An Improved Random Matrix Prediction Model for Manoeuvring Extended Targets

Nathan J. Bartlett, Chris Renton, Adrian G. Wills

Abstract—This paper proposes an improved prediction update for extended target tracking with the random matrix model. A key innovation is to employ a generalised non-central inverse Wishart distribution to model the state transition density of the target extent; resulting in a prediction update that accounts for kinematic state dependent transformations. Moreover, the proposed prediction update offers an additional tuning parameter c.f. previous works, requires only a single Kullback-Leibler divergence minimisation, and improves overall target tracking performance when compared to state-of-the-art alternatives.

Index Terms—Extended target tracking, random matrix model, non-central inverse Wishart, Kullback-Leibler divergence.

I. INTRODUCTION

TARGET tracking is an important practical problem that has received significant research attention since the pioneering works of [1]. Due to an increase in sensor resolution capabilities, it is now commonplace for multiple measurements to be generated for each target per time-step; e.g., modern radar can produce several range measurements for a single target. One approach to handle the generation of multiple measurements is to employ extended-target models, where, in addition to estimating the kinematic state variables, the spatial extent of the target is estimated from available data. In recognition of the utility of extended targets, several models have been proposed over the past decade, see e.g., [2]–[10].

One of the most popular extended-target models, referred to as the random matrix model, was first proposed in the works of Koch [10], and defines the extended target state as the combination of a kinematic state vector and an extent matrix. The extent matrix is assumed to be symmetric positive definite, which in turn enables for the target extent to be represented by an ellipsoid [11]. The kinematic state vector is modelled as a Gaussian distributed random variable, whilst the extent matrix is modelled as an inverse Wishart distributed random variable. Following a Bayesian methodology, the prediction density of each random variable is assumed to belong to the same distribution class of its respective posterior, enabling for the parameters of each density to be updated recursively; consisting of a prediction and a correction stage.

In the early work of Koch and Feldmann et al. [10], [11], the prediction update of the extent matrix was based upon a simple heuristic that artificially increased the covariance whilst preserving the expected value. Koch additionally proposed the use of a Wishart state transition density [10], which was later generalised within [12], [13] to handle temporal evolutions of the target extent. However, even for evolutions independent of the kinematic state, a second-order moment matching technique or Kullback-Leibler divergence minimisation is required to approximate the resulting Generalised Beta Type-II prediction density as an inverse Wishart distribution; instigating information loss in regard to the target extent [14]. To resolve this issue, Bartlett et al. [14] proposed the use of a non-central inverse Wishart state transition density, which was shown to produce an inverse Wishart prediction distribution directly and significantly improve target extent estimation.

Contributions: The key contribution of this paper is the generalisation of the non-central inverse Wishart state transition density to account for kinematic state dependent evolutions of the target extent. This generalisation is motivated by the works of [13], whom, to the best of the authors’ knowledge, were the first to propose a prediction update to handle kinematic state dependent evolutions of the target extent. The main difference here is that by building upon [14], our proposed prediction update requires only a single Kullback-Leibler divergence minimisation, and offers an additional tuning parameter to model uncertainties in target shape more effectively.

The remainder of the paper is organised as follows. In Section II, we provide an overview of the random matrix model and the prediction updates of [10]–[14]. The problem formulation is presented in Section III and the proposed prediction update is given in Section IV. Simulated results comparing the proposed prediction update against state-of-the-art alternatives are presented in Section V. Concluding remarks are given in Section VI. Notation and distributions are summarised in Table I.

II. THE RANDOM MATRIX MODEL

The random matrix model can be accredited to the works of Koch and Feldmann et al. [10], [11]. The model has been used in an wide array of extended-target tracking applications over the last decade, including aircraft tracking with ground radar, pedestrian tracking with laser range sensors, and surface vessel tracking with marine X-band radar; see e.g., [15]–[22].

The random matrix model defines the extended target state \( \xi_k \triangleq (x_k, X_k) \) at time \( t_k \) as the combination of a kinematic state vector \( x_k \in \mathbb{R}^{n_x} \) and an extent matrix \( X_k \in \mathbb{S}^{d}_{++} \). The kinematic state vector consists of states related to the motion of the target centre, whilst the extent matrix models the target extent as a \( d \)-dimensional ellipsoid [14]. Thus, the region of space \( \mathcal{Y} \subseteq \mathbb{R}^d \) occupied by the extended target at time \( t_k \) can be described by the following set:

\[
\mathcal{Y} \triangleq \{ y \in \mathbb{R}^d : (y - H_k x_k)^T X_k^{-1} (y - H_k x_k) \leq 1 \}, \quad (1)
\]
where $H_k$ is a matrix that transforms the kinematic state vector to a target position. In the conditional random matrix model, the kinematic state can only consist of the target position and time derivatives such as velocity and acceleration [10]. In the factorised random matrix model—which shall be utilised in this paper—the kinematic state can also include non-linear kinematics such as turn-rate and heading [11]. The measurements $z_k^r \in Z_k$ at time $t_k$ are assumed to be multivariate Gaussian distributed with covariance related to the extent matrix

$$p(Z_k|\xi_k) = \prod_{r=1}^{n_{k}} \mathcal{N}(z_k^r|H_kx_k, \lambda X_k + R_k),$$

where $R_k$ is the sensor noise covariance, and $\lambda$ is the scaling factor to the spread contribution of the target extent $\Gamma_k$. Hence, the factorised random matrix model defines the posterior of the extended target state $\xi_k$ as the product of a Gaussian kinematic state density and an inverse Wishart matrix density [11].

More specifically, the posterior distribution is approximated as

$$p(\xi_k|Z_k^k) \approx p(x_k|Z_k^k)p(X_k|Z_k^k),$$

where $Z_k^k = \{Z_1, \ldots, Z_k\}$ is the set of all measurement sets up to and including time $t_k$, and

$$p(x_k|Z_k^k) = \mathcal{N}(x_k|m_{k|k}, P_{k|k}),$$

$$p(X_k|Z_k^k) = \mathcal{IW}_d(X_k|\nu_{k|k}, V_{k|k}).$$

In Bayesian estimation, it is often desired for the prediction density to belong to the same distribution class as the posterior [3]; i.e., we want

$$p(\xi_{k+1}|Z_k^k) \approx p(x_{k+1}|Z_k^k)p(X_{k+1}|Z_k^k),$$

where

$$p(x_{k+1}|Z_k^k) = \mathcal{N}(x_{k+1}|m_{k+1|k}, P_{k+1|k}),$$

$$p(X_{k+1}|Z_k^k) = \mathcal{IW}_d(X_{k+1}|\nu_{k+1|k}, V_{k+1|k}).$$

Assuming this condition is satisfied, the prediction update corresponds to obtaining the parameters $\{m_{k+1|k}, P_{k+1|k}\}$, and $\{\nu_{k+1|k}, V_{k+1|k}\}$ of the predicted Gaussian kinematic state vector density [4d] and the predicted inverse Wishart extent matrix density [4e], respectively.

In the early work of Feldmann et al. [11], an extended Kalman filter prediction is utilised to obtain the parameters of the predicted kinematic state vector density [4b]. Moreover, $R_k \approx 0_k$ to prove conjugacy [10]. In scenarios where this approximation is invalid, additional heuristics are required to obtain a closed-form correction; see [11] for further details.

### Table I: Notation

- $I_n$ is a $n \times n$ identity matrix, and $0_m$ is a $m \times m$ zero matrix.
- $\mathbb{R}^n$ is the set of real column vectors of length $n$, $\mathbb{R}^{m \times n}$ is the set of real $m \times n$ matrices, $\mathbb{S}^n_+$ is the set of symmetric positive definite $n \times n$ matrices, $\mathbb{S}^n$ is the set of symmetric positive semi-definite $n \times n$ matrices, $N$ is the set of natural numbers, $SL(n, \mathbb{R})$ is the special linear group of $n \times n$ matrices with determinant equal to one, and $\mathbb{O}^{m \times n}$ is the set of $m \times n$ semi-orthogonal matrices with $m \leq n$.
- $\mathcal{N}(\mathbf{x}, \mathbf{P})$ denotes a multivariate Gaussian pdf defined over the vector $\mathbf{x} \in \mathbb{R}^n$, with expectation $\mathbf{m} \in \mathbb{R}^n$ and covariance matrix $\mathbf{P} \in \mathbb{S}^n_+$,

$$\mathcal{N}(\mathbf{x}|\mathbf{m}, \mathbf{P}) = \frac{1}{(2\pi)^{d/2}|\mathbf{P}|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T\mathbf{P}^{-1}(\mathbf{x} - \mathbf{m})\right),$$

where $|\cdot|$ denotes the matrix determinant.
- $\mathcal{IW}_d(\mathbf{X}|\nu, \mathbf{V})$ denotes an inverse Wishart pdf defined over the matrix $\mathbf{X} \in \mathbb{S}^d_+$ with scalar degrees of freedom $\nu > 2d$ and parameter matrix $\mathbf{V} \in \mathbb{S}^d_+$. \[23\] Definition 3.4.1.

$$\mathcal{IW}_d(\mathbf{X}|\nu, \mathbf{V}) = \frac{\exp(-\frac{1}{2}\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X} - \frac{d\nu}{2} - \frac{1}{2}\ln|\mathbf{V}| - \ln|\mathbf{X}|^2)}{\Gamma_d(\nu)} \mathcal{W}(\nu, \mathbf{V}^{-1}, \mathbf{X})^{-1/2},$$

where $\Gamma_d(\cdot)$ represents the determinant of the matrix trace, and $\mathcal{W}(\cdot, \cdot, \cdot)$ is the multivariate Gamma function which can be expressed as a product of ordinary gamma functions \[23\] Theorem 1.4.1. The expectation of $\mathcal{W}(\cdot, \cdot, \cdot)$ is $\mathcal{W}(\cdot, \cdot, \cdot)$ [23] Theorem 3.4.3.

- $\mathcal{W}_d(\mathbf{X}|\nu, \mathbf{W})$ denotes a Wishart pdf defined over the matrix $\mathbf{X} \in \mathbb{S}^d_+$ with scalar degrees of freedom $\nu > d - 1$ and parameter matrix $\mathbf{W} \in \mathbb{S}^d_+$. \[23\] Definition 3.2.1.

$$\mathcal{W}_d(\mathbf{X}|\nu, \mathbf{W}) = \frac{\exp(-\frac{1}{2}\mathbf{X}^T\mathbf{W}^{-1}\mathbf{X} - \frac{d\nu}{2} - \frac{1}{2}\ln|\mathbf{W}| - \ln|\mathbf{X}|^2)}{\Gamma_d(\nu)\mathcal{W}(\nu, \mathbf{W}^{-1}, \mathbf{X})^{-1/2}}.$$
value and covariance of the extent matrix are always well-defined [24].

Koch also proposed to solve the Chapman-Kolmogorov equation for a Wishart state transition density, and then to approximate the resulting Generalised Beta Type-II prediction density as an inverse Wishart distribution [10]. This idea was further built upon in [12], where an invertible parameter matrix is introduced to describe transformations of the extent matrix that are independent of the kinematic state vector. Moreover, a second-order moment matching technique is proposed to perform the inverse Wishart approximation [12], [25].

Granström et al. further generalised the idea of using a Wishart state transition density, presenting a prediction update that enables the evolution of the extent matrix to be functionally dependent upon the kinematic state vector [13]:

\[ p(X_{k+1}|X_k, V_k) = W_d\left(X_{k+1}|n_{k+1}, M_{x_k}X_k M_{x_k}^T M_{x_k}^{-n_{k+1}}\right). \] (6)

Here, the scalar design parameter \( n_{k+1} > d-1 \), and the matrix transformation \( M_{x_k} \triangleq M(x_k) \) such that \( M: \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{d \times d} \) is a nonsingular matrix-valued function of the kinematic state vector [13]. In order to account for kinematic state uncertainty, the use of (6) leads to the following integral representation of the predicted extent matrix density:

\[ p(X_{k+1}|Z_k) = \int N(x_k|m_{k|k}, P_{k|k}) \times GB_d^{n_k}(X_{k+1}|n_{k+1}, \frac{\nu_{k|k}}{2} - 2d - 2, M_{x_k}V_{k|k}M_{x_k}^T, 0_d, dx_k. \] (7)

Unfortunately, the above integral has no analytical solution [13]. Hence, a series of Kullback-Leibler divergence minimisations is employed to approximate (7) as an inverse Wishart distribution [26]. To summarise:

\[ p(X_{k+1}|Z_k) \approx \mathcal{W}_d(X_{k+1}|n_{k+1}, V_{k+1}|k), \] (8)

where, the scalar degrees of freedom \( \nu_{k+1|k} \) and parameter matrix \( V_{k+1|k} \) are functions of the expected values

\[ C_1 = \int (M_{x_k} V_{k|k} M_{x_k}^T)^{-1} N(x_k|m_{k|k}, P_{k|k}) dx_k, \] (9a)

\[ C_2 = \int M_{x_k} V_{k|k} M_{x_k}^T N(x_k|m_{k|k}, P_{k|k}) dx_k. \] (9b)

Note that the above definition of \( C_1 \) was first presented in [27], and results in a more efficient implementation of the prediction update of [13]. See [27, Table IV] for further details.

Granström et al. showed that the use of (6) offers significant improvement in the tracking of extended targets within unknown turn-rates when compared to [10]–[12], [25]. Nevertheless, even under the assumption that the time evolution is independent of the kinematic state vector, i.e., \( M_{x_k} = M_k \), the prediction update requires a Kullback-Leibler divergence minimisation or moment matching technique to approximate the Generalised Beta Type-II density as an inverse Wishart distribution [13], [27]. In order to remove the need for such density approximations, Bartlett et al. [14] proposed the following non-central inverse Wishart state transition density:

\[ p(X_{k+1}|X_k, Z_k) = \mathcal{W}_d^{nc}(X_{k+1}|\nu_{k+1}, \Sigma_{k+1}, \Sigma_{k+1} \Theta_{k+1}(X_k)), \] (10)

where the degrees of freedom \( \nu_{k+1} \), parameter matrix \( \Sigma_{k+1} \), and non-centrality matrix \( \Theta_{k+1}(X_k) \) are defined as follows:

\[ \nu_{k+1} = (2d, \nu_{k|k}) \cap \mathbb{N}, \] (11a)

\[ \Sigma_{k+1} = M_{k+1}Q_{k+1}^{-1} M_{k+1}^T, \] (11b)

\[ \Theta_{k+1}(X_k) = M_{k+1}^{-1} X_k^{-1} M_{k+1}^T. \] (11c)

Here, the nonsingular \( d \times d \) transition matrix \( M_{k+1} \) is used to model transformations independent of the kinematic state vector, and \( Q_{k+1} \) is used to model uncertainties in the extent matrix evolution; offering an additional \( d(d+1)/2 \) tunable parameters than previous works [14]. Moreover, the use of (10) guarantees the density density is of the desired inverse Wishart form [45] with the following scalar degrees of freedom and parameter matrix:

\[ \nu_{k+1|k} = \nu_{k+1}, \] (12a)

\[ V_{k+1|k} = M_{k+1} V_{k|k}(I_d + Q_k V_{k|k})^{-1} M_{k+1}^T. \] (12b)

The main contribution of this paper is to propose a generalisation of (10) that allows for kinematic state dependent evolutions of the extent matrix; e.g., rotations and scaling. In doing so, the proposed prediction update does not suffer from the same levels of information loss as [13], [27], requires only a single Kullback-Leibler divergence minimisation, and offers an additional tuning parameter to model uncertainties in target shape more effectively.

III. PROBLEM FORMULATION

In Bayesian filtering, the prediction step consists of solving the following integral

\[ p(\xi_{k+1}|\xi_k, Z_k) = \int p(\xi_{k+1}|\xi_k, Z_k)p(\xi_k|Z_k) d\xi_k. \] (13)

In order to obtain a closed-form solution to this integral, it is assumed in [14] that the evolution of the extent matrix is independent of the kinematic state vector. Albeit true for non-maneuvring behaviours, this assumption is often violated during constant or variable turn manoeuvres—in which the target extent rotates as a function of the turn-rate [13]. Therefore, in order to improve the tracking performance of manoeuvring targets, we will now account for such dependency.

Inspired by [13], the non-Markov state transition density \( p(\xi_{k+1}|\xi_k, Z_k) \) is expanded as follows:

\[ p(\xi_{k+1}|\xi_k, Z_k) = p(X_{k+1}|X_{k+1}, X_k, Z_k) \times \]

\[ p(X_{k+1}|X_k, X_k, Z_k), \] (14a)

\[ \approx p(X_{k+1}|X_k)p(X_{k+1}|X_k, Z_k). \] (14b)

Following [14], the state transition density is non-Markov, and thus retains the dependency upon the measurement set \( Z_k \). This action enables for greater flexibility in the selection of state transition parameters to model target shape uncertainties; see [14, Section IV.D] for further details.
The evolution of the kinematic state vector is assumed to be independent of the extent matrix. Hence, phenomena dictated by the target extent are deemed negligible; for example, wind resistance [17].

Given posterior (13a) and non-Markov state transition density [14b], equation [13] becomes
\[ p(\xi_{k+1}|Z^k) = \int p(x_{k+1}|x_k) \int p(x_{k+1}|x_k, X_k) \times p(X_k|Z^k)p(x_k|Z^k)dX_kdx_k. \] (15)

As previously stated, it is often desired in Bayesian filtering for the prediction density to belong to the same distribution class as the posterior—enabling for a finite set of statistics to be updated in Bayesian recursion rather than the entire distribution [28]. Therefore, given (13a), we require that the resulting prediction density of (15) possesses the same functional form as (14a). Unfortunately, this condition cannot be proven to hold in general. Thus, as in the pioneering works of Granström et al. [13], we approximate (15) to be the product of two independent equations: one for the kinematic state vector, and another for the extent matrix:

\[ p(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}|f_{k+1}(x_k), D_{k+1}), \] (16a)

\[ p(x_{k+1}|x_k, X_k) = \int p(x_{k+1}|x_k, X_k, Z^k)p(X_k|Z^k)dX_kdx_k. \] (16b)

We must now determine an appropriate state transition density for both the kinematic state vector and extent matrix such that the resulting prediction densities of (16a) and (16b) are multivariate Gaussian and inverse Wishart distributions respectively. Following [13, 29], the state transition density of the kinematic state vector is given by,

\[ p(x_{k+1}|x_k) = \mathcal{N}(x_{k+1}|f_{k+1}(x_k), D_{k+1}), \] (17)

where \( f_{k+1} : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{n_x} \) is the nonlinear state transition function, and \( D_{k+1} \in \mathbb{S}_{n_x}^{+} \) is the dynamic noise covariance matrix [20, 31]. Substituting the state transition density (17) and posterior (13) into (16a), the extended Kalman filter is then used to approximate the resulting prediction density as a multivariate Gaussian distribution [32, 33]. That is,

\[ p(x_{k+1}|x_k) \approx \mathcal{N}(x_{k+1}|m_{k+1|k}, P_{k+1|k}), \] (18a)

where the mean and covariance are given by
\[ m_{k+1|k} = f_{k+1}(m_{k|k}), \] (18b)
\[ P_{k+1|k} = F_{k+1}F_{k+1}^T + D_{k+1}, \] (18c)
\[ F_{k+1} = \nabla_x f_{k+1}(x)|_{x=m_{k|k}}. \] (18d)

The problem considered in this work is to derive an analogous closed-form prediction update for the extent matrix. More specifically, to determine a suitable non-Markov state transition density \( p(X_{k+1}|x_k, X_k, Z^k) \) that can be used in (16b) to obtain the desired inverse Wishart distribution [42] with minimal approximations. In addition, the chosen state transition density must offer a high degree of modelling flexibility, be physically interpretable, and encapsulate a wide subset of possible kinematic state dependencies.

IV. PREDICTION UPDATE

In this section, we introduce the state transition density and present the new prediction update of the extent matrix. We show that via a single Kullback-Leibler divergence minimisation, the resulting prediction density of (16b) is an inverse Wishart distribution; given the state transition density is a non-central inverse Wishart distribution. All supporting lemmata and corollaries are given in the Appendix.

A. The Extent Matrix State Transition Density

Motivated by [13, 14], we define the non-Markov state transition density of the extent matrix as the following non-central inverse Wishart distribution:

\[ p(x_{k+1}|x_k, X_k, Z^k) = \mathcal{I}\mathcal{W}_d^{nc}(X_{k+1}|V_{k+1}, \Sigma_{k+1}(x_k), \Theta_{k+1}(x_k, X_k)), \] (19)

where the scalar degrees of freedom \( v_{k+1} \), parameter matrix \( \Sigma_{k+1}(x_k) \), and non-centrality matrix \( \Theta_{k+1}(x_k, X_k) \) are defined as follows:

\[ v_{k+1} \in (2d, v_{k+1}] \cap \mathbb{N}, \] (20a)
\[ \Sigma_{k+1}(x_k) = M_{x_k}Q_{k+1}^{-1}M_{x_k}^T, \] (20b)
\[ \Theta_{k+1}(x_k, X_k) = M_{x_k}^{-1}X_kM_{x_k}^{-1}. \] (20c)

The tuneable parameters of the non-Markov state transition density are thereby the degrees of freedom \( v_{k+1} \), the symmetric positive definite noise matrix \( Q_{k+1} \in \mathbb{S}_{d}^{+} \), and the matrix transformation \( M_{x_k} \triangleq M(x_k) \); where \( M : \mathbb{R}^{n_x} \rightarrow \mathbb{R}^{d \times d} \) is a nonsingular matrix-valued function of the kinematic state. To provide adequate meaning to these parameters in regard to extended target tracking, we shall now briefly discuss the underlying state transition model.

B. The Extent Matrix State Transition Model

The proposed state transition density is a generalisation of [14]. From Lemma [1] the state transition model governing (19) and describing the evolution of the extent matrix from time \( t_k \) to \( t_{k+1} \) is

\[ X_{k+1}^{-\frac{1}{2}} = M_{x_k}^{-\frac{1}{2}}(X_k^{-\frac{1}{2}} + n_{k+1}^{-\frac{1}{2}}W_{k+1}), \] (21a)
\[ W_{k+1} \sim \mathcal{N}(0, Q_{k+1}/n_{k+1}), \] (21b)

where \( n_{k+1} = v_{k+1} - d - 1 \). Although model (21) seems rather complex, its physical interpretation can be separated into two simple components: the injection of Wishart distributed process noise \( W_{k+1} \), and the extent evolution described by the transition matrix \( M_{x_k} \) [14]. We shall now discuss each of these components to highlight the properties of each state transition parameter.

In accordance with [33, Theorem 3.3.15], the expectation and variance of the process noise \( W_{k+1} \) is given by:

\[ \mathbb{E}[W_{k+1}] = Q_{k+1}, \] (22a)
\[ \text{Var}(W_{k+1}) = \frac{Q_{k+1}Q_{k+1}}{v_{k+1}-d-1} + \frac{\text{tr}(Q_{k+1})Q_{k+1}}{v_{k+1}-d-1}. \] (22b)
By (22), the primary parameter that governs the size, shape, and expected value of the process noise is the noise matrix $Q_{k+1}$ \cite{14}. Therefore, contrary to \cite{13, 27}, our model offers an additional $d(d + 1)/2$ tunable parameters to describe the effects of process noise on the extent matrix evolution. For example, varying levels of process noise can be applied to each principle axis of the target extent; which can aid in the tracking of group targets such as truck convoys \cite{14}.

Furthermore, as $Q_{k+1}$ approaches $0_{d,d}$, the expected value and variance of the process noise also approach $0_{d,d}$ \cite{22}. This implies that, although non-linearly, smaller elements of $Q_{k+1}$ results in a more deterministic evolution of the target extent \cite{14}. Hence, the noise matrix $Q_{k+1}$ is analogous to the dynamic noise covariance matrix $D_{k+1}$ \cite{17}, which models the uncertainty in the evolution of the kinematic state vector $x_k$ from time $t_k$ to $t_{k+1}$. In Section IV-D, we shall provide methodologies for tuning $v_{k+1}$ and $Q_{k+1}$ for extended target tracking purposes.

Once the process noise $W_{k+1}$ has been injected, the state transition model then performs the evolution described by $M_{x_k}$. To highlight the effects of $M_{x_k}$, consider the ideal scenario in which the process noise is zero. Then, (21a) is equivalent to the deterministic model

$$X_{k+1} = M_{x_k} X_k M^T_{x_k}. \quad (23)$$

Thus, as in the works of \cite{13}, the main motivation for $M_{x_k}$ is to model rotations of the target extent; however, in general, the function $M_{x_k}$ can be selected as any arbitrary transformation—provided the output is a non-singular $d \times d$ matrix \cite{14}. For example, in group target tracking the extent matrix $M_{x_k}$ is thereby analogous to the kinematic state transition function $f_{k+1}$ \cite{17}; which models the evolution of the kinematic state vector $x_k$ from time $t_k$ to $t_{k+1}$. Within our work, we shall consider $M_{x_k}$ to be of the following form:

$$M_{x_k} = \begin{bmatrix} \cos(T \omega_k) & -\sin(T \omega_k) \\ \sin(T \omega_k) & \cos(T \omega_k) \end{bmatrix}, \quad (24)$$

where $T$ is the time interval, $x_k = [x_k, y_k, \dot{x}_k, \dot{y}_k, \omega_k]^T$, and $\omega_k$ is the turn-rate of the extended target. Note that (24) is commonly utilised to model target extent rotations; see e.g., \cite{13, 16, 27, 29}.

C. The Generalised Extent Matrix Prediction Update

Substituting the posterior $\overline{x}_k$ and the state transition density \cite{19} into (16b), the prediction update of the extent matrix is equivalent to

$$p(X_{k+1}|Z^k) = \int \int \int_{\Omega} T W_d(x_{k+1}|v_{k+1}, \Sigma_{k+1}(x_k), \Sigma_{k+1}(x_k)\Theta(x_k, X_k)) \times \mathcal{N}(x_k|m_{k|k}, P_{k|k}) \mathcal{N}(x_k|m_{k|k}, P_{k|k}) dX_k dX_{k+1} dX_{k+1}, \quad (25)$$

which, given \cite{14} Theorem 1, yields the intermediate integral

$$p(X_{k+1}|Z^k) = \int \mathcal{N}(x_k|m_{k|k}, P_{k|k}) T W_d(x_{k+1}|v_{k+1}, M_{x_k} \bar{V}_{k+1} M^T_{x_k}) dX_{k+1}, \quad (26a)$$

where the intermediate parameter matrix $\bar{V}_{k+1}$ is

$$\bar{V}_{k+1} = V_{k|k}(I_d + Q_{k+1} V_{k|k})^{-1}. \quad (26b)$$

Unfortunately, the above integral (26a) has no analytical solution. Hence, motivated by \cite{13}, we resort to approximating the resulting prediction density with an inverse Wishart distribution through Kullback-Leibler divergence minimisation.

Kullback-Leibler divergence is considered the optimal difference measure when approximating distributions in a maximum likelihood sense \cite{34–36}. For the two distributions $p(X)$ and $q(X)$, the Kullback-Leibler divergence is defined as follows \cite{37}:

$$\text{KL}(p(X)||q(X)) = \int p(X) \log \left( \frac{p(X)}{q(X)} \right) dX. \quad (27)$$

By defining $p(X)$ as (26a) and $q(X)$ by (28c), Lemma 2 proves that the best and unique global approximation of (26a) that minimises (27) is given by

$$p(X_{k+1}|Z^k) \approx T W_d(x_{k+1}|v_{k+1}|k, V_{k+1|k}), \quad (28a)$$

where the parameter matrix $V_{k+1|k}$ is equal to

$$V_{k+1|k} = \left( \frac{\nu_{k+1|k} - d - 1}{\nu_{k+1|k} - d - 1} \right) C_1^{-1}, \quad (28b)$$

and the scalar degrees of freedom $\nu_{k+1|k}$ is the solution to

$$d \ln \left( \frac{v_{k+1|k} - d - 1}{\nu_{k+1|k} - d - 1} \right) + \sum_{i=1}^{d} \psi_0 \left( \frac{v_{k+1|k} - d - i}{2} \right) - C_3 - \ln (|C_1|) = 0. \quad (28c)$$

Furthermore, $C_1$ and $C_3$ are the expectations

$$C_1 = \int (M_{x_k} \bar{V}_{k+1} M^T_{x_k})^{-1} N(x_k|m_{k|k}, P_{k|k}) dX_k, \quad (28d)$$

$$C_3 = \int \ln (|M_{x_k} \bar{V}_{k+1} M^T_{x_k}|) N(x_k|m_{k|k}, P_{k|k}) dX_k. \quad (28e)$$

The optimal value of the scalar degrees of freedom $\nu_{k+1|k}$ can be found by applying a numerical root-finding algorithm to (28c). Examples include the Newton-Raphson algorithm and Halley’s method \cite{38}. Nevertheless, to improve the computational efficiency of the prediction update, we shall utilise the following theorem to obtain a closed-form solution for $\nu_{k+1|k}$.

Theorem 1. The expected values of (26a) and (28a) will have the same volume if the scalar degrees of freedom $\nu_{k+1|k}$ is

$$\nu_{k+1|k} = 2d + 2 + \frac{(d + 1) \rho_{k+1}}{(\rho_{k+1} + d + 1) C_1 C_2 \frac{d}{2} - \rho_{k+1}}, \quad (29)$$

where $\rho_{k+1} = v_{k+1} - 2d - 2$, and the expectation

$$C_2 = \int M_{x_k} \bar{V}_{k+1} M^T_{x_k} N(x_k|m_{k|k}, P_{k|k}) dX_k. \quad (30)$$
Proof. Let $\mathbb{E}[X_{k+1} | Z^k]$ and $\mathbb{E}[X_{k+1} | Z^k]$ denote the expected values of (26a) and (28a) respectively. By [23] Theorem 3.4.3,

$$
\mathbb{E}[X_{k+1} | Z^k] = \frac{C_2}{\nu_{k+1} - 2d - 2},
$$
(31a)

$$
\mathbb{E}[X_{k+1} | Z^k] = \frac{C_2}{\nu_{k+1} - 2d - 2}.
$$
(31b)

The volume of an ellipsoid $X$ is equal to $c_d |X|$ where $c_d$ is the volume of a $d$-dimensional unit sphere [14]. Matching the volume of the expected values is thereby equivalent to matching the determinants

$$
\frac{V_{k+1}^{|k|}}{\nu_{k+1} - 2d - 2} = \frac{C_2}{\nu_{k+1} - 2d - 2}.
$$
(32)

Substituting (28b) into (32) and re-arranging yields (29). \qed

As stated previously, the main motivation for $M_{x_k}$ is to model rotations of the target extent. Thus, to compare the optimal (28c) and closed-form (29) solutions of $\nu_{k+1}|k|$, we define $M_{x_k}$ by (24). Furthermore, motivated by [13] Corollary 2, we approximate the expectations $C_1$, $C_2$, and $C_3$ with third-order Taylor series expansions. That is, given $V_{x_k} \triangleq M_{x_k} V_{k+1} M_{x_k}^T$:

$$
C_1 \approx \left( \nabla_{x_k} - 1 + \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \frac{\partial^2 \nabla_{x_k}}{\partial x_k^i \partial x_k^j} P_{k|k}^{ij} \right) |_{x_k = m_{k|k}},
$$
(33a)

$$
C_2 \approx \nabla_{x_k} + \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \frac{\partial^2 \nabla_{x_k}}{\partial x_k^i \partial x_k^j} P_{k|k}^{ij} \right) |_{x_k = m_{k|k}},
$$
(33b)

$$
C_3 \approx \left( \ln(\nabla_{x_k}) + \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \frac{\partial^2 \ln(\nabla_{x_k})}{\partial x_k^i \partial x_k^j} P_{k|k}^{ij} \right) |_{x_k = m_{k|k}},
$$
(33c)

where $x_k^i$ denotes the $i$th element of $x_k$, and $P_{k|k}^{ij}$ denotes the $(i, j)^{th}$ element of $P_{k|k}$.

Figure 1 shows the optimal and closed-form solutions of $\nu_{k+1}|k|$ for differing levels of turn-rate variance $P_{\omega_k}^{\omega_k}$ and state transition degrees of freedom $\nu_{k+1}$. The intermediate parameter matrix $\tilde{V}_{k+1} = \text{diag}((10^2, 5^2))$, and the expected value of the turn-rate $\omega_{k|k} = 10^6/T$. From Figure 1, it is observed that the maximum relative error between the optimal and closed-form solutions is less than ten percent. We remark here that relative error experienced between the optimal and closed-form solutions proposed in [13] is typically on the order of one tenth of a degree of freedom; see [13] Corollary 1. We thereby conclude that, as in the works of [13], our closed-form solution is an adequate approximation to the optimal solution.

In addition to the above discussion, the closed-form solution (29) ensures the volume of the expected value of the extent matrix remains constant over the inverse Wishart approximation. Furthermore, by Jensen’s inequality, the expected value of the extent matrix is always well-defined. That is, the closed-form solution guarantees $\nu_{k+1}|k| > 2d + 2$. We remark

Here that the same cannot be said for the optimal solution; which only guarantees $\nu_{k+1}|k| > 2d$. The proposed prediction update is presented in Table II. To avoid numerical root-finding and ensure the expected value is well-defined, the update uses Theorem I.

D. Parameter Selection

In this section, we discuss methodologies to select the state transition parameters to model practical phenomena. Before presenting said methodologies however, we first remark that the extent matrix state transition density (19) is non-Markov. This enables for the state transition parameters $\nu_{k+1}$ and $Q_{k+1}$ to be functionally dependent upon the posterior parameters $V_{k|k}$ and $V_{k|k}$.

The first setting introduced for the state transition degrees of freedom is

$$
\nu_{k+1} = 2d + 2 + (\nu_{k|k} - 2d - 2) |I_d + Q_{k+1} V_{k|k}|^{-\frac{1}{2}}.
$$
(35)

In accordance with [14], the above setting ensures that the volume of the expected value of the extent matrix is preserved for all $Q_{k+1}$. Nevertheless, the volume is still dependent upon the matrix transformation $M_{x_k}$. Specifically, by Lemma 9\footnote{By [24], the matrix transformation $M_{x_k}$ is a function of the turn-rate $\omega_k$ and prediction time interval $T$. Therefore, no covariance element other than turn-rate variance will impact the expectations $C_1$, $C_2$, and $C_3$.}

$$
\text{Vol}(\mathbb{E}[X_{k+1} | Z^k]) = \text{Vol}(\int M_{x_k} \mathbb{E}[X_{k} | Z^k] H M_{x_k}^T N(x_k | m_{k|k}, P_{k|k}) dx_k),
$$
(36)

where $H \triangleq \text{diag}(I_d + Q_{k+1} V_{k|k})^{-1/2} |I_d + Q_{k+1} V_{k|k}|^{-1/2} \in \mathbb{S}_+(d, \mathbb{R})$.

From (36), it can be seen that the variance of the kinematic state vector influences the volume of the expected value. As the variance increases, so will the expected volume. The setting is thereby suitable for modelling changes in target size that are dependent upon the kinematic state vector. For example, the spread of a truck convoy may be dependent upon the convey

| TABLE II: Proposed Prediction Update |
|---|
| **Input:** Previous kinematic state estimate, covariance, degrees of freedom, and parameter matrix $\{m_{k|k}, P_{k|k}, \nu_{k|k}, V_{k|k}\}$. Kinematic state transition function and covariance $\{f_{k+1}, D_{k+1}\}$. Extent matrix state transition parameters and transformation $\{\nu_{k+1}, Q_{k+1}, M_{x_k}\}$. |
| **Output:** $\{m_{k+1|k}, P_{k+1|k}, V_{k+1|k}\}$. |

$$
m_{k+1|k} = f_{k+1}(m_{k|k}),
P_{k+1|k} = F_{k+1} P_{k|k} F_{k+1}^T + D_{k+1},
$$
$$
\nu_{k+1|k} = 2d + 2 + (\nu_{k|k} - 2d - 2) |I_d + Q_{k+1} V_{k+1|k}|^{-\frac{1}{2}} - \rho_{k+1},
$$
$$
V_{k+1|k} = \left( \frac{\nu_{k+1|k} - d - 1}{\nu_{k+1|k} - d - 1} \right) C_1 C_2^{-1} - \rho_{k+1},
$$
$$
\rho_{k+1} = \nu_{k+1|k} - 2d - 2,
$$
$$
F_{k+1} = \nabla_{x_k} f_{k+1}(x_k) |_{x_k = m_{k|k}},
$$
$$
C_1 = \int (M_{x_k} \tilde{V}_{k+1} M_{x_k}^T)^{-1} N(x_k | m_{k|k}, P_{k|k}) dx_k,
$$
$$
C_2 = \int M_{x_k} \tilde{V}_{k+1} M_{x_k}^T N(x_k | m_{k|k}, P_{k|k}) dx_k.
$$

By the same token, if $M_{x_k}$ is functionally dependent upon the posterior parameters $V_{k|k}$ and $V_{k|k}$, the volume remains constant over the inverse Wishart approximation.
speed or acceleration. Assuming $M_{x_k}$ is an appropriate scale matrix, equation (35) ensures that the size of the truck convoy is adequately updated based upon the current estimate of the kinematic state. Similarly, the shape of the truck convoy can be adequately modified through the noise matrix $Q_{k+1}$ to handle any unforeseen transformations of the group extent. We refer the reader to [14, Section IV-D] for further details on the selection of noise matrix $Q_{k+1}$ to model practical phenomena.

In contrast to the above discussion, the volume of an extended target may be time invariant. For example, the volume of a surface vessel or ground vehicle will not change over time. Hence, it would be desirable to ensure the volume of the expected value of the extent matrix is preserved from time $t_k$ to $t_{k+1}$. This gives motivation for the following setting of the state transition degrees of freedom:

$$v_{k+1} = 2d + 2 + (v_{k|k} - 2d - 2)|C_{2}^{-1}V_{k|k}|^{-\frac{1}{2}}.$$  (37)

The above setting ensures that the volume of the expected value of the extent matrix is preserved for all $Q_{k+1}$ and $M_{x_k}$. That is,

$$\text{Vol}(\mathbb{E}[X_{k+1}|Z^k]) = \text{Vol}(\mathbb{E}[X_k|Z^k]).$$  (38)

Given setting (37), it is assumed that the kinematic state vector only influences the shape and orientation of the target extent. All influences in size are neglected. We remark here that any unforeseen evolutions in target shape can still be adequately modelled by the noise matrix $Q_{k+1}$. Furthermore, if $M_{x_k}$ is independent of $x_k$ and $|M_{x_k}| = 1$, (37) reduces to (35).

V. SIMULATIONS

This section presents several simulated results that compare the proposed prediction update to [13], [14]. The works of [12] have been excluded from this section, as the prediction update of [14] was shown to outperform [12] whenever the turn-rate is unknown to the observer; see [14, Section VII] for further discussion.

A single extended target followed the trajectory shown in Figure 2. The target extent was represented by an ellipsoid with diameters 50m and 16m respectively. The speed of the extended target was assumed constant at 30 m/s. At time-step $k = 18$, the extended target performed a constant turn manoeuvre with a turn-rate of $10^{\circ}/s$. Scattering measurement centers were uniformly distributed over the target extent $X_k$, and the measurement noise was assumed to be zero-mean Gaussian distributed with covariance $R_k = \text{diag}(1.5^2, 1.5^2)$. The number of measurements at each time step was Poisson distributed with mean 10, and the scan rate was $T = 1s$.

Herein, the prediction of [14] is denoted by M1. A white noise acceleration model was used to update the kinematic state vector; which contained the Cartesian position and velocity of the extended target. That is, $x_k = [x_k, y_k, \dot{x}_k, \dot{y}_k]^T$. We remark here that M1 assumes a conditional random matrix model, therefore, unlike its competitors, the turn-rate of the extended target cannot be included within the kinematic state vector. The three discrete models of M1 were chosen to adopt...
Furthermore, for M3, the state transition degrees of freedom denoted the standard deviation of the target acceleration.

It was assumed no prior information was known in regard to the turn-rate manoeuvre. Thus, the transition matrix $M_k$ was chosen to be 

$$M_{k+1} = \begin{cases} 1 & \text{if } \theta = 0^\circ, \\
\frac{1}{\cos(\theta)} & \text{otherwise} \end{cases}$$

with $\sigma_{k+1} = 1$ and $\theta = 0^\circ$ for each model. Henceforth, the estimators using the prediction update of [13] and the proposed prediction update are denoted by M2 and M3 respectively. To remove any potential bias, the constant turn model was used to update the kinematic state vector of M2 and M3; which contained the Cartesian position, velocity, and turn-rate of the extended target. That is, $x_k = [x_k, y_k, \dot{x}_k, \dot{y}_k, \omega_k]^T$. The parameters of the two estimators were chosen as follows:

1) M2: Medium kinematic and extent process noise; $\sigma_a = 2.0$, $\sigma_\omega = 0.1^\circ$, and $n_{k+1} = 30$.
2) M3: Medium kinematic and extent process noise; $\sigma_a = 2.0$, $\sigma_\omega = 0.1^\circ$, and $Q_{k+1} = 0.333V_{k+1}^{-1}$,

where $\sigma_a$ and $\sigma_\omega$ denote the standard deviation of the target acceleration and turn-rate respectively; see [31] for details. Furthermore, for M3, the state transition degrees of freedom $v_{k+1}$ was set according to [37].

4Here, $\bar{\sigma}$ denotes the process noise intensity, and is required when using the discrete-time equivalent white noise acceleration model; see [11] for details. In accordance with Feldmann et al. [11], we define $\bar{\sigma} \triangleq 0.75\sigma_a$, where $\sigma_a$ denotes the standard deviation of the target acceleration.

Fig. 3: The Gaussian Wasserstein distance (left) and ANEES (right) over 900 Monte Carlo runs. The red dashed line is approach M1, the blue dash-dot line is approach M2 and the yellow solid line is approach M3.

To evaluate the tracking performance of each estimator, we exploit two credibility measures over $N = 900$ Monte Carlo runs. The first credibility measure is the Gaussian Wasserstein distance $d_{GW}$, which is the recommended metric for comparing elliptical extended targets [17], [39]. The square of the Gaussian Wasserstein distance is given by:

$$d_{GW}^2 = \frac{1}{N} \sum_{i=1}^{N} \left( \text{Tr}(X_k + \tilde{X}_{k|k} - 2\tilde{X}_k) + ||m_{k|k} - x_k||^2 \right), \quad (39)$$

where $\tilde{X}_{k|k} = V_{k|k}/(\nu_{k|k} - 2d - 2)$, and $\tilde{X}_k = (X_k^T \tilde{X}_{k|k} X_k^T)^{1/2}$.

The second credibility measure is the average normalised estimation error squared (ANEES), which measures how confident an estimator is in its estimation quality [11]. Values greater than one indicate that the estimator is overly confident, whilst values less than one indicate that the estimator is too pessimistic. The ANEES of the kinematic state vector and the extent matrix are calculated as follows:

$$\text{ANEES}_X = \frac{1}{Nn_x} \sum_{i=1}^{N} (m_{k|k} - x_k)^T P_{k|k} (m_{k|k} - x_k), \quad (40a)$$

$$\text{ANEES}_X = \frac{1}{N} \sum_{i=1}^{N} \text{Tr}((X_{k|k} - X_k)^2)/e_{k|k}, \quad (40b)$$

where $e_{k|k} \triangleq \text{Tr}([\text{Var}(X_{k|k}|Z_k)])^{-1/2}$.

Figure 3 shows the results of each estimator. Due to the use of a white noise acceleration model, M1 offers the best target tracking performance during the constant velocity motion. Nevertheless, during the turn-rate manoeuvre, both M2 and M3 offer significant improvement in target tracking performance than M1. This performance increase can be attributed to two key points. Firstly, the use of a constant-turn model to...
update the kinematic state vector, and secondly, the use of the matrix transformation $M_{x_k}$ [4] to update the orientation of target extent. Moreover, the ANEES of the target extent is significantly larger for M1 than for M2 and M3; particularly for the target extent.

**B. Variable Turn Manoeuvre**

In this simulation, the extended target followed the trajectory presented in Figure 4. As this trajectory involves long sequences of variable and constant turn-maneuvres, we have restricted our simulation to compare only the factorised estimators M2 and M3. The parameters of each estimator were set to the same values as in the previous simulation.

Figure 5 shows the results of each estimator. Interestingly, M3 possesses a lower Gaussian Wasserstein distance than M2, with a maximum relative difference of $\approx 7\%$ (left). This could be a benefit of using a single Kullback-Leibler divergence minimisation as opposed to several, or due to preserving the volume of the expected extent matrix via the state transition degrees of freedom $v_{k+1}$. In regard to the ANEES, both estimators appear to be too over-confident in their estimation quality of the kinematic state vector (top right); particularly in the last sequence of constant velocity and constant turn motions. For the target extent, the ANEES of both estimators never exceeds a value of one. We also note that the ANEES of M3 is much larger than M2; with a maximum relative difference of $\approx 32.7\%$ (bottom right). This indicates that M3 is less pessimistic than M2, and therefore possesses a better level of confidence in its estimation quality of the target extent.

**VI. CONCLUDING REMARKS**

In this contribution, we generalised the prediction update presented [14] to enable for kinematic state evolutions of the target extent. In contrast to [13], the newly proposed update required the use of only a single Kullback-Leibler divergence minimisation, and offers an additional tuning parameter to model target shape uncertainties—the noise matrix $Q_{k+1}$. To avoid numerical root-finding, we additionally provided a closed-form solution for the predicted degrees of freedom $v_{k+1|k}$. Further, we presented two alternatives on the selection of the state transition degrees of freedom $v_{k+1}$ to model practical phenomena. Simulated results indicate that the proposed prediction update enables for improved target tracking performance than previous works.

Although not utilised to its full potential, the noise matrix can be utilised to model transformations of the target extent. Future research should be pushed in this direction to determine other potential structures of $Q_{k+1}$ which can aid in the tracking of extended and group targets.

**APPENDIX**

**Lemma 1.** Let the non-Markov state transition density $p(X_{k+1}|X_k, Z^k)$ be the non-central inverse Wishart distribution described in [19]. Then, the state transition model governing the evolution of $X_k$ to $X_{k+1}$ is

$$X_{k+1}^{-\frac{1}{2}} = M_{x_k}^{-T} \left( X_k^{-\frac{1}{2}} + n_{k+1} \frac{1}{2} W_{k+1}^2 \right),$$

$$W_{k+1} \sim \mathcal{W}_d \left( W_{k+1|n_{k+1}, \frac{Q_{k+1}}{n_{k+1}}} \right),$$

where $n_{k+1} \triangleq v_{k+1} - d - 1$.

**Proof.** Let $X_{k+1}^{-1} = Y_k Y_k^T$, where $Y_k$ is a real $d \times n_k$ matrix with dimension $n_k \triangleq v_{k|k} - d - 1$. By [14, Lemma 2],

$$X_k^{-1} = Y_k Y_k^T = (Y_k H_k^T) (Y_k H_k^T)^T,$$

where $H_k \in \mathcal{O}_{n_k|d|}^{n_k + 1 \times n_n}$ with $n_{k+1} \triangleq v_{k+1} - d - 1$. Similarly, let $X_{k+1}^{-1} = Y_{k+1} Y_{k+1}^T$, where $Y_{k+1}$ is a real $d \times n_{k+1}$ matrix. Then, using [14, Lemma 3], (19) is equivalent to

$$p(Y_{k+1}|X_k, Z^k) = \mathcal{N}_{d,n_{k+1}} \left( Y_{k+1} | M_{x_k} Y_k H_k^T, \Sigma_{x_k}^{-1} \otimes I_{n_{k+1}} \right),$$

where $\Sigma_{x_k} = M_{x_k} Q_{k+1} M_{x_k}^T$. The state transition model governing (43) is given by:

$$Y_{k+1} = M_{x_k}^{-T} \left( Y_k H_k^T + \frac{1}{2} U_{k+1} \right),$$

$$U_{k+1} \sim \mathcal{N}_{d,n_{k+1}} \left( U_{k+1} | 0_{d,n_{k+1}}, \frac{Q_{k+1}}{n_{k+1}} \otimes I_{n_{k+1}} \right).$$

Define $X_{k+1}^{-\frac{1}{2}} = Y_k H_k^T$, $X_{k+1}^{-\frac{1}{2}} = Y_{k+1}$, and $W_{k+1}^2 = U_{k+1}$. Then, according to [23, Theorem 3.2.2], the state transition model governing the evolution of $X_k$ to $X_{k+1}$ is given by:

$$X_{k+1}^{-\frac{1}{2}} = M_{x_k}^{-T} \left( X_k^{-\frac{1}{2}} + \frac{1}{2} W_k^2 \right),$$

$$W_{k+1} \sim \mathcal{W}_d \left( W_{k+1|n_{k+1}, \frac{Q_{k+1}}{n_{k+1}}} \right),$$

which is the desired result.

**Lemma 2.** Consider the probability density function

$$p(X) = \int \mathcal{D} \mathcal{W}_d \left( X \mid v, M_X V M_X^T \right) N'(x \mid m, P) dx,$$

where $M_X \triangleq M(x)$ such that $M : \mathbb{R}^n \rightarrow \mathbb{R}^{d \times d}$ is a non-singular matrix-valued function. Moreover, let $(\nu^*, V^*)$ denote the parameter set that minimises the Kullback-Leibler
Proof. By equation (27),

\[
\begin{align*}
(\nu^*, V^*) &= \arg \min_{(\nu, V)} \text{KL}(p(X) || \mathcal{IW}_d(X | \nu, V)), \\
\nu^* &= \arg \max_{(\nu, V)} \int p(X) \ln (\mathcal{IW}_d(X | \nu, V)) \, dX,
\end{align*}
\]

and

\[
\begin{align*}
\nu^* &= \arg \max \int p(X) \ln (\mathcal{IW}_d(X | \nu, V)) \, dX,
\end{align*}
\]

where the objective function

\[
\begin{align*}
f(\nu, V) &= \frac{\nu - d - 1}{2} \left( \ln (|V|) - d \ln(2) \right) - \frac{\nu}{2} \mathbb{E} \left[ \ln (|X|) \right] - \frac{1}{2} \text{tr}(V \mathbb{E}[X^{-1}]) - \ln \left( \Gamma_d \left( \frac{\nu - d - 1}{2} \right) \right).
\end{align*}
\]

And

\[
\begin{align*}
\nabla f(\nu, V) &= \left[ \frac{\partial f(\nu, V)}{\partial \text{vec}(V)^T} \quad \frac{\partial f(\nu, V)}{\partial \nu} \right]^{T}_{(\nu^*, V^*)} = 0. \quad (53)
\end{align*}
\]

Using the notation of Magnus and Neudecker [40], the derivative of the objective function with respect to \( \text{vec}(V) \)

\[
\begin{align*}
\frac{\partial f(\nu, V)}{\partial \text{vec}(V)} &= \frac{\nu - d - 1}{2} \frac{1}{\text{vec}(V)} \ln (|V|) - \frac{1}{2} \frac{\partial \text{tr}(V \mathbb{E}[X^{-1}])}{\partial \text{vec}(V)} \\
&= \frac{\nu - d - 1}{2} \text{vec}(V^{-1}) - \frac{1}{2} \text{vec}(\mathbb{E}[X^{-1}]). \quad (54a)
\end{align*}
\]

Setting (54b) to zero and using Lemma 3

\[
V^* = \left( \frac{\nu^* - d - 1}{v - d - 1} \right) C_1^{-1}, \quad (55)
\]

which is the desired result (48). Similarly, the derivative of the objective function with respect to \( \nu \) is:

\[
\begin{align*}
\frac{\partial f(\nu, V)}{\partial \nu} &= \frac{1}{2} \left( \ln (|V|) - d \ln(2) - \mathbb{E} \left[ \ln (|X|) \right] \right) - \frac{\partial}{\partial \nu} \frac{\ln \left( \Gamma_d \left( \frac{\nu - d - 1}{2} \right) \right)}{2}, \quad (56)
\end{align*}
\]

where the objective function

\[
\begin{align*}
f(\nu, V) &= \frac{\nu - d - 1}{2} \left( \ln (|V|) - d \ln(2) \right) - \frac{\nu}{2} \mathbb{E} \left[ \ln (|X|) \right] - \frac{1}{2} \text{tr}(V \mathbb{E}[X^{-1}]) - \ln \left( \Gamma_d \left( \frac{\nu - d - 1}{2} \right) \right).
\end{align*}
\]

The first-order necessary condition for optimality requires that the gradient of the objective function (52) will be zero. Therefore, \((\nu^*, V^*)\) must satisfy
which, through the use of Lemma 4 and Theorem 1.4.1, is equivalent to

\[
\frac{\partial f(\nu, V)}{\partial \nu} = \frac{1}{2} \left( \ln |V| + \sum_{i=1}^{d} \psi_0 \left( \frac{\nu - d - i}{2} \right) - \sum_{i=1}^{d} \psi_0 \left( \frac{\nu - d - i}{2} - C_3 \right) \right). \tag{57}
\]

Substitution of (55) into (57) and solving for zero yields (49).

To complete the proof, we must now show that the objective function is concave, and that \((\nu^*, V^*)\) is the unique global maximum. For the objective function to be concave, the hessian matrix, denoted by \(\nabla^2 f(\nu, V)\), must be symmetric negative definite, or conversely \(-\nabla^2 f(\nu, V)\) is symmetric positive definite if and only if

\[
\nabla^2 f(\nu, V) = \begin{bmatrix}
\frac{\partial^2 f(\nu, V)}{\partial \nu \partial \text{vec}(V)^T} & \frac{\partial^2 f(\nu, V)}{\partial \text{vec}(V) \partial \nu} \\
\frac{\partial^2 f(\nu, V)}{\partial \text{vec}(V) \partial \nu} & \frac{\partial^2 f(\nu, V)}{\partial \nu \partial \text{vec}(V)^T}
\end{bmatrix} \tag{58}
\]

Using equations (54b), (57) and Section 8.4, the negative of the hessian matrix is equal to

\[
-\nabla^2 f(\nu, V) = \begin{bmatrix}
\frac{(V^{-1} \otimes V^{-1})}{2(\nu - d - 1)^{-1}} & \frac{1}{4} \sum_{i=1}^{d} \psi_0 \left( \frac{\nu - d - i}{2} \right) \\
\frac{1}{4} \sum_{i=1}^{d} \psi_0 \left( \frac{\nu - d - i}{2} \right) & \frac{2}{\nu - d - 1} - \frac{1}{2} \text{vec}(V^{-1})^T \text{vec}(V^{-1})
\end{bmatrix}, \tag{59}
\]

where the Kronecker product is symmetric positive definite for all \(\nu > 2d\) and \(V \in \mathbb{S}^{d+1}_{++}\). Hence, using the Schur complement\(^7\) \(-\nabla^2 f(\nu, V)\) is symmetric positive definite if and only if

\[
\sum_{i=1}^{d} \psi_0 \left( \frac{\nu - d - i}{2} \right) > \frac{2}{\nu - d - 1} - \frac{1}{2} \text{vec}(V^{-1})^T \text{vec}(V^{-1}), \tag{60}
\]

which holds true by Corollary 1. Hence, the objective function is concave, and \((\nu^*, V^*)\) is the global minimum of the Kullback-Leibler divergence between \(p(X)\) and all \(\mathcal{IW}_d\) distributions. Moreover, this minimum is unique. \(\square\)

**Lemma 3.** Let \(p(X)\) be the probability density function described in (46). Then,

\[
\mathbb{E} [X^{\nu-1}] = (\nu - d - 1) C_1, \tag{61}
\]

where \(C_1\) denotes the expectation

\[
C_1 = \int (M_x \bar{V} M_x^T)^{-1} \mathcal{N}(x|\mathbf{m}, P) dx. \tag{62}
\]

**Proof.** Let \(X = W^{-1}\). By Theorem 3.4.1,

\[
\mathbb{E} [W] = \int W p(W) dW, \tag{63a}
\]

\[
\mathbb{E} [W] = \int \int W \mathcal{IV}_d(W) |\nu - d - 1, (M_x \bar{V} M_x^T)^{-1} | \mathcal{N}(x|\mathbf{m}, P) dW dx,
\]

\[
= \int (\nu - d - 1) (M_x \bar{V} M_x^T)^{-1} \mathcal{N}(x|\mathbf{m}, P) dx, \tag{63c}
\]

where the last equality is a consequence of Theorem 3.3.15. Substitution of (62) yields the desired result. \(\square\)

**Lemma 4.** Let \(p(X)\) be the probability density function described in (46). Then,

\[
\mathbb{E} [\ln |X|] = C_3 - d \ln(2) - \sum_{i=1}^{d} \psi_0 \left( \frac{\nu - d - i}{2} \right), \tag{64}
\]

where \(C_3\) denotes the integral

\[
C_3 = \int \ln \left( (M_x \bar{V} M_x^T)^{-1} \right) \mathcal{N}(x|\mathbf{m}, P) dx. \tag{65}
\]

**Proof.** By the law of the unconscious statistician [43, pg. 274],

\[
\mathbb{E} [\ln |X|] = \int \ln |X| p(X) dx,
\]

\[
= \int \ln |X| \mathcal{IW}_d(X|\nu, M_x \bar{V} M_x^T) dX. \tag{66b}
\]

Thus, in order to prove (64), we must first find an analytical expression for the following intermediate expectation:

\[
\mathbb{E}_{\mathcal{IW}} [\ln |X|] \triangleq \int \ln |X| \mathcal{IW}_d(X|\nu, M_x \bar{V} M_x^T) dX. \tag{67}
\]

Using the moment generating function, (67) is equivalent to the following partial derivative:

\[
\mathbb{E}_{\mathcal{IW}} [\ln |X|] = \frac{\partial}{\partial s} \mathbb{E}_{\mathcal{IW}} [X^s] \bigg|_{s=0}, \tag{68}
\]

where the expectation

\[
\mathbb{E}_{\mathcal{IW}} [X^s] = \int |X|^s \mathcal{IW}_d(X|\nu, M_x \bar{V} M_x^T) dX, \tag{69a}
\]

\[
= \frac{\Gamma_d \left( \frac{\nu - 2s - d - 1}{2} \right)}{\Gamma_d \left( \frac{\nu - d - 1}{2} \right)} |M_x \bar{V} M_x^T|^s. \tag{69b}
\]

Substituting (69b) into (68), and using Lemmas 7 and 8

\[
\mathbb{E}_{\mathcal{IW}} [\ln |X|] = \ln \left( \frac{|M_x \bar{V} M_x^T|^s}{2d} \right) - \sum_{i=1}^{d} \psi_0 \left( \frac{\nu - d - i}{2} \right). \tag{70}
\]

Substitution of (70) into (66b) yields (64). \(\square\)

**Lemma 5.** Let \(V \in \mathbb{S}^{d+1}_{++}\) where \(d \in \mathbb{N}\). Then,

\[
\text{vec}(V^{-1})^T (V \otimes V) \text{vec}(V^{-1}) = d. \tag{71}
\]

**Proof.** From Theorem 1.2.22

\[
\text{vec}(V^{-1})^T (V \otimes V) \text{vec}(V^{-1}) = \text{vec}(V^{-1})^T \text{vec}(V). \tag{72}
\]

Moreover, by Theorem 1.2.10, there exists an orthogonal matrix \(H_1 \in \mathbb{O}^{d \times d}\) and diagonal matrix \(\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_d)\) such that \(V = H_1 \Lambda H_1^T\). Therefore,

\[
\text{vec}(V^{-1})^T \text{vec}(V) = \text{vec}(H_1^{-1} H_1^T)^T \text{vec}(H_1 \Lambda H_1^T), \tag{73a}
\]

\[
= \text{vec}(\Lambda^{-1} H_1^T (I_d \otimes I_d) \text{vec}(\Lambda), \tag{73b}
\]

\[
= \sum_{i=1}^{d} \lambda_i \tag{73c}
\]

\[
= d. \tag{73d}
\]
Substitution of (73d) into (72) yields (71).

**Lemma 6.** Let \( \nu > 2d \) where \( d \in \mathbb{N} \). Then,
\[
\sum_{i=1}^{d} \psi_{1}\left(\frac{\nu-d-i}{2}\right) > \frac{2d}{\nu-d-1}. \tag{74}\]

**Proof.** Consider the following upper bound:
\[
\sum_{i=1}^{d} \frac{2}{\nu-d-i} > \frac{2d}{\nu-d-1}. \tag{75}\]
Thus, rather then solving (74) directly, it suffices to prove
\[
\sum_{i=1}^{d} \left(\psi_{1}\left(\frac{\nu-d-i}{2}\right) - \frac{2}{\nu-d-i}\right) > 0. \tag{76}\]
A sufficient condition for the inequality to hold is for the difference within the summation to remain positive. That is, for the inequality
\[
\psi_{1}(a_i) - \frac{1}{a_i} > 0 \tag{77}\]
to hold true for all \( a_i = \frac{1}{2}(\nu-d-i) \), where \( i = 1, \ldots, d \). By substitution of recurrence relation \( \psi_{1}(a_i+1) = \psi_{1}(a_i) - a_i^{-2} \) into Lemma 10,\[
\psi_{1}(a_i) > \frac{1}{a_i} + \frac{2a_i + 1}{2a_i^2 + 2a_i + \frac{\pi}{4}} > \frac{1}{a_i}, \quad \forall a_i > 0. \tag{78}\]
Hence, inequality (77) holds for all \( \nu > 2d \). □

**Corollary 1.** Let \( V \in S_{d+\nu}^{\nu} \) and \( \nu > 2d \) where \( d \in \mathbb{N} \). Then,
\[
\sum_{i=1}^{d} \psi_{1}\left(\frac{\nu-d-i}{2}\right) > \frac{2}{\nu-d-1} \times \text{vec}(V^{-1})^T(V \otimes V)\text{vec}(V^{-1}). \tag{79}\]

**Proof.** Substitute Lemma 5 into Lemma 6. □

**Lemma 7.** Let the scalar \( \nu > 2s + 2d \), where \( s \in \mathbb{R} \) and \( d \in \mathbb{N} \). Then,
\[
\frac{\partial}{\partial s} \Gamma_{d}\left(\frac{\nu-2s-d-1}{2}\right) = -\Gamma_{d}\left(\frac{\nu-2s-d-1}{2}\right) \times \sum_{i=1}^{d} \psi_{0}\left(\frac{\nu-2s-d-i}{2}\right). \tag{80}\]

**Proof.** In accordance with logarithmic differentiation,
\[
\frac{\partial}{\partial s} \Gamma_{d}\left(\frac{\nu-2s-d-1}{2}\right) = \Gamma_{d}\left(\frac{\nu-2s-d-1}{2}\right) \times \frac{\partial}{\partial s} \ln \left[\Gamma_{d}\left(\frac{\nu-2s-d-1}{2}\right)\right]. \tag{81}\]
where, via [23] Theorem 1.4.1,
\[
\frac{\partial}{\partial s} \ln \left[\Gamma_{d}\left(\frac{\nu-2s-d-1}{2}\right)\right] = \frac{d(d-1)}{4} \ln(\pi) + \sum_{i=1}^{d} \frac{\partial}{\partial s} \ln \left[\Gamma\left(\frac{\nu-2s-d-i}{2}\right)\right], \tag{82a}\]
\[
= -\sum_{i=1}^{d} \psi_{0}\left(\frac{\nu-2s-d-i}{2}\right). \tag{82b}\]

Substitution of (82b) into (81) yields (80). □

**Lemma 8.** Let \( s \in \mathbb{R} \) and \( V \in S_{d+\nu}^{\nu} \) where \( d \in \mathbb{N} \). Then,
\[
\frac{\partial}{\partial s} \left[\frac{M_{X}V M_{X}^T}{2ds}\right] = \ln \left[\frac{M_{X}V M_{X}^T}{2ds}\right] \times \frac{M_{X}V M_{X}^T}{2ds}. \tag{83}\]

**Proof.** Given \( \alpha^s = \exp(s \ln(\alpha)) \),
\[
\frac{\partial}{\partial s} \left[\frac{M_{X}V M_{X}^T}{2ds}\right] = \frac{\partial}{\partial s} \exp \left(s \ln \left[\frac{M_{X}V M_{X}^T}{2ds}\right]\right), \tag{84a}\]
\[
= \ln \left[\frac{M_{X}V M_{X}^T}{2ds}\right] \times \frac{M_{X}V M_{X}^T}{2ds}. \tag{84b}\]

**Lemma 9.** Let the state transition degrees of freedom \( v_{k+1} \) be defined by (35). Then,
\[
\text{Vol\{\mathbb{E}[X_{k+1}|Z^k]\}} = \text{Vol} \left(\int M_{X,k} \mathbb{E}[X_{k}|Z^k] H M_{X,k}^T N(X_k|m_{k|k},P_{k|k})dx_k\right), \tag{85}\]
where \( H = (I_d + Q_{k+1}V_{k|k})^{-1}(I_d + Q_{k+1}V_{k|k})^{-1} \subseteq S_{d+\nu}(\mathbb{R}) \).

**Proof.** Let \( \mathbb{E}[X_{k+1}|Z^k] \) denote the expected value of (26a). In accordance with [23] Theorem 3.4.3,
\[
\mathbb{E}[X_{k+1}|Z^k] = \int M_{X,k} \bar{V}_{k+1} M_{X,k}^T N(X_k|m_{k|k},P_{k|k})dx_k. \tag{86}\]
However, from (26b) and (35),
\[
\overline{V}_{k+1} = \frac{V_{k|k}(I_d + Q_{k+1}V_{k|k})^{-1}}{v_{k+1}-2d-2} \tag{87a}\]
\[
= \mathbb{E}[X_{k}|Z^k] \frac{(I_d + Q_{k+1}V_{k|k})^{-1}}{v_{k+1} - v_{k+1} - 2d - 2}. \tag{87b}\]
Substituting (87b) into (86) and defining the special linear group matrix \( H = (I_d + Q_{k+1}V_{k|k})^{-1}(I_d + Q_{k+1}V_{k|k})^{-1} \subseteq S_{d+\nu}(\mathbb{R}) \), the expected value becomes:
\[
\mathbb{E}[X_{k+1}|Z^k] = \int M_{X,k} \mathbb{E}[X_{k}|Z^k] H M_{X,k}^T \times N(X_k|m_{k|k},P_{k|k})dx_k. \tag{88}\]
Finally, from Theorem 1, the following equality holds true:
\[
\text{Vol\{\mathbb{E}[X_{k+1}|Z^k]\}} = \text{Vol}(\mathbb{E}[X_{k+1}|Z^k]). \tag{89}\]
Substitution of (88) into (89) yields (85). □

**References**

[1] Y. Bar-Shalom, *Tracking and data association*. Academic Press Professional, Inc., 1987.

[2] K. Gilholm and D. Salmond, “Spatial distribution model for tracking extended objects,” *IEEE Proceedings-Radar, Sonar and Navigation*, vol. 152, no. 3, pp. 364–371, 2005.

[3] Y. Boers, H. Driessen, J. Torstensson, M. Triebs, R. Karlsson, and F. Gustafsson, “Track-before-detect algorithm for tracking extended targets,” *IEEE Proceedings-Radar, Sonar and Navigation*, vol. 153, no. 4, pp. 345–351, 2006.

[4] D. Angelova, L. Mihaylova, N. Petrov, and A. Gning, “A convolution particle filtering approach for tracking elliptical extended objects,” in *Proceedings of the 16th International Conference on Information Fusion*. IEEE, 2013, pp. 1542–1549.

[5] R. Mahler, “PHD filters for nonstandard targets: I. Extended targets,” 2009 12th International Conference on Information Fusion, no. July 2009, pp. 448–452, 2009.
