identifying *if* something is wrong, e.g., model errors, rather than pointing out *how* it is wrong, e.g., which components the model errors affect. This means that we are mainly interested in the eigenvalue statistics. However, in one of the applications we briefly analyze the corresponding eigenvectors which contain information about in which components the models errors contribute.

V. WISHART EIGENVALUE STATISTICS

If $\mathbf{L}^{-1} \tilde{\mathbf{x}} \sim \mathcal{N}_{n_x}(\mathbf{0}, \mathbf{I})$ and $\mathbf{B}^{-1} \tilde{\mathbf{y}} \sim \mathcal{N}_{n_y}(\mathbf{0}, \mathbf{I})$, then $\hat{\mathbf{\Xi}} \sim \mathcal{W}_{n_x}(M, \mathbf{I})$ and $\hat{\mathbf{\Pi}} \sim \mathcal{W}_{n_y}(K, \mathbf{I})$. Hence, we can utilize Wishart statistics to draw conclusions about, credibility, filter consistency, and the models used in a target tracking system.

In this section statistical properties of $\lambda_{\min}(\mathbf{V})$ and $\lambda_{\max}(\mathbf{V})$ are analyzed, where $\mathbf{V} \sim \mathcal{W}_m(n, \mathbf{I})$. It is assumed that $n \geq m$ such that $\mathbf{V} \succ \mathbf{0}$.

A. Joint Probability Distribution

Let $\Gamma(z)$ denote the gamma function, $\gamma(z, a, b) = \int_a^b t^{z-1} \exp(-t) dt$ be the generalized incomplete gamma function, and $r(z, a, b) = \frac{1}{\Gamma(a)} \gamma(z, a, b)$ denote the generalized regularized incomplete gamma

function. Define $\Gamma_m(z) = \pi^{m(m-1)/4} \prod_{i=1}^m \Gamma(z - (i-1)/2)$ and $g(z, t) = t^z \exp(-t)$.

Let $\mathbf{V} \sim \mathcal{W}_m(n, \mathbf{I})$ and $\lambda_{\max} = \lambda_1 \geq \dots \geq \lambda_m = \lambda_{\min}$ be the ordered eigenvalues of \mathbf{V} . The joint PDF of $\boldsymbol{\lambda} = [\lambda_1 \dots \lambda_m]$ is given by [16, 17]

$$f_{\boldsymbol{\lambda}}(\xi_1, \dots, \xi_m) = K_J \prod_{i=1}^m \exp(-\xi_i/2) \xi_i^\alpha \prod_{i < j}^m (\xi_i - \xi_j), \quad (22)$$

where $\alpha = (n - m - 1)/2$, $\xi_1 \geq \dots \geq \xi_m$, and the normalization constant K_J is given by

$$K_J = \frac{\pi^{m^2/2}}{2^{mn/2} \Gamma_m(m/2) \Gamma_m(n/2)}. \quad (23)$$

B. Exact Marginal Probability Distributions

The exact probability that all eigenvalues of $\mathbf{V} \sim \mathcal{W}_m(n, \mathbf{I})$ lie within an arbitrary interval is developed in [18]. The cumulative density functions (CDFs) of the smallest and largest eigenvalues of \mathbf{V} are then obtained as special cases⁵.

The probability that all eigenvalues of $\mathbf{V} \sim \mathcal{W}_m(n, \mathbf{I})$ lie within an interval $[a, b] \subseteq [0, \infty)$ is [18]

$$\begin{aligned} \psi(a, b) &= \Pr(a \leq \lambda_{\min}(\mathbf{V}), \lambda_{\max}(\mathbf{V}) \leq b) \\ &= K_{\lambda} \sqrt{\det(\mathbf{A}(a, b))}, \end{aligned} \quad (24)$$

where

$$\begin{aligned} K_{\lambda} &= K_J 2^{\alpha m + m(m+1)/2} \prod_{i=1}^m \Gamma(\alpha + i) \\ &= \frac{\pi^{m^2/2}}{\Gamma_m(m/2) \Gamma_m(n/2)} \prod_{i=1}^m \Gamma(\alpha + i), \end{aligned} \quad (25)$$

and $\mathbf{A}(a, b)$ is a skew symmetric matrix. A recursive formula for $\psi(a, b)$ is provided in Algorithm 1.

With $\psi(a, b)$, the CDFs $F_{\lambda_{\min}}$ and $F_{\lambda_{\max}}$ for the smallest and largest eigenvalues of \mathbf{V} , respectively, are given by

$$F_{\lambda_{\min}}(a) = \Pr(\lambda_{\min} \leq a) = 1 - \psi(a, \infty), \quad (26a)$$

$$F_{\lambda_{\max}}(b) = \Pr(\lambda_{\max} \leq b) = \psi(0, b). \quad (26b)$$

since ψ has a positive support.

⁵Pioneering work on the marginalization of the extreme eigenvalues of Wishart distributed matrices are found in [16, 19–21].

Algorithm 1 Probability $\psi(a, b)$ that all eigenvalues of $\mathbf{V} \sim \mathcal{W}_m(n, \mathbf{I})$ lie within $[a, b]$ [18]

Input: $m, n, a,$ and b

$\mathbf{A} = \mathbf{0}_{m \times m}$ $\triangleright m \times m$ matrix of zeros
 $\alpha_\ell = \alpha + \ell$ $\triangleright \ell$ is an integer

$K_\lambda = \frac{\pi^{m^2/2}}{\Gamma_m(m/2)\Gamma_m(n/2)} \prod_{i=1}^m \Gamma(\alpha + i)$

for $i = 1, \dots, m-1$ **do**

for $j = i, \dots, m-1$ **do**

$$\begin{aligned} [\mathbf{A}]_{i,j+1} &= [\mathbf{A}]_{i,j} \\ &+ \frac{2^{1-\alpha_i-\alpha_j} \Gamma(\alpha_i + \alpha_j)}{\Gamma(\alpha_j + 1) \Gamma(\alpha_i)} r(\alpha_i + \alpha_j, a, b) \\ &- \frac{g(\alpha_j, a/2) + g(\alpha_j, b/2)}{\Gamma(\alpha_j + 1)} r(\alpha_i, a/2, b/2) \end{aligned}$$

if m is odd **then**

$\mathbf{c} = \mathbf{0}_{m \times 1}$ \triangleright column vector of zeros

for $i = 1, \dots, m$ **do**

$$[\mathbf{c}]_i = r(\alpha_i, a/2, b/2)$$

$$\mathbf{A} \leftarrow \begin{bmatrix} \mathbf{A} & \mathbf{c} \\ \mathbf{0}_{1 \times m} & 0 \end{bmatrix}$$

$\mathbf{A} \leftarrow \mathbf{A} - \mathbf{A}^\top$

Output: $\psi(a, b) = K_\lambda \sqrt{\det(\mathbf{A})}$

C. Approximate Marginal Probability Distributions

The CDFs $F_{\lambda_{\min}}$ and $F_{\lambda_{\max}}$ computed using Algorithm 1 are exact. However, for large m, n the asymptotic behavior is often sufficient. In addition, numerical issues might arise when m and n (or n alone) are large⁶. In these situations approximate CDFs are useful. It is known that the smallest and largest eigenvalues converges to a shifted *Tracy-Widom* distribution as $m, n \rightarrow \infty$ [18, 23]. Here, we will use the simpler approximations proposed in [18, 22] which are based upon shifted gamma distributions.

Let $\mathbf{V} \sim \mathcal{W}_m(n, \mathbf{I})$ and let $r(z, a)$ be the lower regularized gamma function. Moreover, let μ_1, σ_1^2 , and s_1 be the mean, variance, and skewness of the Tracy-Widom distribution⁷ of type 1. Define

$$\kappa = \frac{4}{s_1^2}, \quad \theta = \frac{\sigma_1 s_1}{2}, \quad \rho = \kappa \theta - \mu_1. \quad (27)$$

The CDF $F_{\lambda_{\min}}$ is approximated using the result from [18]

$$\Pr(\lambda_{\min}(\mathbf{V}) \leq a) \approx r\left(\kappa, \frac{\max(0, -a' + \rho)}{\theta}\right), \quad (28)$$

⁶ m, n on the order of ≥ 100 are large in this context [22].

⁷For details about these parameters, see, e.g., [23].

where

$$a' = \frac{a - \mu_{\min}}{\sigma_{\min}}, \quad (29a)$$

$$\mu_{\min} = (\sqrt{n + c_n} - \sqrt{m + c_m})^2, \quad (29b)$$

$$\sigma_{\min} = \sqrt{\mu_{\min}} \left(\frac{1}{\sqrt{m + c_m}} - \frac{1}{\sqrt{n + c_n}} \right)^{\frac{1}{3}}, \quad (29c)$$

and where c_m and c_n are tuning parameters, here set to $c_m = c_n = -1/2$ following [22].

Similarly, $F_{\lambda_{\max}}$ is approximated using [22]

$$\Pr(\lambda_{\max}(\mathbf{V}) \leq b) \approx r\left(\kappa, \frac{\max(0, b' + \rho)}{\theta}\right), \quad (30)$$

where

$$b' = \frac{b - \mu_{\max}}{\sigma_{\max}}, \quad (31a)$$

$$\mu_{\max} = (\sqrt{m + c_m} + \sqrt{n + c_n})^2, \quad (31b)$$

$$\sigma_{\max} = \sqrt{\mu_{\max}} \left(\frac{1}{\sqrt{m + c_m}} + \frac{1}{\sqrt{n + c_n}} \right)^{\frac{1}{3}}. \quad (31c)$$

D. Expected Values

Assume that z is a random variable with nonnegative support and CDF $F_z(\zeta)$. Then [24]

$$\mathbb{E}(z) = \int_0^\infty (1 - F_z(\zeta)) d\zeta. \quad (32)$$

Hence, since both λ_{\min} and λ_{\max} have positive support only, their expected values are given by

$$\mathbb{E}(\lambda_{\min}) = \int_0^\infty (1 - F_{\lambda_{\min}}(\xi)) d\xi, \quad (33a)$$

$$\mathbb{E}(\lambda_{\max}) = \int_0^\infty (1 - F_{\lambda_{\max}}(\xi)) d\xi. \quad (33b)$$

Let $\mathbf{V} \sim \mathcal{W}_m(n, \mathbf{I})$. The expected values of $\lambda_{\min}(\mathbf{V})$ and $\lambda_{\max}(\mathbf{V})$ are plotted in Fig. 2 for $m = 3$ and different values of n . The inverse CDFs $F_{\lambda_{\min}}^{-1}$ and $F_{\lambda_{\max}}^{-1}$ are also plotted, corresponding to one-sided 95% confidence intervals for λ_{\min} and λ_{\max} , respectively. All curves are normalized with n . The curves approaches 1 as n tends to infinity.

E. Relation to χ^2 Statistics

The Wishart distribution is a multivariate generalization of the χ^2 distribution. Let $\mathbf{Z} =$

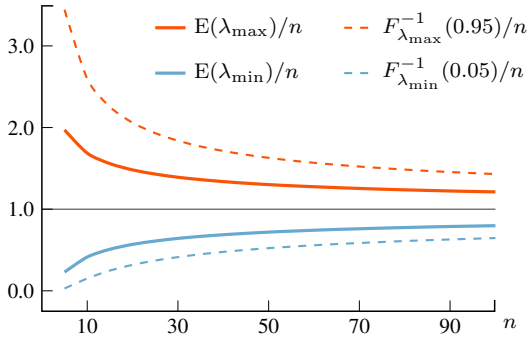


Fig. 2. Expected values and inverse CDFs of λ_{\min} and λ_{\max} as functions of n . The curves are normalized with n .

$[\mathbf{z}_1 \ \dots \ \mathbf{z}_n]$, where $\mathbf{z}_i \sim \mathcal{N}_m(\mathbf{0}, \mathbf{I})$ are i.i.d.. Then $\mathbf{Z}\mathbf{Z}^T \sim \mathcal{W}_m(n, \mathbf{I})$ and

$$\begin{aligned} \text{tr}(\mathbf{Z}\mathbf{Z}^T) &= \text{tr}\left(\sum_{i=1}^n \mathbf{z}_i \mathbf{z}_i^T\right) = \sum_{i=1}^n \text{tr}(\mathbf{z}_i \mathbf{z}_i^T) \\ &= \left(\sum_{i=1}^n \mathbf{z}_i^T \mathbf{z}_i\right) \sim \chi_{mn}^2. \end{aligned}$$

Hence, the χ^2 statistics is closely related to the Wishart statistics. However, it is not possible to reconstruct the Wishart statistics from the χ^2 statistics. Still, the χ^2 statistics is relevant when it comes to the evaluation of scalar properties such as trace-conservativeness, cf. (3), in which case the scalar NEES can be used.

F. Example: Switching Target Dynamics

We will now use a target tracking example to demonstrate the proposed statistics and compare them to their scalar analogs. A target is first tracked using correct models of the dynamics using a KF. After a certain time, the target dynamics change without changing the models in the KF. The goal is to be able to detect these changes. To this end, we analyze the NEES matrix and the NIS matrix before and after the change in the target dynamics.

Assume two spatial dimensions and let T_k be the sampling time. The target state x_k evolves according to a discrete time (nearly) *constant velocity* (CV) model

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{G}_k \mathbf{w}_k, \quad (34)$$

where $\mathbf{w}_k \sim \mathcal{N}_2(\mathbf{0}, \mathbf{Q}_k)$ is the process noise, \mathbf{Q}_k the process noise covariance, and⁸

$$\mathbf{F}_k = \begin{bmatrix} 1 & 0 & T_k & 0 \\ 0 & 1 & 0 & T_k \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{G}_k = \begin{bmatrix} \frac{T_k^2}{2} & 0 \\ 0 & \frac{T_k^2}{2} \\ T_k & 0 \\ 0 & T_k \end{bmatrix}. \quad (35)$$

A measurement $\mathbf{y}_k \in \mathbb{R}^2$ at time k is given according to the linear measurement model

$$\mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \mathbf{x}_k + \mathbf{v}_k, \quad (36)$$

where $\mathbf{v}_k \sim \mathcal{N}_2(\mathbf{0}, \mathbf{R}_k)$ is the measurement noise and $\mathbf{R}_k = \sigma_v^2 \mathbf{I}$ the measurement noise covariance.

We will simulate \mathbf{x}_k according to (34) using a \mathbf{Q}_k that switches at time k_{switch} . Let \mathbf{u}_k^{\parallel} and \mathbf{u}_k^{\perp} be two-dimensional unit vectors, where \mathbf{u}_k^{\parallel} is longitudinal and \mathbf{u}_k^{\perp} is lateral to the target velocity at time k . At $k = 1, \dots, k_{\text{switch}}$

$$\mathbf{Q}_k = q^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad (37)$$

and at $k = k_{\text{switch}} + 1, \dots, 20$

$$\mathbf{Q}_k = q^2 \begin{bmatrix} \mathbf{u}_k^{\parallel} & \mathbf{u}_k^{\perp} \\ \mathbf{u}_k^{\perp} & \mathbf{u}_k^{\parallel} \end{bmatrix} \begin{bmatrix} 10^{-6} & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{u}_k^{\parallel} & \mathbf{u}_k^{\perp} \end{bmatrix}^T, \quad (38)$$

where q is the magnitude of the random acceleration. The target hence first evolves according to isotropic random accelerations and then according to nearly central random accelerations.

The problem is evaluated using MC simulations. We compute the NEES matrix $\hat{\Xi}$ using Definition 1 and the NIS matrix $\hat{\Pi}'$ using Definition 3. Both are averaged over the MC simulations for each time k . The results are summarized in Fig. 3. It is seen that λ_{\min} and λ_{\max} respond very quickly when the switch occurs. Since the change in dynamics is such that the acceleration decreases in the longitudinal component and increases in the lateral component, λ_{\min} falls and λ_{\max} rises. This is true for both the NEES and NIS statistics. However, $\bar{\lambda}$ is approximately the same after the k_{switch} . Hence, it would be difficult to observe the change by merely looking at the scalar-valued NEES and NIS. Confidence intervals derived from the inverse CDFs are also included. By $F_{\chi_{Mm}^2}^{-1}$ we denote the inverse CDF of the χ^2 distribution with Mm degrees of freedom.

⁸This corresponds to a sample-and-hold model.

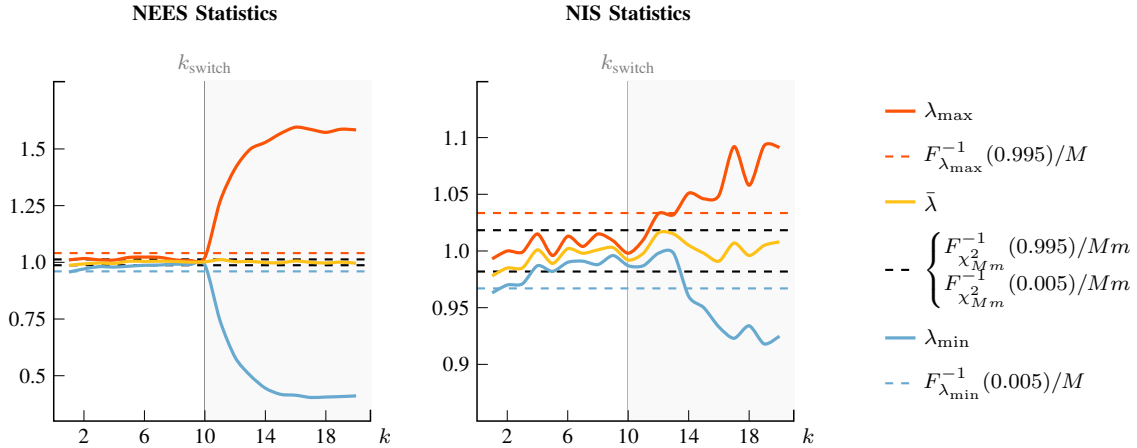


Fig. 3. Switching dynamics example. At k_{switch} , the target dynamics switches from a CV model with isotropic random accelerations to a CV model with central random accelerations. The switch is captured by λ_{\min} and λ_{\max} for both $\hat{\Xi}$ and $\hat{\Pi}'$, but not clearly by $\bar{\lambda}$. The confidence intervals given by F^{-1} have been normalized for comparison reason.

VI. TARGET TRACKING APPLICATIONS

In this section we demonstrate two important applications for the proposed matrix-valued measures. In the first application we consider track fusion design where the task is to choose a track fusion method *offline* evaluation of the NEES matrix. In the second application we use the NIS matrix for *online* detection of process model mismatch. In both applications the eigenvalues of NEES/NIS matrix are evaluated by utilizing the Wishart statistics presented in the previous section.

MATLAB[®] source code for the applications is available at <https://github.com/robinforstling/dtt/>. The repository also contains the functionality described in the previous section.

A. Distributed Track Fusion Design

Track fusion is a type of data fusion. It is an integral part of network-centric target tracking systems⁹ where multiple agents track overlapping sets of targets [15]. The goal with this example is to illustrate how the NEES matrix statistics is used to evaluate conservativeness.

1) *Scenario and Models:* Assume a target tracking scenario where two agents track a common target in two spatial dimensions. The target state x_k is assumed to evolve according to the CV model

$$\mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}_4(\mathbf{0}, \mathbf{Q}_k), \quad (39)$$

⁹For instance, target tracking in distributed sensor networks.

where \mathbf{F}_k is given in (35) and

$$\mathbf{Q}_k = q^2 \begin{bmatrix} \frac{T_k^3}{3} & 0 & \frac{T_k^2}{2} & 0 \\ 0 & \frac{T_k^3}{3} & 0 & \frac{T_k^2}{2} \\ \frac{T_k^2}{2} & 0 & T_k & 0 \\ 0 & \frac{T_k^2}{2} & 0 & T_k \end{bmatrix}, \quad (40)$$

with $T_k = 1$. At each time k the agents filters their local measurements using a KF. A nonlinear measurement model is assumed for both agents, where a measurement \mathbf{y}_k^i in Agent i at time k is generated according to

$$\mathbf{y}_k^i = \mathbf{h}(\mathbf{x}_k, \mathbf{s}_k^i) + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}_2(\mathbf{0}, \mathbf{R}_k), \quad (41)$$

where \mathbf{s}_k^i is the position of Agent i , $\mathbf{h}(\cdot)$ is a mapping from Cartesian to polar coordinates with origin in \mathbf{s}_k^i , and $\mathbf{R}_k = \text{diag}(\sigma_r^2, \sigma_\phi^2)$ with σ_r^2 and σ_ϕ^2 denoting the variances of the radial and azimuthal error, respectively. At odd k Agent 1 shares its local track with Agent 2 who fuses the tracks. At even k a local estimates is shared in the opposite direction for track fusion.

Assume now that a local extended Kalman filter (EKF, [25]), which only uses local measurements and no track fusion, has already been tuned in a satisfactory way for this particular problem. To utilize a received track a track fusion method is needed. The task is here to select a track fusion method based on performance and uncertainty assessment obtained using an MC study. For this simple example we consider *covariance intersection* (CI, [26]) and the *largest ellipsoid* (LE, [27]) method as the two candidate track fusion methods. For implementation

TABLE I
TRACK FUSION DESIGN PARAMETERS

Parameter	Description
$M = 10\,000$	number of MC runs
$q = 5$	process noise parameter [$\text{ms}^{-\frac{3}{2}}$]
$\sigma_r = 100$	standard deviation of radial uncertainty [m]
$\sigma_\phi = 2$	standard deviation of azimuthal uncertainty [$^\circ$]

details and comparisons, see, e.g., [28]. Simulation parameters are summarized in Table I.

2) *Measures for Estimation Quality:* Let $(\hat{\mathbf{x}}_k^i, \mathbf{P}_k^i)$ denote the local estimate, at time k and in MC run i , after track fusion in one of the agents. Performance is evaluated using the *root mean trace* (RMT) defined as

$$\text{RMT}_k = \sqrt{\frac{1}{M} \sum_{i=1}^M \text{tr}(\mathbf{P}_k^i)}. \quad (42)$$

A robust design of the track fusion must also take into account the uncertainty assessment. To this end we consider conservativeness, which basically means that we want a track fusion method that is able to ensure conservative estimates or at least does not violate the conservativeness property too much. Conservativeness is evaluated using $\lambda_{\max}(\hat{\mathbf{\Xi}})$, where $\hat{\mathbf{\Xi}}$ is the NEES matrix given in Definition 1. In particular, $\lambda_{\max}(\hat{\mathbf{\Xi}})$ is compared to a predetermined value of $F_{\lambda_{\max}}^{-1}(p)$. For $\mathbf{V} \sim \mathcal{W}_m(M, \mathbf{I})$, the confidence parameter p corresponds to the probability that $\lambda_{\max}(\mathbf{V}) \leq F_{\lambda_{\max}}^{-1}(p)$. Hence, since $\hat{\mathbf{\Xi}}$ is normalized by M , we compare $\lambda_{\max}(\hat{\mathbf{\Xi}})$ with $F_{\lambda_{\max}}^{-1}(p)/M$. If $\lambda_{\max}(\hat{\mathbf{\Xi}}) \leq F_{\lambda_{\max}}^{-1}(p)/M$, the estimator is considered conservative.

3) *Results:* The NEES matrix statistics, computed over all MC runs, for each k are displayed in Fig. 4. The gray area and curve correspond to $\{\lambda_{\min}, \bar{\lambda}, \lambda_{\max}\}$ for the local EKF (LKF). Using CI for track fusion results in conservative but rather pessimistic estimates as λ_{\max} is below $1 < F_{\lambda_{\max}}^{-1}(p)/M$. On the other hand, we cannot say that LE is conservative with respect to (w.r.t.) the confidence p . It is interesting that $\bar{\lambda}$ for LE does not deviate considerably from the LKF which by assumption is satisfactorily tuned. Moreover, both are below 1. Hence, if we had only looked at $\bar{\lambda}$, or equivalently $\text{NEES} = n_x \bar{\lambda}$, then we would probably have arrived at the conclusion that also LE is conservative.

The RMT results are presented in Fig. 5¹⁰. The

¹⁰Only the results for Agent 1 are presented. The results for Agent 2 are almost identical.

curves have been normalized by the *Cramér-Rao lower bound* (CRLB, [5]) such that $\text{RMT}_k = 1$ is optimal. It is clearly seen that LE outperforms CI w.r.t. RMT.

In summary, LE shows better performance than CI, but at the cost of not being conservative. The main point is that it requires the NEES matrix statistics to be able to detect that LE is not conservative—the NEES, cf. (8), is not sufficient in this case.

B. Filter Model Mismatch Detection

Using a representative process model is key to the performance of any target tracking system. In practice, the assumed process model used in a tracking filter almost always deviates from the true dynamics of the tracked target. We will now demonstrate how the NIS matrix can be used online to detect a process model mismatch.

1) *Scenario and Models:* The considered scenario is similar to the example in Sec. V-F, but without the switching dynamics. In that example, a linear KF was used to track a single target. The actual target state \mathbf{x} evolves in continuous time according to

$$\dot{\mathbf{x}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \mathbf{x} + \begin{bmatrix} 0 \\ 0 \\ a_x \\ a_y \end{bmatrix}, \quad (43)$$

where $\dot{\mathbf{x}}$ is the time derivative of \mathbf{x} , a_x and a_y are continuous white noise accelerations along the x -axis and y -axis, respectively. It is assumed that

$$\mathbf{Q}^a = \text{cov} \left(\begin{bmatrix} a_x \\ a_y \end{bmatrix} \right) = q^2 \begin{bmatrix} \alpha^2 & 0 \\ 0 & 1/\alpha^2 \end{bmatrix}, \quad (44)$$

where $\alpha = 2$. Note that, for all $\alpha \neq 0$, $\det(\mathbf{Q}^a) = q^2$. Discretizing the continuous time model in (43) results in \mathbf{F}_k according to (35) and [29]

$$\mathbf{Q}_k = q^2 \begin{bmatrix} \frac{\alpha^2 T_k^3}{3} & 0 & \frac{\alpha^2 T_k^2}{2} & 0 \\ 0 & \frac{T_k^3}{3\alpha^2} & 0 & \frac{T_k^2}{2\alpha^2} \\ \frac{\alpha^2 T_k^2}{2} & 0 & \alpha^2 T_k & 0 \\ 0 & \frac{T_k^2}{2\alpha^2} & 0 & \frac{T_k}{\alpha^2} \end{bmatrix}. \quad (45)$$

If $\alpha = 1$, then this \mathbf{Q}_k reduces to (40).

The linear sensor model in (36) is assumed, where $\mathbf{R}_k = \sigma_v^2 \mathbf{I}$. The simulation parameters are summarized in Table II.

The target is simulated using \mathbf{Q}_k with $\alpha = 2$ but the KF uses \mathbf{Q}_k with $\alpha = 1$. Apart from that, all models used in the KF are correct. The task is to

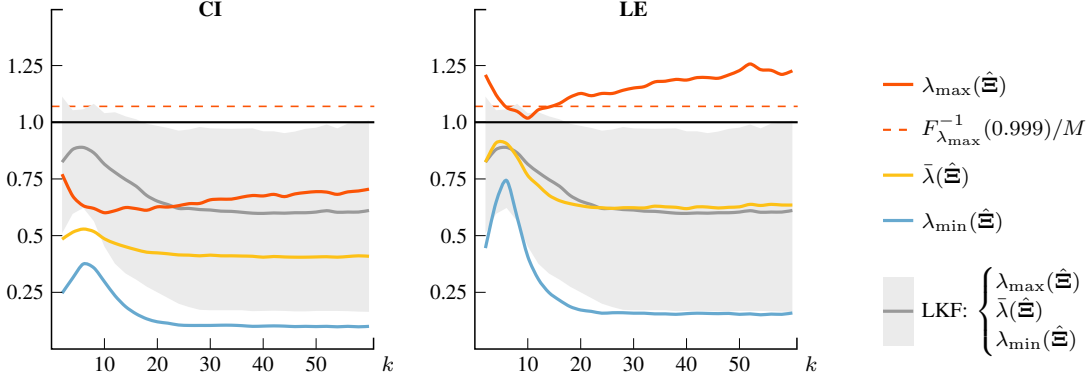


Fig. 4. NEES matrix results for the track fusion design. To evaluate if a track fusion method leads to conservative estimates, $\lambda_{\max}(\hat{\Xi})$ is compared with $F_{\lambda_{\max}}^{-1}/M$. For convenience, also $\bar{\lambda}(\hat{\Xi}) = \text{NEES}/n_x$ and $\lambda_{\min}(\hat{\Xi})$ are included.

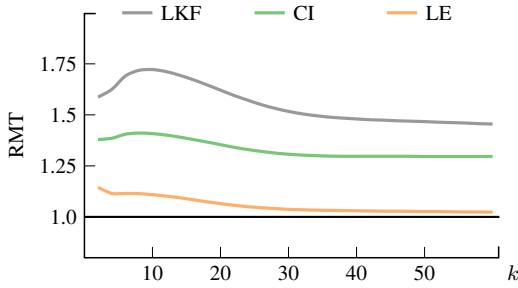


Fig. 5. RMT results for the track fusion design. The RMT curves have been normalized by the CRLB.

TABLE II
MODEL MISMATCH DETECTION PARAMETERS

Parameter	Description
$M = 10\,000$	number of MC runs
$q = 10$	process noise parameter [$\text{ms}^{-\frac{3}{2}}$]
$\sigma_v = 10$	standard deviation of measurement error [m]

detect the process model mismatch. Essentially, we want to test the null hypothesis

$$\mathcal{H}_0: \mathbf{\Pi}_k = \mathbf{I}. \quad (46)$$

Note, we will not design an actual detection algorithm, but instead compute statistics related to the model mismatch.

2) *Measures for Model Mismatch Detection:* The NIS matrix $\hat{\mathbf{\Pi}}$ is used for online evaluation of the model assumption. We compute an accumulated $\hat{\mathbf{\Pi}}_k$ in a single run according to

$$\hat{\mathbf{\Pi}}_k = \frac{1}{k} \sum_{l=1}^k \mathbf{B}_l^{-1} \tilde{\mathbf{y}}_l \tilde{\mathbf{y}}_l^T \mathbf{B}_l^{-T}. \quad (47)$$

The single runs are evaluated separately using MC simulations to obtain good statistics. The performance is evaluated using the probability p_{det} of detecting a model mismatch. For a certain probability parameter $p \in [0, 1]$, we define

$$p_{\text{det}}^{\mathcal{W}} = \Pr((\lambda_{\min} < a_{\lambda_{\min}}) \vee (\lambda_{\max} > b_{\lambda_{\max}})), \quad (48)$$

where $a_{\lambda_{\min}} = F_{\lambda_{\min}}^{-1}(1-p)$, $b_{\lambda_{\max}} = F_{\lambda_{\max}}^{-1}(p)$, and \vee denotes logical or. This corresponds to the probability that at least one $\lambda(\hat{\mathbf{\Pi}})$ is *outside* a $100(2p-1)\%$ confidence interval. As a reference, we define the corresponding probability for the $\bar{\lambda}(\hat{\mathbf{\Pi}})$ statistics accordingly as

$$p_{\text{det}}^{\chi^2} = \Pr((\bar{\lambda} < a_{\chi^2}) \vee (\bar{\lambda} > b_{\chi^2})), \quad (49)$$

where $a_{\chi^2} = F_{\chi^2}^{-1}(1-p)$ and $b_{\chi^2} = F_{\chi^2}^{-1}(p)$. For a fixed p , larger p_{det} means a more sensitive detector.

3) *Complementary Measures:* We also analyze the eigenvectors of $\hat{\mathbf{\Pi}}$ to investigate if they add any complementary information related to the model mismatch. The idea is that the eigenvectors should contain information about which directions there is a mismatch in the process noise. For instance, in the x -axis the actual white noise accelerations have a variance of $q^2\alpha^2$ but the filter is based on q^2 .

Let $\mathbf{u}_{\max,k}$ be the eigenvector associated with $\lambda_{\max}(\hat{\mathbf{\Pi}}_k)$. By construction, $\hat{\mathbf{\Pi}}_k$ is computed in a transformed domain due to the \mathbf{B}_k^{-1} . We therefore define

$$\mathbf{b}_{\max,k} = \mathbf{B}_k \mathbf{u}_{\max,k}. \quad (50)$$

Let θ_k be the angle between $\mathbf{b}_{\max,k}$ and the x -axis, which might be both positive and negative, computed in each of the single runs. We compare $\mathbf{b}_{\max,k}$ and the x -axis since for the true dynamics the process

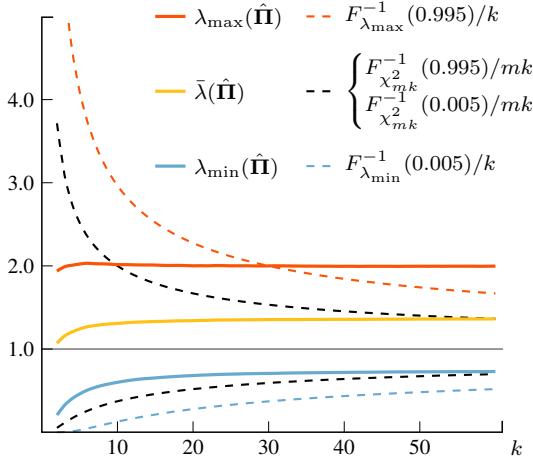


Fig. 6. Filter model mismatch detection, where mean values of λ_{\max} , $\bar{\lambda}$, and λ_{\min} are plotted. Dashed lines refer to normalized inverse CDFs of the computed quantities.

noise is larger in the x -component which should be captured by λ_{\max} . For each time k , we will examine $\bar{\theta}_k$ and σ_{θ_k} , corresponding to the mean and standard deviation of θ_k , respectively, obtained by averaging over the MC runs. Note that, since the eigenvectors are orthogonal, identical results would be obtained by making the corresponding comparison with the eigenvector associated with $\lambda_{\min}(\hat{\mathbf{\Pi}}_k)$ and the y -axis.

4) *Results:* Fig. 6 illustrates the single run statistics with $p = 0.995$. The thick solid curves represent mean values of λ_{\max} , $\bar{\lambda}$, and λ_{\min} , averaged over the MC runs for each time k . The dashed curves illustrate normalized inverse CDFs under \mathcal{H}_0 . We see that λ_{\max} crosses $F_{\lambda_{\max}}^{-1}(0.995)/k$ somewhere around $k = 30$. This indicates that there is a significant level of probability to detect the model mismatch. The same cannot be said for $\bar{\lambda}$ which relates to detecting the model mismatch using χ^2 statistics.

The probabilities $p_{\text{det}}^{\mathcal{W}}$ and $p_{\text{det}}^{\chi^2}$ are approximated by their sample means. That is, for each k , we average the logical expressions inside $\Pr(\cdot)$ in (48) and (49) over the MC runs. The results are plotted in Fig. 7, where $\hat{p}_{\text{det}}^{\mathcal{W}}$ and $\hat{p}_{\text{det}}^{\chi^2}$ refer to the sampled approximations of (48) and (49), respectively. It is clear that using Wishart statistics the detection performance is significantly improved compared to using the χ^2 statistics.

The results related to $\mathbf{b}_{\max,k}$ and θ_k are presented in Fig. 8. These are the single run results which have been averaged over the MC for easier interpretation. While $\bar{\theta}$ is approximately zero-mean over all k , the standard deviation σ_{θ} is initially very high. However,

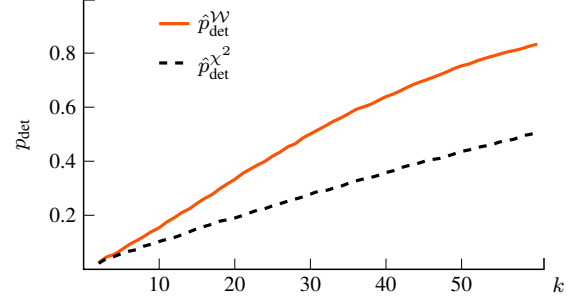


Fig. 7. Filter model mismatch detection, where $\hat{p}_{\text{det}}^{\mathcal{W}}$ and $\hat{p}_{\text{det}}^{\chi^2}$ are sampled approximations of (48) and (49), respectively.

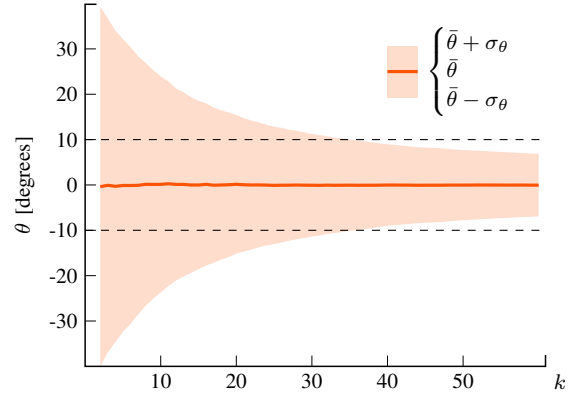


Fig. 8. Filter model mismatch detection. The angle θ represents the deviation of \mathbf{b}_{\max} from the x -axis.

as k increases, σ_{θ} decreases and somewhere between $k = 30$ and $k = 40$ it becomes less than 10° . Hence, it seems like the eigenvectors of $\hat{\mathbf{\Pi}}$ contain some additional information, although noisy, that can be used to draw conclusions about which components the assumed model fails to match the actual process. This opens up for the possibility to use the eigenvalue statistics to say whether there is a filter model mismatch at all, and then use the eigenvectors to decide how the filter can be retuned, in online applications. However, we consider this to be future work.

VII. CONCLUSIONS

We have proposed matrix-valued measures, the *NEES matrix* and the *NIS matrix*, with applications to the design and evaluation of target tracking systems. In particular, it has been shown how the eigenvalues of the NEES and NIS matrices and the associated eigenvalue statistics can be used to draw conclusions about properties such as credibility, filter consistency, and conservativeness. The applicability of the proposed measures was demonstrated using two target

tracking problems: (i) distributed track fusion design; and (ii) filter model mismatch detection.

While the focus of this paper has been on specific target tracking applications, we argue that the proposed measures are useful in essentially all types of estimation problems. For instance, the NIS matrix can be used to evaluate the correctness of landmark initializations in simultaneous localization and mapping (SLAM). It can also serve as an online computable quality measure for, e.g., localization and decision-making problems in general. It would also be interesting to integrate the proposed measures in an auto-tuning framework such as [4]. To this end it might be useful to further elaborate on how the eigenvectors of the NIS matrix can be exploited.

REFERENCES

- [1] S. S. Blackman and R. Popoli, *Design and analysis of modern tracking systems*. Norwood, MA, USA: Artech House, 1999.
- [2] J. Duník, O. Kost, O. Straka, and E. Blasch, "Covariance estimation and Gaussianity assessment for state and measurement noise," *J. Guid., Control, Dyn.*, vol. 43, no. 1, pp. 132–139, 2020.
- [3] Z. Chen, C. Heckman, S. J. Julier, and N. Ahmed, "Time dependence in Kalman filter tuning," in *Proc. 24th IEEE Int. Conf. Inf. Fusion*, Sun City, South Africa, Nov. 2021, pp. 1–8.
- [4] Z. Chen, H. Biggie, N. Ahmed, S. J. Julier, and C. Heckman, "Kalman filter auto-tuning with consistent and robust Bayesian optimization," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 2, pp. 2236–2250, 2024.
- [5] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. New York, NY, USA: John Wiley & Sons, Ltd, 2001.
- [6] J. Wishart, "The generalised product moment distribution in samples from a normal multivariate population," *Biometrika*, vol. 20A, no. 1/2, pp. 32–52, 1928.
- [7] Y. Gao, X.-R. Li, and E. Song, "Robust linear estimation fusion with allowable unknown cross-covariance," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 9, pp. 1314–1325, 2016.
- [8] R. Forsling, A. Hansson, F. Gustafsson, Z. Sjanic, J. Löfberg, and G. Hendeby, "Conservative linear unbiased estimation under partially known covariances," *IEEE Trans. Signal Process.*, vol. 70, pp. 3123–3135, Jun. 2022.
- [9] Y. Bar-Shalom and K. Birmiwal, "Consistency and robustness of PDAF for target tracking in cluttered environments," *Automatica*, vol. 19, no. 4, pp. 431–437, 1983.
- [10] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960.
- [11] X.-R. Li, Z. Zhao, and V. P. Jilkov, "Practical measures and test for credibility of an estimator," in *Proc. Workshop on Estimation, Tracking and Fusion: A tribute to Yaakov Bar-Shalom*, Monterey, CA, USA, May 2001.
- [12] —, "Estimator's credibility and its measures," in *Proc. 15th Triennial IFAC World Congress*, Barcelona, Spain, Jul. 2002.
- [13] X.-R. Li and Z. Zhao, "Measuring estimator's credibility: Noncredibility index," in *Proc. 9th IEEE Int. Conf. Inf. Fusion*, Florence, Italy, Jul. 2006.
- [14] Z. Chen, C. Heckman, S. J. Julier, and N. Ahmed, "Weak in the NEES?: Auto-tuning Kalman filters with Bayesian optimization," in *Proc. 21st IEEE Int. Conf. Inf. Fusion*, Cambridge, UK, Jul. 2018, pp. 1072–1079.
- [15] R. Forsling, "The dark side of decentralized target tracking: Unknown correlations and communication constraints," Dissertations. No. 2359, Linköping University, Linköping, Sweden, Nov. 2023.
- [16] A. T. James, "Distributions of matrix variates and latent roots derived from normal samples," *Ann. Math. Stat.*, vol. 35, no. 2, pp. 475–501, Jun. 1964.
- [17] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. New York, NY, USA: John Wiley & Sons, Inc., 1982.
- [18] M. Chiani, "On the probability that all eigenvalues of Gaussian, Wishart, and double Wishart random matrices lie within an interval," *IEEE Trans. Inf. Theory*, vol. 63, no. 7, pp. 4521–4531, 2017.
- [19] C. G. Khatri, "Distribution of the largest or the smallest characteristic root under null hypothesis concerning complex multivariate normal populations," *Ann. Math. Stat.*, vol. 35, no. 4, pp. 1807–1810, 1964.
- [20] A. Edelman, "The distribution and moments of the smallest eigenvalue of a random matrix of Wishart type," *Linear Algebra Its Appl.*, vol. 159, pp. 55–80, 1991.
- [21] A. Zanella, M. Chiani, and M. Z. Win, "On the marginal distribution of the eigenvalues of Wishart matrices," *IEEE Trans. Commun.*, vol. 57, no. 4, pp. 1050–1060, 2009.
- [22] M. Chiani, "Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the TracyWidom distribution," *J. Multivar. Anal.*, vol. 129, pp. 69–81, 2014.
- [23] C. A. Tracy and H. Widom, "The distributions of random matrix theory and their applications," in *New Trends Math. Phys.*, Dordrecht, Netherlands, 2009, pp. 753–765.
- [24] C. R. Rao, *Linear Statistical Inference and its Applications*, 2nd ed. New York, NY, USA: John Wiley & Sons, Inc., 1973.
- [25] A. Jazwinski, *Stochastic processes and filtering theory*. New York, NY, USA: Academic Press, 1970.
- [26] S. J. Julier and J. K. Uhlmann, "A non-divergent estimation algorithm in the presence of unknown correlations," in *Proc. 1997 Amer. Control Conf.*, Albuquerque, NM, USA, Jun. 1997, pp. 2369–2373.
- [27] A. R. Benaskeur, "Consistent fusion of correlated data sources," in *Proc. 28th Ann. Conf. IEEE Ind. Electron. Soc.*, Sevilla, Spain, Nov. 2002, pp. 2652–2656.
- [28] R. Forsling, B. Noack, and G. Hendeby, "A quarter century of covariance intersection: Correlations still unknown?" *IEEE Control Syst. Mag.*, vol. 44, no. 2, pp. 81–105, 2024.
- [29] C. Van Loan, "Computing integrals involving the matrix exponential," *IEEE Trans. Autom. Control*, vol. 23, no. 3, pp. 395–404, 1978.



Robin Forsling received the M.Sc. degree in engineering physics in 2016 from Umeå University, Umeå, Sweden, and the Ph.D. degree in automatic control in 2023 from Linköping University, Linköping, Sweden. Since 2016 he is employed at Saab Aeronautics in Linköping, Sweden, where he currently works with mission functions and tactical integration in Sweden's future combat air systems programme. His research interests include target tracking and data fusion for network-centric operations, and robust decision-making under uncertainty.



Simon J. Julier (Member, IEEE) received the Ph.D. degree in computer science from the University of Oxford in 1997. Between 1997 and 2006 he worked at the Naval Research Laboratory, Washington DC, where he worked on tracking and estimation in nonlinear systems and in mobile augmented reality systems. Since 2006 he has been a member of the Computer Science Department at University College London. His research

interests include user interfaces, distributed data fusion, nonlinear estimation, and simultaneous localization and mapping.



Gustaf Hendeby (Senior Member, IEEE) received the M.Sc. degree in applied physics and electrical engineering in 2002 and the Ph.D. degree in automatic control in 2008, both from Linköping University, Linköping, Sweden. He is Associate Professor and Docent at the division of Automatic Control, Department of Electrical Engineering, Linköping University. His main research interests

include sensor fusion and stochastic signal processing, with applications to nonlinear problems, target tracking, and simultaneous localization and mapping.