On Accumulated State Densities with Applications to Out-of-Sequence Measurement Processing

Wolfgang Koch
German Defence Research Establishment
Wachtberg, Germany
w.koch@fgan.de

Abstract – *In target tracking applications, the full information on the kinematic target states accumulated over a certain time window up to the present time is contained in the joint probability density function of these state vectors, given the time series of all sensor data. This joint density may also be called an Accumulated State Density (ASD) and provides a unified treatment of filtering and retrodiction insofar as by marginalizing the ASD, the standard filtering and retrodiction densities are obtained. In addition, ASDs fully describe the correlations between the state estimates produced for different instants of time. The paper discusses the notion of ASDs and closed formulae for calculating them. The practical usefulness of considering ASDs is illustrated by applications, where out-of-sequence (OoS) measurements are to be processed within the framework of a centralized measurement fusion architecture, i.e. when the sensor data do not arrive in the temporal order, in which they were produced. The approach can be applied to Kalman-, MHT-, and IMM filtering.*

Keywords: Accumulated target states, out-of-sequence measurements (OoS), Kalman/MHT/IMM filtering.

1 Introduction

Practical needs from both civilian and military applications as well as the rapid development of information and communications technology are driving factors for creating ever increasing sensor networks. To a degree never known before, decision makers in a net-centric world have access to vast amounts of data. For effective use of this information potential in real-world applications, however, the data streams must not overwhelm the decision makers involved. On the contrary, the data must be fused in such a way that high quality information for situation pictures results.

Situation pictures are produced by spatio-temporally processing various pieces of sensor information that in themselves often have only limited value for understanding the underlying situation. In this context, target 'tracks' are of particular importance [1, 2]. Tracking faces an omnipresent aspect in real-world application insofar as it is dealing with fusion of data produced at different instants of time; i.e. tracking is important in all applications where particular emphasis is placed on sensor data given by time series.

In most target tracking algorithms, the characteristics of conditional probability densities $p(X_l|Z^k)$ of target states X_l are calculated, which describe the available knowledge of the target properties at a certain instant of time t_l, given a time series Z^k of imperfect sensor data accumulated up to time t_k. In certain applications, however, the kinematic target states $\mathbf{x}_k, \dots, \mathbf{x}_m$, $m \leq k$, accumulated over a certain time window from a past instant of time t_m up to the present time t_k is of interest. The statistical properties of the accumulated state vectors are completely described the joint probability density function of them, , $p(\mathbf{x}_k, \dots, \mathbf{x}_m|Z^k)$, which is conditioned by the time series Z^k. These densities may be called *Accumulated State Densities (ASDs)*. By marginalizing them, the standard filtering and retrodiction densities directly result; in other words, ASDs provide a unified description of filtering and retrodiction. In addition, ASDs fully describe the correlations between the state estimates at different instants of time.

In [3], ASDs are considered to provide a more comprehensive treatment of issues in particle filtering. To some extent, the notion of ASDs might be considered as a step backwards insofar as in the very old days it was known that one could express a linear-Gaussian estimation problem in a joint ("accumulated") fashion, while Kalman's approach was a way to find a recursive solution. Nevertheless, the chief contribution of this paper is a *recursive* algorithm to find the parameters of an ASD.

ASDs are useful in tracking applications, where out-of-sequence (OoS) measurements are to be processed, i.e. when the sensor data do not arrive in the temporal order, in which they have been produced. The OoS problem is unavoidable in any real-world multiple sensor tracking application. To the author's knowledge, Y. Bar-Shalom was the first, who picked up the problem and provided an exact solution in the case of Kalman filtering [4]. For the subsequent development and generalizations see for example [5–9]. To avoid storing and reprocessing the entire time series of sensor data as well as to avoid the temporal delay connected to it, OoS measurements have to be inserted into the running tracking process in a particular way.

The proposed paper is organized as follows. Section 2 summarizes basic facts on the Bayesian tracking paradigm. In section 3, the notion of an Accumulated State Density is introduced. The discussion in section 4 provides closed-formulae for the parameters of the ASD in the case, where Kalman filtering can be applied to tracking. The concept of ASDs is generalized to multiple hypothesis and interaction multiple model trackers (MHT, IMM) in section 5. In section 6, ASDs are applied the out-of-sequence tracking problem. A summary concludes the paper. For the sake of notational simplicity we confine the discussion to the case of synchronous sensors. The generalization of the methodological framework to asynchronous sensors is straightforward.

2 Bayesian Tracking Approach

A Bayesian tracking algorithm is an iterative updating scheme for calculating conditional probability density functions $p(X_l|Z^k)$ that represent all available knowledge on the object states X_l at discrete instants of time t_l. The densities are explicitly conditioned by the sensor data Z^k accumulated up to some time t_k, typically the present time. Implicitly, they are also determined by context knowledge on the sensor characteristics, the dynamical object properties, the object environment, topographical maps, or tactical rules governing the objects' over-all behavior.

With respect on the instant of time t_l at which estimates of the object states X_l are required, the density iteration is referred to as *prediction* ($t_l > t_k$), *filtering* ($t_l = t_k$), or *retrodiction* ($t_l < t_k$). The propagation of the probability densities involved is given by three basic update equations.

2.1 Prediction

The prediction density $p(X_k|Z^{k-1})$ is obtained by combining the evolution model $p(X_k|X_{k-1})$ with the previous filtering density $p(X_{k-1}|Z^{k-1})$:

$$p(X_{k-1}|Z^{k-1}) \xrightarrow[\text{constraints}]{\text{evolution model}} p(X_k|Z^{k-1}) \tag{1}$$

$$p(X_k|Z^{k-1}) = \int \mathrm{d}X_{k-1}\, \underbrace{p(X_k|X_{k-1})}_{\text{evolution model}}\, \underbrace{p(X_{k-1}|Z^{k-1})}_{\text{previous filtering}}.$$

2.2 Filtering

The filtering density $p(X_k|Z^k)$ is obtained by combining the sensor model $p(Z_k|X_k)$, the "likelihood function", with the prediction density $p(X_k|Z^{k-1})$ according to:

$$p(X_k|Z^{k-1}) \xrightarrow[\text{sensor model}]{\text{current sensor data}} p(X_k|Z^k) \tag{2}$$

$$p(X_k|Z^k) = \frac{p(Z_k, m_k|X_k)\, p(X_k|Z^{k-1})}{\int \mathrm{d}X_k\, \underbrace{p(Z_k, m_k|X_k)}_{\text{sensor model}}\, \underbrace{p(X_k|Z^{k-1})}_{\text{prediction}}}.$$

2.3 Retrodiction

The retrodiction density $p(X_l|Z^k)$ is obtained by combining the object evolution model $p(X_{l+1}|X_l)$ with the previous prediction and filtering densities according to:

$$p(X_{l-1}|Z^k) \xleftarrow[\text{evolution model}]{\text{filtering, prediction}} p(X_l|Z^k) \tag{3}$$

$$p(X_l|Z^k) = \int \mathrm{d}X_{l+1} \frac{\overbrace{p(X_{l+1}|X_l)}^{\text{evolution}}\, \overbrace{p(X_l|Z^l)}^{\text{prev. filtering}}}{\underbrace{p(X_{l+1}|Z^l)}_{\text{prev. prediction}}}\, \underbrace{p(X_{l+1}|Z^k)}_{\text{prev. retrodiction}}.$$

Being the natural antonym of "prediction", the technical term "retrodiction" was introduced by Oliver Drummond in a series of papers [10]. Adopting the standard terminology [11], we could speak of *fixed-interval* retrodiction.

According to this paradigm, an *object track* represents all relevant knowledge on a time varying object state of interest, including its history and measures that describe the quality of this knowledge. As a technical term, 'track' is therefore either a synonym for the collection of densities $p(X_l|Z^k)$, $l = 1, \ldots, k, \ldots$, or of suitably chosen parameters characterizing them, such as estimates related to appropriate risk functions and the corresponding estimation error covariance matrices.

3 Accumulated State Densities

All information on the object states accumulated over a time window $t_k, t_{k-1}, \ldots, t_n$ of length $n + 1$,

$$X_{k:n} = (X_k, \ldots, X_n) \tag{4}$$

that can be extracted from the time series of accumulated sensor data Z^k up to and including time t_k is contained in a joint density function $p(X_{k:n}|Z^k)$, which may be called *Accumulated State Density (ASD)*. Via marginalizing over $X_k, \ldots, X_{l+1}, X_{l-1}, \ldots, X_n$,

$$p(X_l|Z^k) = \int \mathrm{d}X_k, \ldots, \mathrm{d}X_{l+1}, \mathrm{d}X_{l-1}, \ldots, \mathrm{d}X_n \; p(X_k, \ldots, X_n|Z^k), \tag{5}$$

the filtering density $p(X_k|Z^k)$ for $l = k$ and the retrodiction densities $p(X_l|Z^k)$ for $l < k$ result from the ASD. ASDs thus in a way unify the notions of filtering and retrodiction. In addition, ASDs also contain all mutual correlations between the individual object states at different instants of time. Bayes' Theorem provides a recursion formula for updating accumulated state densities:

$$p(X_{k:n}|Z^k) = \frac{p(Z_k, m_k|X_k)\, p(X_k|X_{k-1})\, p(X_{k-1:n}|Z^{k-1})}{\int \mathrm{d}X_{k-1}\, p(Z_k, m_k|X_k)\, p(X_k|X_{k-1})\, p(X_{k-1:n}|Z^{k-1})}. \tag{6}$$

The sensor data Z_k explicitly appear in this representation. A little formalistically speaking, 'sensor data processing' means nothing else than to achieve by certain reformulations that the sensor data are no longer be explicitly present.

4 Closed-form Representation

Under conditions, where Kalman filtering is applicable (perfect data sensor-data-to-track association, linear Gaußian sensor and evolution models), a closed-form representation of $p(\mathbf{x}_{k:n}|Z^k)$ can be derived. In this case, let the likelihood function be given by:

$$p(Z_k, m_k|X_k) = \mathcal{N}\big(\mathbf{z}_k;\, \mathbf{H}_k\mathbf{x}_k,\, \mathbf{R}_k\big), \qquad (7)$$

where $Z_k = \mathbf{z}_k$ denotes the vector of sensor measurements at time t_k, $X_k = \mathbf{x}_k$ the kinematic state vector of the object, $\mathbf{H}_k$ the measurement matrix, and $\mathbf{R}_k$ the measurement error covariance matrix, while the Markovian evolution model of the target is represented by:

$$p(X_k|X_{k-1}) = \mathcal{N}\big(\mathbf{x}_k;\, \mathbf{F}_{k|k-1}\mathbf{x}_{k-1},\, \mathbf{D}_{k|k-1}\big) \qquad (8)$$

with an evolution matrix $\mathbf{F}_{k|k-1}$ and a corresponding evolution covariance matrix $\mathbf{D}_{k|k-1}$.

A repeated use of a well-known product formula for Gaußians (see appendix, Eq. 55) directly yields a product representation of the augmented state density:

$$p(\mathbf{x}_{k:n}|Z^k) = \mathcal{N}\big(\mathbf{x}_k;\, \mathbf{x}_{k|k},\, \mathbf{P}_{k|k}\big) \times \prod_{l=n}^{k-1} \mathcal{N}\big(\mathbf{x}_l;\, \mathbf{h}_{l|l+1}(\mathbf{x}_{l+1}),\, \mathbf{R}_{l|l+1}\big), \qquad (9)$$

where the auxiliary quantities $\mathbf{h}_{l|l+1}$, $\mathbf{R}_{l|l+1}$, $l \leq k$, are defined by:

$$\mathbf{h}_{l|l+1}(\mathbf{x}_{l+1}) = \mathbf{x}_{l|l} + \mathbf{W}_{l|l+1}(\mathbf{x}_{l+1} - \mathbf{x}_{l+1|l}) \qquad (10)$$

$$\mathbf{R}_{l|l+1} = \mathbf{P}_{l|l} - \mathbf{W}_{l|l+1}\mathbf{P}_{l|l+1}\mathbf{W}_{l|l+1}^\top \qquad (11)$$

and a "retrodiction gain" matrix

$$\mathbf{W}_{l|l+1} = \mathbf{P}_{l|l}\mathbf{F}_{l+1|l}^\top\mathbf{P}_{l+1|l}^{-1}. \qquad (12)$$

Note that $\mathcal{N}\big(\mathbf{x}_l;\, \mathbf{h}_{l|l+1}(\mathbf{x}_{l+1}),\, \mathbf{R}_{l|l+1}\big)$ can be interpreted in analogy to a Gaußian likelihood function with a a linear measurement function $\mathbf{h}_{l|l+1}(\mathbf{x}_{l+1})$. $\mathbf{h}_{l|l+1}$, $\mathbf{R}_{l|l+1}$ are defined by the parameters of $p(\mathbf{x}_l|Z^l) = \mathcal{N}(\mathbf{x}_l;\, \mathbf{x}_{l|l},\, \mathbf{P}_{l|l})$,

$$\mathbf{x}_{l|l} = \begin{cases} \mathbf{x}_{l|l-1} + \mathbf{W}_{l|l-1}(\mathbf{z}_l - \mathbf{H}_l\mathbf{x}_{l|l-1}) \\ \mathbf{P}_{l|l}(\mathbf{P}_{l|l-1}^{-1}\mathbf{x}_{l|l-1} + \mathbf{H}_l^\top\mathbf{R}_l^{-1}\mathbf{z}_l) \end{cases} \qquad (13)$$

$$\mathbf{P}_{l|l} = \begin{cases} \mathbf{P}_{l|l-1} - \mathbf{W}_{l|l-1}\mathbf{S}_{l|l-1}\mathbf{W}_{l|l-1}^\top \\ (\mathbf{P}_{l|l-1}^{-1} + \mathbf{H}_l^\top\mathbf{R}_l^{-1}\mathbf{H}_l)^{-1}. \end{cases} \qquad (14)$$

Note that there exist two equivalent formulations of the Kalman update formulae according to the two versions of the product formula (Eq. 55). The innovation covariance matrix $\mathbf{S}_{l|l-1}$ and the *Kalman Gain* matrix are given by:

$$\mathbf{S}_{l|l-1} = \mathbf{H}_l\mathbf{P}_{l|l-1}\mathbf{H}_l^\top + \mathbf{R}_l. \qquad (15)$$

$$\mathbf{W}_{l|l-1} = \mathbf{P}_{l|l-1}\mathbf{H}_{l|l-1}^\top\mathbf{S}_{l|l-1}^{-1}. \qquad (16)$$

Also the parameters of the prediction density $p(\mathbf{x}_{l+1}|Z^l) = \mathcal{N}(\mathbf{x}_{l+1};\, \mathbf{x}_{l+1|l},\, \mathbf{P}_{l+1|l})$,

$$\mathbf{x}_{l|l-1} = \mathbf{F}_{l|l-1}\mathbf{x}_{l-1|l-1} \qquad (17)$$

$$\mathbf{P}_{l|l-1} = \mathbf{F}_{l|l-1}\mathbf{P}_{l-1|l-1}\mathbf{F}_{l|l-1}^\top + \mathbf{D}_{l|l-1}, \qquad (18)$$

enter into the product representation in Eq. 9. With $\mathbf{x}_{l|k}$, $\mathbf{P}_{l|k}$, $\mathbf{W}_{l|l+1}$ known from the Rauch-Tung-Striebel recursion,

$$\mathbf{x}_{l|k} = \mathbf{x}_{l|l} + \mathbf{W}_{l|l+1}\big(\mathbf{x}_{l+1|k} - \mathbf{x}_{l+1|l}\big) \qquad (19)$$

$$\mathbf{P}_{l|k} = \mathbf{P}_{l|l} + \mathbf{W}_{l|l+1}\big(\mathbf{P}_{l+1|k} - \mathbf{P}_{l+1|l}\big)\mathbf{W}_{l|l+1}^\top, \qquad (20)$$

we can rewrite $p(\mathbf{x}_{k:n}|Z^k)$ by the following product:

$$p(\mathbf{x}_{k:n}|Z^k) = \mathcal{N}\big(x_k;\, \mathbf{x}_{k|k},\, \mathbf{P}_{k|k}\big) \prod_{l=n}^{k-1} \mathcal{N}\big(x_l - \mathbf{W}_{l|l+1}x_{l+1};\, \mathbf{x}_{l|k} - \mathbf{W}_{l|l+1}\mathbf{x}_{l+1|k},\, \mathbf{Q}_{l|k}\big), \qquad (21)$$

where we used the abbreviation:

$$\mathbf{Q}_{l|k} = \mathbf{P}_{l|k} - \mathbf{W}_{l|l+1}\mathbf{P}_{l+1|k}\mathbf{W}_{l|l+1}^\top. \qquad (22)$$

Due to elementary matrix algebra manipulations it can be shown that this product can be represented by a single Gaußian,

$$p(\mathbf{x}_{k:n}|Z^k) = \mathcal{N}\big(\mathbf{x}_{k:n};\, \mathbf{x}_{k:n}^k,\, \mathbf{P}_{k:n}^k\big), \qquad (25)$$

with a joint expectation vector $\mathbf{x}_{k:n}^k$ defined by:

$$\mathbf{x}_{k:n}^k = (\mathbf{x}_{k|k}^\top, \mathbf{x}_{k-1|k}^\top, \ldots, \mathbf{x}_{n|k}^\top)^\top, \qquad (26)$$

while the corresponding joint covariance matrix $\mathbf{P}_{k:n}^k$ can be written as an inverse of a tridiagonal block matrix (see Eq. 23 on the subsequent page).

This can be seen by considering projectors $\mathbf{\Pi}_l$ defined by:

$$\mathbf{\Pi}_l\mathbf{x}_{k:n} = \begin{cases} (\mathbf{1}, \mathbf{O}, \ldots, \mathbf{O})\mathbf{x}_{k:n}, & l = k \\ (\mathbf{O}, \ldots, -\mathbf{W}_{l|l+1}, \mathbf{1}, \ldots, \mathbf{O})\mathbf{x}_{k:n}, & n \leq l < k \end{cases}$$

$$= \begin{cases} \mathbf{x}_k, & l = k \\ \mathbf{x}_l - \mathbf{W}_{l|l+1}\mathbf{x}_{l+1}, & n \leq l < k. \end{cases} \qquad (27)$$

Using $\mathbf{\Pi}_l$ and $\mathbf{Q}_{l|k}$, $l = 1, \ldots, k$, the ASD can be rewritten:

$$p(\mathbf{x}_{k:n}^k|Z^k) = \prod_{l=n}^{k} \mathcal{N}\big(\mathbf{\Pi}_l\mathbf{x}_{k:n};\, \mathbf{\Pi}_l\mathbf{x}_{k:n}^k,\, \mathbf{Q}_{l|k}\big) \qquad (28)$$

$$= \prod_{l=n}^{k} \mathcal{N}\big(\mathbf{x}_{k:n};\, \mathbf{x}_{k:n}^k,\, (\mathbf{\Pi}_l^\top\mathbf{Q}_{l|k}^{-1}\mathbf{\Pi}_l)^{-1}\big) \qquad (29)$$

$$= \mathcal{N}\big(\mathbf{x}_{k:n};\, \mathbf{x}_{k:m}^k,\, \mathbf{P}_{k:n}^k\big) \qquad (30)$$

$$\mathbf{P}_{k:n}^{k} = \begin{pmatrix} \mathbf{T}_{k|k} & -\mathbf{W}_{k-1|k}^{\top}\mathbf{Q}_{k-1|k}^{-1} & \mathbf{O} & \cdots & \mathbf{O} \\ -\mathbf{Q}_{k-1|k}^{-1}\mathbf{W}_{k-1|k} & \mathbf{T}_{k-1|k} & -\mathbf{W}_{k-2|k}^{\top}\mathbf{Q}_{k-2|k}^{-1} & \ddots & \vdots \\ \mathbf{O} & -\mathbf{Q}_{k-2|k}^{-1}\mathbf{W}_{k-2|k} & \ddots & \ddots & \mathbf{O} \\ \vdots & \ddots & \ddots & \mathbf{T}_{m+1|k} & -\mathbf{W}_{m|k}^{\top}\mathbf{Q}_{m|k} \\ \mathbf{O} & \cdots & \mathbf{O} & -\mathbf{Q}_{m|k}\mathbf{W}_{m|k} & \mathbf{T}_{m|k} \end{pmatrix}^{-1} \tag{23}$$

$$= \begin{pmatrix} \mathbf{P}_{k|k} & \mathbf{P}_{k|k}\mathbf{W}_{k-1|k}^{\top} & \mathbf{P}_{k|k}\mathbf{W}_{k-2|k}^{\top} & \cdots & \mathbf{P}_{k|k}\mathbf{W}_{1|k}^{\top} \\ \mathbf{W}_{k-1|k}\mathbf{P}_{k|k} & \mathbf{P}_{k-1|k} & \mathbf{P}_{k-1|k}\mathbf{W}_{k-2|k-1}^{\top} & * & \mathbf{P}_{k-1|k}\mathbf{W}_{1|k-1}^{\top} \\ \mathbf{W}_{k-2|k}\mathbf{P}_{k|k} & \mathbf{W}_{k-2|k-1}\mathbf{P}_{k-1|k} & \mathbf{P}_{k-2|k} & * & \vdots \\ \vdots & * & * & * & \mathbf{P}_{2|k}\mathbf{W}_{1|2}^{\top} \\ \mathbf{W}_{1|k}\mathbf{P}_{k|k} & \mathbf{W}_{1|k-1}\mathbf{P}_{k-1|k} & \cdots & \mathbf{W}_{1|2}\mathbf{P}_{2|k} & \mathbf{P}_{1|k} \end{pmatrix} \tag{24}$$

Block-matrix representations of the ASD covariance matrices.

with a covariance matrix $\mathbf{P}_{k:n}^{k}$, which is given by an harmonic mean according to the product formula for Gaußians:

$$\mathbf{P}_{k:n} = \Big(\sum_{l=n}^{k} \mathbf{\Pi}_l^{\top}\mathbf{Q}_{l|k}^{-1}\mathbf{\Pi}_l\Big)^{-1}. \tag{31}$$

The summation of the matrices $\mathbf{\Pi}_l^{\top}\mathbf{Q}_{l|k}^{-1}\mathbf{\Pi}_l$ directly yields the inverse ASD covariance matrix as a tridiagonal block matrix displayed on the top of page 3. Here, the auxiliary quantities $\mathbf{T}_{l|k}$, $m \leq l \leq k$ are defined by:

$$\mathbf{T}_{l|k} = \begin{cases} \mathbf{Q}_{n|k}^{-1} & \text{for } l = n \\ \mathbf{P}_{k|k}^{-1} + \mathbf{W}_{l-1|l}^{\top}\mathbf{Q}_{l-1|k}^{-1}\mathbf{W}_{l-1|l} & \text{for } l = k \\ \mathbf{Q}_{l|k}^{-1} + \mathbf{W}_{l-1|l}^{\top}\mathbf{Q}_{l-1|k}^{-1}\mathbf{W}_{l-1|l} & \text{else} \end{cases}.$$

The tridiagonal structure is a consequence of the Markov property of the underlying evolution model. This representation of the inverse of $\mathbf{P}_{k:n}^{k}$ is useful in calculations.

By a repeated use of the matrix inversion lemma (see appendix) and an induction argument, the inverse of this tridiagonal block matrix can be calculated. The resulting block matrix is displayed in Eq. 24 on the subsequent page, where the following abbreviations were used:

$$\mathbf{W}_{l|k} = \prod_{\lambda=l}^{k-1} \mathbf{W}_{\lambda|\lambda+1} = \prod_{\lambda=l}^{k-1} \mathbf{P}_{\lambda|\lambda}\mathbf{F}_{\lambda+1|\lambda}^{\top}\mathbf{P}_{\lambda+1|\lambda}^{-1}. \tag{32}$$

The densities $\{\mathcal{N}(\mathbf{x}_l;\, \mathbf{x}_{l|k},\, \mathbf{P}_{l|k})\}_{l=m}^{k}$ are directly obtained via marginalizing, since the covariance matrices $\mathbf{P}_{l|k}$, $n \leq l \leq k$, appear on the diagonal of this block matrix. Note that the ASD is completely defined by the results of prediction, filtering, and retrodiction obtained for the time window $t_k, \dots, t_n$, i.e. it is a by-product for Kalman filtering and Rauch-Tung-Striebel smoothing.

5 Discussion of Generalizations

These considerations can be generalized to the case of ambiguity with respect to the origin of the sensor data or with respect to the evolution model, which is currently in effect, i.e. to Multiple Hypothesis Tracking (MHT) and Interacting Multiple Model filters (IMM).

5.1 ASDs for MHT Filtering

A sensor output at time t_k, consisting of m_k measurements collected in the set Z_k, can be ambiguous, i.e. the origin of the sensor data has to be explained by a set of data interpretations, which are assumed to be exhaustive and mutually exclusive. As an example, let us consider measurements $Z_k = \{\mathbf{z}_k^j\}_{j=1}^{m_k}$ possibly related to the kinematic state $\mathbf{x}_k$ of well-separated objects. 'Well-separated' here means that measurements potentially originated by one object could not have been originated by another. Even in this simplified situation, ambiguity can arise from imperfect detection and due to false measurements, often referred to as *clutter*, or of measurements from unwanted objects. Let the detection properties of the sensor be summarized by its detection probability P_D and the clutter background by the spatial false return density ρ_F.

Let $j_k = 0$ denote the data interpretation hypothesis that the object has not been detected at all by the sensor at time t_k, i.e. all sensor data have to be considered as false measurements, while $1 \leq j_k \leq m_k$ represents the hypothesis that the object has been detected, $\mathbf{z}_k^{j_k} \in Z_k$ being the corresponding measurement of the object properties, the remaining sensor data being false. Evidently, $\{0, \dots, m_k\}$ denotes a set of mutually exclusive and exhaustive data interpretations. Standard reasoning yields for this simple example a likelihood function for ambiguous data given by a weighted sum of Gaußians and a constant (see [13], e.g.):

$$p(Z_k, m_k|\mathbf{x}_k) = \sum_{j_k=0}^{m_k} p(Z_k, m_k|j_k, \mathbf{x}_k)\, p(j_k) \quad (33)$$

$$\propto (1 - P_D)\rho_F + P_D \sum_{j_k=0}^{m_k} \mathcal{N}\big(\mathbf{z}_{j_k};\, \mathbf{H}_k\mathbf{x}_k,\, \mathbf{R}_k\big). \quad (34)$$

Data interpretation hypotheses are the basis for *Multiple Hypothesis Tracking* techniques (MHT, see [12], e.g.). In such situations, the origin of a time series $Z^k = \{Z_k, m_k, Z^{k-1}\}$ of sensor data accumulated up to the time t_k can be interpreted by interpretation histories,

$$\mathbf{j}_k = (j_k, \ldots, j_1), \quad 0 \le j_k \le m_k, \quad (35)$$

that assume at each data collection time t_l, $1 \le l \le k$, a certain data interpretation j_l to be true. Via marginalizing, the previous filtering density $p(\mathbf{x}_{k-1}|Z^{k-1})$ can be written as a mixture over such interpretation histories $\mathbf{j}_{k-1}$: $p(\mathbf{x}_{k-1}|Z^{k-1}) = \sum_{\mathbf{j}_{k-1}} p_{\mathbf{j}_{k-1}}\, \mathcal{N}\big(\mathbf{x}_{k-1};\, \mathbf{x}^{\mathbf{j}_{k-1}}_{k-1|k-1},\, \mathbf{P}^{\mathbf{j}_{k-1}}_{k-1|k-1}\big)$. By making use of likelihood functions for uncertain data, we directly obtain: $p(\mathbf{x}_k|Z^k) = \sum_{\mathbf{j}_k} p_{\mathbf{j}_k}\, \mathcal{N}\big(\mathbf{x}_k;\, \mathbf{x}^{\mathbf{j}_k}_{k|k},\, \mathbf{P}^{\mathbf{j}_k}_{k|k}\big)$. Mixture reduction techniques, such as described in [14], keep the number of mixture components involved manageable.

According to these preliminary considerations and using the data interpretation histories $\mathbf{j}_k$, the accumulated state density is given by:

$$p(\mathbf{x}_{k:n}|Z^k) = \sum_{\mathbf{j}_k} p(\mathbf{j}_k|Z^k)\, p(\mathbf{x}_{k:n}|\mathbf{j}_k, Z^k) \quad (36)$$

$$= \sum_{\mathbf{j}_k} p^{\mathbf{j}_k}_{k|k}\, \mathcal{N}\big(\mathbf{x}_{k:n};\, \mathbf{x}^{\mathbf{j}_k}_{k:n},\, \mathbf{P}^{\mathbf{j}_k}_{k:n}\big), \quad (37)$$

i.e. the ASD for MHT applications is simply a weighted sum of individual ASDs, which are completely defined by the results of prediction, filtering, and retrodiction along a certain branch of the hypothesis tree defined by a particular interpretation history $\mathbf{j}_k$. The corresponding weighting factor is given by the probability of $\mathbf{j}_k$ being true at time t_k given the data: $p(\mathbf{j}_k|Z^k) = p^{\mathbf{j}_k}_{k|k}$. The ASD for MHT application is thus a by product of MHT tracking and retrodiction. Since PDA filtering [2] can be considered as a second-order approximation to MHT, an approximate representation of the corresponding ASD by a single Gaußian can be derived.

5.2 ASDs for IMM Filtering

In practical applications, it may be uncertain, which evolution model out of a set of r possible alternatives is currently in effect (different flight phases such as no turn, slight maneuver, high-g, turn, e.g.). The maneuvering class $1 \le i_k \le r$, an object belongs to at time t_k can thus be considered as a part of its state. Markovian IMM evolution models (see [15] and the literature cited therein) for object states $X_k = (\mathbf{x}_k, i_k)$ have the form:

$$p(x_k, i_k|x_{k-1}, i_{k-1}) = p_{i_k i_{k-1}} \times\ \mathcal{N}\big(\mathbf{x}_k;\, \mathbf{F}^{i_k}_{k|k-1}\mathbf{x}_{k-1},\, \mathbf{D}^{i_k}_{k|k-1}\big). \quad (38)$$

IMM models are thus characterized r by kinematic linear Gaußian transition densities $p(\mathbf{x}_k|\mathbf{x}_{k-1}, i_k)$ and class transition probabilities $p_{i_k i_{k-1}} = p(i_k|i_{k-1})$ that are to be specified and part of the modeling assumptions.

By making use of the Total Probability Theorem, the IMM approach can easily be combined with Kalman or MHT filtering. In the probability density

$$p(\mathbf{x}_k|Z^k) = \sum_{\mathbf{j}_k} p(\mathbf{x}_k, \mathbf{j}_k|Z^k) \quad (39)$$

at each step k of the filtering loop (Eq. 2), the individual terms of the sum become mixture densities themselves,

$$p(\mathbf{x}_k, \mathbf{j}_k|Z^k) = \sum_{i_k,\ldots i_1} p(i_k, \ldots i_1, \mathbf{j}_k|Z^k) \times\ p(\mathbf{x}_k|i_k, \ldots i_1, \mathbf{j}_k, Z^k). \quad (40)$$

Hence, in the optimal approach to IMM filtering the conditional densities $p(\mathbf{x}_k, \mathbf{j}_k|Z^k)$ of the kinematic state $\mathbf{x}_k$ of the target are sums over every possible sequence of dynamics models $i_k, \ldots i_1$ from the initial observation through the most recent measurement at scan k (“Dynamics Histories”). As the number of terms in the sum (40) exponentially increases with increasing k, various techniques have been developed that approximately represent the densities (40) by mixtures with a *constant* or fluctuating but small number of components.

Let us denote the dynamics histories “m scans back” by $\mathbf{i}_k$, an m-tuple of indices,

$$\mathbf{i}_k = (i_k, i_{k-1}, \ldots i_{k-m-1}). \quad (41)$$

In particular, we are looking for approximations by Gaußian mixtures,

$$p(x_k, \mathbf{j}_k|Z^k) \approx \sum_{\mathbf{i}_k} p^{\mathbf{i}_k,\mathbf{j}_k}_{k|k}\, \mathcal{N}\big(\mathbf{x}_k;\, \mathbf{x}^{\mathbf{i}_k,\mathbf{j}_k}_{k|k},\, \mathbf{P}^{\mathbf{i}_k,\mathbf{j}_k}_{k|k}\big). \quad (42)$$

The weighting factors of the data association history $\mathbf{j}_k$, $p^{\mathbf{j}_k}_{k|k} = p(\mathbf{j}_k|Z^k)$, are given by $p^{\mathbf{j}_k}_{k|k} = \sum_{\mathbf{i}_k} p^{\mathbf{i}_k,\mathbf{j}_k}_{k|k}$. Due to Bayes' Theorem, the expectation vectors $\mathbf{x}^{\mathbf{i}_k,\mathbf{j}_k}_{k|k}$ and covariance matrices $\mathbf{P}^{\mathbf{i}_k,\mathbf{j}_k}_{k|k}$ of the mixtures are iteratively obtained by formulae that are essentially based on Kalman filtering. Also the weighting factors $p^{\mathbf{i}_k,\mathbf{j}_k}_{k|k}$ obey simple update formulae. For $n = 1$ and assuming well-separated targets, the density $p(\mathbf{x}_k, \mathbf{j}_k|Z^k)$ is approximated by a mixture with r components according to the r dynamics models used. GPB and IMM algorithms are possible realizations of this scheme [15]. The IMM approach may easily be adopted to fixed-interval retrodiction. In direct analogy to Eq. 42, the densities $p(\mathbf{x}_l|Z^k) = \sum_{\mathbf{j}_k} p(\mathbf{x}_l, \mathbf{j}_k|Z^k)$ in the retrodiction loop (Eq. 3) are approximately represented by the same class of functions previously used in the filtering loop:

$$p(\mathbf{x}_l, \mathbf{j}_k|Z^k) \approx \sum_{\mathbf{i}_i} p^{\mathbf{i}_l\mathbf{j}_k}_{l|k}\, \mathcal{N}\big(\mathbf{x}_l;\, \mathbf{x}^{\mathbf{i}_l\mathbf{j}_k}_{l|k},\, \mathbf{P}^{\mathbf{i}_l\mathbf{j}_k}_{l|k}\big) \quad (43)$$

with $l < k$. The backward iteration is initialized by the filtering result at the present scan k: $p_{k|k}^{\mathbf{i}_k\mathbf{j}_k}$, $\mathbf{x}_{k|k}^{\mathbf{i}_k\mathbf{j}_k}$, $\mathbf{P}_{k|k}^{\mathbf{i}_k\mathbf{j}_k}$. According to [17], approximate update formulae for the parameters defining $p(\mathbf{x}_l, \mathbf{j}_k|Z^k)$ can be derived. According to these considerations and for the same reasons as before, the accumulated state density is as mixture of individual ASDs for each data interpretation and model history:

$$p(\mathbf{x}_{k:n}|Z^k) = \sum_{\mathbf{i}_k\mathbf{j}_k} p_{k|k}^{\mathbf{i}_k\mathbf{j}_k}\ \mathcal{N}\big(\mathbf{x}_{k:n};\, \mathbf{x}_{k:n}^{\mathbf{j}_k\mathbf{j}_k},\, \mathbf{P}_{k:n}^{\mathbf{i}_k\mathbf{j}_k}\big). \tag{44}$$

Each ASD component is defined by the results of prediction, filtering, and retrodiction given these histories. The corresponding weighting factor results from the filtering step. In the case of standard IMM filtering with r evolution models, the ASD is approximately given via marginalization:

$$p(\mathbf{x}_{k:n}, i_k, \dots, i_n) \approx \prod_{l=n} p_{l|k}^{i_l} \\ \times\ \mathcal{N}\big(\mathbf{x}_{k:n};\, \mathbf{x}_{k:n}^{i_k,\dots,i_n},\, \mathbf{P}_{k:n}^{i_k,\dots,i_n}\big), \tag{45}$$

where $\mathbf{x}_{k:n}^{i_k,\dots,i_n}$ is given by the expectation vectors resulting from standard IMM filtering and retrodiction,

$$\mathbf{x}_{k:n}^{i_k,\dots,i_n} = \big(\mathbf{x}_{k|k}^{i_k\top}, \dots, \mathbf{x}_{n|k}^{i_k\top}\big)^\top, \tag{46}$$

while $\mathbf{P}_{k:n}^{i_k,\dots,i_n}$ essentially has the same structure as the Kalman ASD excepting the covariances of IMM filtering and retrodiction are used. The weighing factors $p_{l|k}^{i_l}$ are the same as in standard IMM filtering and retrodiction.

6 Out-of-Sequence Measurements

In any real-world application of sensor data fusion, we have to be aware of *out-of-sequence measurements*. Due to latencies in the underlying communication infrastructure, for example, such measurements arrive at a processing node in a distributed data fusion system "too late", i.e. after sensor data with a time stamp newer than the time stamp of an out-of-sequence measurement have already been processed. Accumulated object state densities are useful means for dealing with this type of sensor data, which may provide valuable information on an object state of interest in spite of their latency, especially if the sensor involved is of a high quality.

Under conditions, where Kalman filtering is applicable, let us consider a measurement $\mathbf{z}_m$ produced at time t_m with $n \leq m < k$, i.e. before the 'present' time t_k, where the time series Z^k is available and has been exploited. We wish to understand the impact this new, but late sensor information has on the present and the past object states $\mathbf{x}_l$, $n \leq l \leq k$, i.e. on the accumulated object state $\mathbf{x}_{k:n}$. Let $\mathbf{z}_m$ be a measurement of the object state $\mathbf{x}_m$ at time t_m characterized by a Gaußian likelihood function, which is defined by a measurement matrix $\mathbf{H}_m$ and a corresponding measurement error covariance matrix $\mathbf{R}_m$ of a Gaußian likelihood function. We furthermore introduce a projection matrix $\mathbf{\Pi}_m$, defined by $\mathbf{\Pi}_m \mathbf{x}_{k:n} = \mathbf{x}_m$, which extracts the object state $\mathbf{x}_m$ from the accumulated state vector $\mathbf{x}_{k:n}$. The likelihood function of the out-of-sequence measurement with respect to the accumulated object state is thus given by:

$$p(\mathbf{z}_m|\mathbf{x}_{k:n}) = \mathcal{N}\big(\mathbf{z}_m;\, \mathbf{H}_m \mathbf{\Pi}_m \mathbf{x}_{k:n},\, \mathbf{R}_m\big). \tag{47}$$

Standard Bayesian reasoning and the product formula for Gaussians (Eq. 55) directly yields for the accumulated state density:

$$p(\mathbf{x}_{k:n}|\mathbf{z}_m, Z^k) = \frac{p(\mathbf{z}_m|\mathbf{x}_{k:n})\, p(\mathbf{x}_{k:n}|Z^k)}{\int \mathrm{d}\mathbf{x}_{k:n}\, p(\mathbf{z}_m|\mathbf{x}_{k:n})\, p(\mathbf{x}_{k:n}|Z^k)} \tag{48}$$

$$= \mathcal{N}\big(\mathbf{x}_{k:n};\, \mathbf{x}_{k:m:n},\, \mathbf{P}_{k:m:n}\big) \tag{49}$$

with parameters obtained by a version of the Kalman update equations:

$$\mathbf{x}_{k:m:n} = \mathbf{x}_{k:n} + \mathbf{W}_{k:m:n}(\mathbf{z}_m - \mathbf{H}_m \mathbf{\Pi}_m \mathbf{x}_{k:n}) \tag{50}$$

$$\mathbf{P}_{k:m:n} = \mathbf{P}_{k:n} - \mathbf{W}_{k:M:n} \mathbf{S}_{k,m,1} \mathbf{W}_{k:m:n}^\top, \tag{51}$$

where the corresponding Kalman gain and innovation matrices are given by:

$$\mathbf{S}_{k:m:n} = \mathbf{H}_m \mathbf{\Pi}_m \mathbf{P}_{k:n} \mathbf{\Pi}_m^\top \mathbf{H}_m^\top + \mathbf{R}_m \tag{52}$$

$$\mathbf{W}_{k:m:n} = \mathbf{P}_{k:n} \mathbf{\Pi}_m^\top \mathbf{H}_m^\top \mathbf{S}_{k:m:n}^{-1}. \tag{53}$$

Note that the matrix $\mathbf{S}_{k:m:n}$ to be inverted when calculating the Kalman gain matrix has the same dimension as the measurement vector $\mathbf{z}_m$, i.e. is low-dimensional matrix, just as in standard Kalman filtering. Nevertheless, the processing of an out-of-sequence measurement $\mathbf{z}_m$ has impact on all state estimates and the related error covariance matrices in the time window considered. The strongest impact is observed for the time t_m, where the measurement has actually been produced, while it declines the further we proceed to the present time $t_k \geq t_l > t_m$ or deeper into the past $t_m > t_l \geq t_n$.

Accumulated state densities are therefore well suited to quantitatively discuss the question to what extent an OoS measurement is still useful or not, a phenomenon that is sometimes called "information aging". If we are interested in the updated state estimates for the time t_l, $n \leq l \leq k$, we simply have to consider the density:

$$p(\mathbf{x}_l|\mathbf{z}_m, Z^k) = \mathcal{N}\big(\mathbf{x}_l;\, \mathbf{\Pi}_l \mathbf{x}_{k,m,1},\, \mathbf{\Pi}_l \mathbf{P}_{k,m,1} \mathbf{\Pi}_l^\top\big) \tag{54}$$

which results from applying the projection matrix $\mathbf{\Pi}_l$ previously introduced. In a practical application, we will usually be interested in the effect of out-of-sequence measurements, which were produced not too long ago, on the present time and the most recent past. It is therefore sufficient to consider accumulated state densities $p(\mathbf{x}_{k:n}|Z^k)$ characterized by lower dimensional parameters $\mathbf{x}_{k:n}$, $\mathbf{P}_{k:n}$. To determine the actual size of n to be taken into account is an important task in designing sensor networks.

By using the accumulated state densities for MHT and IMM filtering, these considerations can directly be generalized to the case of ambiguous sensor data. Their practical implementation, however, needs a more extended analysis of several numerical and algorithmic issues than can be provided here.

Appendix

Product Formula for Gaußians For matrices of suitable dimensions the following formula for products of Gaußians holds:

$$\begin{aligned}\mathcal{N}\big(\mathbf{z};\,\mathbf{H}\mathbf{x},\,\mathbf{R}\big)\,\mathcal{N}\big(\mathbf{x};\,\mathbf{y},\,\mathbf{P}\big) &= \mathcal{N}\big(\mathbf{z};\,\mathbf{H}\mathbf{y},\,\mathbf{S}\big)\\ &\times \begin{cases}\mathcal{N}\big(\mathbf{x};\,\mathbf{y}+\mathbf{W}\boldsymbol{\nu},\,\mathbf{P}-\mathbf{W}\mathbf{S}\mathbf{W}^\top\big)\\ \mathcal{N}\big(\mathbf{x};\,\mathbf{Q}(\mathbf{P}^{-1}\mathbf{y}+\mathbf{H}^\top\mathbf{R}^{-1}\mathbf{z}),\,\mathbf{Q}\big)\end{cases}\end{aligned} \tag{55}$$

with the following abbreviations:

$$\boldsymbol{\nu} = \mathbf{z} - \mathbf{H}\mathbf{y} \tag{56}$$

$$\mathbf{S} = \mathbf{H}\mathbf{P}\mathbf{H}^\top + \mathbf{R} \tag{57}$$

$$\mathbf{W} = \mathbf{P}\mathbf{H}^\top\mathbf{S}^{-1} \tag{58}$$

$$\mathbf{Q}^{-1} = \mathbf{P}^{-1} + \mathbf{H}^\top\mathbf{R}^{-1}\mathbf{H}. \tag{59}$$

The vector $\mathbf{x}$ in both factors of the left side of Equation 55 only exists in one of the factors on the right side.

Sketch of a proof: Interpret $\mathcal{N}\big(\mathbf{z};\,\mathbf{H}\mathbf{x},\,\mathbf{R}\big)\,\mathcal{N}\big(\mathbf{x};\,\mathbf{y},\,\mathbf{P}\big)$ as a joint density:

$$p(\mathbf{z},\mathbf{x}) = p(\mathbf{z}|\mathbf{x})p(\mathbf{x}). \tag{60}$$

It can be written as a Gaußian, from which the marginal and conditional densities $p(\mathbf{z})$, $p(\mathbf{x}|\mathbf{z})$ can be derived. In the calculations we make use of known formulae for the inverse of a partitioned matrix (see [2, p. 22], e.g.). From

$$p(\mathbf{z},\mathbf{x}) = p(\mathbf{x}|\mathbf{z})p(\mathbf{z}) \tag{61}$$

the formula results.

Inverse of Block Matrices. The inversion of the inverse ASD covariance matrix makes use of well-known results on block matrices [18]. Die Inverse of a partitioned symmetric matrix is given by:

$$\begin{aligned}&\begin{pmatrix}\mathbf{A} & \mathbf{C}\\ \mathbf{C}^\top & \mathbf{B}\end{pmatrix}^{-1} =\\ &= \begin{pmatrix}\mathbf{A}^{-1}+\mathbf{A}^{-1}\mathbf{C}\mathbf{S}^{-1}\mathbf{C}^\top\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{C}\mathbf{S}^{-1}\\ -\mathbf{S}^{-1}\mathbf{C}^\top\mathbf{A}^{-1} & \mathbf{S}^{-1}\end{pmatrix}\end{aligned} \tag{62}$$

where the matrix $\mathbf{S} = \mathbf{B} - \mathbf{C}^\top\mathbf{A}^{-1}\mathbf{C}$ is called the *Schur Complement* of the matrix $\mathbf{A}$.

Gaußian ASDs: Details. While the expectation vector $\mathbf{x}_{k:n}^k$ of the accumulated object states $\mathbf{x}_{k:n}$ is defined by:

$$\mathbf{x}_{k:n}^k = (\mathbf{x}_{k|k}^\top, \mathbf{x}_{k-1|k}^\top, \ldots, \mathbf{x}_{n+1|k}^\top, \mathbf{x}_{n|k}^\top)^\top,$$

the corresponding covariance matrix $\mathbf{P}_{k:n}^k$ can recursively be written as:

$$\mathbf{P}_{l:n}^k = \begin{pmatrix}\mathbf{P}_{l|k} & \mathbf{P}_{l|k}\mathbf{W}_{l-1:n}^\top\\ \mathbf{W}_{l-1:n}\mathbf{P}_{l|k} & \mathbf{P}_{l-1:n}^k\end{pmatrix},$$

$n+1 \le l \le k$, with $\mathbf{P}_{n:n}^k = \mathbf{P}_{n|k}$ and $\mathbf{W}_{l:n}$ given by:

$$\mathbf{W}_{l:n} = \begin{pmatrix}\mathbf{W}_{l|l+1}\\ \mathbf{W}_{l-1:n}\mathbf{W}_{l|l+1}\end{pmatrix}$$

with $\mathbf{W}_{n:n} = \mathbf{W}_{n|n+1}$ and the retrodiction gain matrices $\mathbf{W}_{l|l+1}$ given by:

$$\begin{aligned}\mathbf{W}_{l|l+1} &= \mathbf{P}_{l|l}\mathbf{F}_{l+1|l}^\top\mathbf{P}_{l+1|l}^{-1}\\ &= \mathbf{P}_{l|l}\mathbf{F}_{l+1|l}^\top(\mathbf{F}_{l+1|l}\mathbf{P}_{l|l}\mathbf{F}_{l+1|l}^\top + \mathbf{D}_{l+1|l})^{-1}.\end{aligned}$$

This statement directly follows from a straightforward induction argument, though the necessary calculations are perhaps somewhat tedious. The proposition holds for $k = n$. Let us assume that it is true at time t_k. Due to standard assumptions, the ASD at time t_{k+1} can be represented by:

$$\begin{aligned}&p(\mathbf{x}_{k+1:n}|Z^{k+1}) =\\ &\frac{p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})\,p(\mathbf{x}_{k+1}|\mathbf{x}_k)\,p(\mathbf{x}_{k:n}|Z^k)}{\int \mathrm{d}\mathbf{x}_{k+1:n}\,p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})\,p(\mathbf{x}_{k+1}|\mathbf{x}_k)\,p(\mathbf{x}_{k:n}|Z^k)}.\end{aligned} \tag{63}$$

Using the projection matrices $\boldsymbol{\Pi}_k = (\mathbf{1}, \mathbf{O}, \ldots, \mathbf{O})$ defined by:

$$\boldsymbol{\Pi}_k\mathbf{x}_{k:n} = \mathbf{x}_k \tag{64}$$

and $\boldsymbol{\Pi}_{k:n} = (-\mathbf{W}_{k:n}, \mathbf{1})$ defined by:

$$\boldsymbol{\Pi}_{k:n}(\mathbf{x}_{k+1}^\top, \mathbf{x}_{k:n}^\top)^\top = -\mathbf{W}_{k:n}\mathbf{x}_{k+1} + \mathbf{x}_{k:n}, \tag{65}$$

a repeated use of the product formula 55 yields:

$$\begin{aligned}&p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})\,p(\mathbf{x}_{k+1}|\mathbf{x}_k)\,p(\mathbf{x}_{k:n}|Z^k)\\ &\quad= \mathcal{N}\big(\mathbf{z}_k;\,\mathbf{H}_k\mathbf{x}_{k+1},\,\mathbf{R}_{k+1}\big)\;\mathcal{N}\big(\mathbf{x}_{k+1};\,\mathbf{x}_{k+1|k},\,\mathbf{P}_{k+1|k}\big)\\ &\qquad\times\;\mathcal{N}\big(\mathbf{x}_{k:n};\,\mathbf{x}_{k:n}^k + \mathbf{W}_{k:n}(\mathbf{x}_{k+1} - \mathbf{x}_{k+1|k}),\,\mathbf{R}_{k:n}\big)\\ &\quad= \mathcal{N}\big(\mathbf{z}_k;\,\mathbf{H}_k\mathbf{x}_{k+1|k},\,\mathbf{S}_{k+1|k}\big)\\ &\qquad\times\;\mathcal{N}\big(\mathbf{x}_{k+1};\,\mathbf{x}_{k+1|k+1},\,\mathbf{P}_{k+1|k+1}\big)\\ &\qquad\times\;\mathcal{N}\big(\boldsymbol{\Pi}_{k:n}\mathbf{x}_{k+1:n};\,\boldsymbol{\Pi}_{k:n}(\mathbf{x}_{k+1|k}^\top, \mathbf{x}_{k:n}^k)^\top,\,\mathbf{R}_{k:n}\big)\end{aligned}$$

with $\mathbf{W}_{k:n}$ and $\mathbf{R}_{k:n}$ given by:

$$\begin{aligned}\mathbf{W}_{k:n} &= \mathbf{P}_{k:n}^k\boldsymbol{\Pi}_k^\top\mathbf{F}_{k+1|k}^\top\mathbf{P}_{k+1|k}^{-1}\\ &= \begin{pmatrix}\mathbf{W}_{k|k+1}\\ \mathbf{W}_{k-1:n}\mathbf{W}_{k|k+1}\end{pmatrix}\\ \mathbf{R}_{k:n} &= \mathbf{P}_{k:n}^k - \mathbf{W}_{k:n}\mathbf{P}_{k+1|k}\mathbf{W}_{k:n}^\top.\end{aligned}$$

We thus obtain:

$$\boldsymbol{\Pi}_{k:n}^\top\mathbf{R}_{k:n}^{-1}\boldsymbol{\Pi}_{k:n} = \begin{pmatrix}\mathbf{W}_{k:n}^\top\mathbf{R}_{k:n}^{-1}\mathbf{W}_{k:n} & -\mathbf{W}_{k:n}^\top\mathbf{R}_{k:n}^{-1}\\ -\mathbf{R}_{k:n}^{-1}\mathbf{W}_{k:n} & \mathbf{R}_{k:n}^{-1}\end{pmatrix}.$$

By a second use of the product formula for Gaußians, we thus obtain up to a constant independent of the state vectors:

$$\begin{aligned}&p(\mathbf{z}_{k+1}|\mathbf{x}_{k+1})\,p(\mathbf{x}_{k+1}|\mathbf{x}_k)\,p(\mathbf{x}_{k:n}|Z^k) \propto\\ &\mathcal{N}(\mathbf{z}_k;\,\mathbf{H}_k\mathbf{x}_{k+1|k},\,\mathbf{S}_{k+1|k})\,\mathcal{N}(\mathbf{x}_{k+1:n};\,\mathbf{x}_{k+1:}^{k+1},\,\mathbf{P}_{k+1:n}^{k+1}),\end{aligned}$$

where the covariance matrix $\mathbf{P}_{k+1:n}^{k+1}$ it is given by:

$$\mathbf{P}_{k+1:n}^{k+1} = \left(\mathbf{\Pi}_{k+1}^\top \mathbf{P}_{k+1|k+1}^{-1} \mathbf{\Pi}_{k+1} + \mathbf{\Pi}_{k:n}^\top \mathbf{R}_{k:n}^{-1} \mathbf{\Pi}_{k:n}\right)^{-1}$$
$$= \begin{pmatrix} \mathbf{P}_{k+1|k+1}^{-1} + \mathbf{W}_{k:n}^\top \mathbf{R}_{k:n}^{-1} \mathbf{W}_{k:n} & -\mathbf{W}_{k:n}^\top \mathbf{R}_{k:n}^{-1} \\ -\mathbf{R}_{k:n}^{-1} \mathbf{W}_{k:n} & \mathbf{R}_{k:n}^{-1} \end{pmatrix}^{-1}.$$

This block matrix can directly be inverted by using Eq. 62. The corresponding Schur Complement is particularly simple and given by:

$$\mathbf{T} = \mathbf{P}_{k+1|k+1}^{-1} + \mathbf{W}_{k:n}^\top \mathbf{R}_{k:n}^{-1} \mathbf{W}_{k:n} - \mathbf{W}_{k:n}^\top \mathbf{R}_{k:n}^{-1} \mathbf{R}_{k:n} \mathbf{R}_{k:n}^{-1} \mathbf{W}_{k:n} = \mathbf{P}_{k+1|k+1}^{-1}.$$

We thus obtain for the covariance matrix of the ASD:

$$\mathbf{P}_{k+1:n}^{k+1} = \begin{pmatrix} \mathbf{P}_{k+1|k+1} & \mathbf{P}_{k+1|k+1}\mathbf{W}_{k:n}^\top \\ \mathbf{W}_{k:n}\mathbf{P}_{k+1|k+1} & \mathbf{R}_{k:n} + \mathbf{W}_{k:n}\mathbf{P}_{k+1|k+1}\mathbf{W}_{k:n}^\top \end{pmatrix}$$
$$= \begin{pmatrix} \mathbf{P}_{k+1|k+1} & \mathbf{P}_{k+1|k+1}\mathbf{W}_{k:n}^\top \\ \mathbf{W}_{k:n}\mathbf{P}_{k+1|k+1} & \mathbf{P}_{k:n}^{k} + \mathbf{W}_{k:n}(\mathbf{P}_{k+1|k+1} - \mathbf{P}_{k+1|k})\mathbf{W}_{k:n}^\top \end{pmatrix}. \quad (66)$$

Using the identity

$$\mathbf{W}_{k|k+1}(\mathbf{P}_{k+1|k+1} - \mathbf{P}_{k+1|k})\mathbf{W}_{k|k+1}^\top = \mathbf{P}_{k|k+1} - \mathbf{P}_{k|k} \quad (67)$$

resulting from the Rauch-Tung-Striebel equations, the matrix $\mathbf{W}_{k:n}(\mathbf{P}_{k+1|k+1} - \mathbf{P}_{k+1|k})\mathbf{W}_{k:n}^\top$ can be transformed yielding:

$$\mathbf{W}_{k:n}(\mathbf{P}_{k+1|k+1} - \mathbf{P}_{k+1|k})\mathbf{W}_{k:n}^\top = \begin{pmatrix} \mathbf{P}_{k|k+1} - \mathbf{P}_{k|k} & (\mathbf{P}_{k|k+1} - \mathbf{P}_{k|k})\mathbf{W}_{k-1:n}^\top \\ \mathbf{W}_{k-1:n}(\mathbf{P}_{k|k+1} - \mathbf{P}_{k|k}) & \mathbf{W}_{k-1:n}(\mathbf{P}_{k|k+1} - \mathbf{P}_{k|k})\mathbf{W}_{k-1:n}^\top \end{pmatrix}.$$

With this result, the block matrix $\mathbf{P}_{k:n}^{k} + \mathbf{W}_{k:n}(\mathbf{P}_{k+1|k+1} - \mathbf{P}_{k+1|k})\mathbf{W}_{k:n}^\top$ on the right-lower corner on the right side of Eq. 66 is given by:

$$\mathbf{P}_{k:n}^{k} + \mathbf{W}_{k:n}(\mathbf{P}_{k+1|k+1} - \mathbf{P}_{k+1|k})\mathbf{W}_{k:n}^\top = \begin{pmatrix} \mathbf{P}_{k|k+1} & \mathbf{P}_{k|k+1}\mathbf{W}_{k-1:n}^\top \\ \mathbf{W}_{k-1:n}\mathbf{P}_{k|k+1} & \mathbf{P}_{k-1:n}^{k} + \mathbf{W}_{k-1:n}(\mathbf{P}_{k|k+1} - \mathbf{P}_{k|k})\mathbf{W}_{k-1:n}^\top \end{pmatrix}.$$

An induction argument for the block matrix on the right-lower corner directly yields:

$$\mathbf{P}_{k:n}^{k} + \mathbf{W}_{k:n}(\mathbf{P}_{k+1|k+1} - \mathbf{P}_{k+1|k})\mathbf{W}_{k:n}^\top = \mathbf{P}_{k:n}^{k+1}. \quad (68)$$

According to the product formula Eq. 55, $\mathbf{x}_{k+1:n}^{k+1}$ is the sum of the following vectors:

$$\mathbf{P}_{k+1:n}^{k+1}\mathbf{\Pi}_{k:n}^\top \mathbf{R}_{k:n}^{-1} \mathbf{\Pi}_{k:n}(\mathbf{x}_{k+1|k}^\top, \mathbf{x}_{k:n}^{k\top})^\top \quad (69)$$
$$= \begin{pmatrix} \mathbf{O} \\ -\mathbf{W}_{k:n}\mathbf{x}_{k+1|k} + \mathbf{x}_{k:n}^{k} \end{pmatrix} \quad (70)$$
$$\mathbf{P}_{k+1:n}^{k+1}\mathbf{\Pi}_{k+1}^\top \mathbf{P}_{k+1|k+1}^{-1} \mathbf{\Pi}_{k+1}\mathbf{x}_{k+1|k+1} \quad (71)$$
$$= \begin{pmatrix} \mathbf{x}_{k+1|k+1} \\ \mathbf{W}_{k:n}\mathbf{x}_{k+1|k+1} \end{pmatrix}. \quad (72)$$

By using an induction argument, we thus obtain:

$$\mathbf{x}_{k+1:n}^{k+1} = \begin{pmatrix} \mathbf{x}_{k+1|k+1} \\ \mathbf{x}_{k|k+1} \\ \mathbf{x}_{k-1:n}^{k} + \mathbf{W}_{k-1:n}(\mathbf{x}_{k|k+1} - \mathbf{x}_{k|k}) \end{pmatrix}. \quad (73)$$

An induction argument concludes the proof.

References

[1] S. Blackman and R. Populi. *Design and Analysis of Modern Tracking Systems.* Artech House, 1999.

[2] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan. *Estimation with Applications to Tracking and Navigation.* Wiley & Sons, 2001.

[3] M.S. Arulampalam, S. Maskell, N. Gordon, T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE T-SP*, Vol. 50, Nr. 2, pp. 174-188 (2002).

[4] Y. Bar Shalom. Update with Out-of-Sequence Measurements in Tracking: Exact Solution. *IEEE T-AES*, Vol. 38, No. 3, p. 769-777, 2002.

[5] M. Mallick. Out-of-sequence track filtering using the decorrelated pseudo measurement approach. *SPIE SMT*, Vol. 5428, 154 (2004).

[6] Y. Bar-Shalom, H. Chen. IMM Estimator with Out-of-Sequence Measurements. *IEEE T-AES*, Vol. 41, No. 1, p. 90-98, 2005.

[7] M. Orton, A. Marrs. Particle Filters for tracking with out-of-sequence measurements. *IEEE T-AES*, Vol. 41, No. 2, p. 693-702, 2005.

[8] R.G. Everitt, S.R. Maskell, R. Wright, M. Briers. Multi-target out-of-sequence data association: Tracking using graphical models. *Information Fusion*, Vol. 7, No. 4, p. 434-447, 2007.

[9] J. Zhen, A. Balasuriya, S. Challa. Sensor fusion-based visual target tracking for autonomous vehicles with the out-of-sequence measurements solution. *Robotics and Autonomous Systems*, Vol. 56, No. 2, p. 157-176, 2008.

[10] O. E. Drummond. Target Tracking with Retrodicted Discrete Probabilities. *SPIE SMT*, Vol. 3163, 249 (1997).

[11] A. Gelb (Ed.). *Applied Optimal Estimation*, MIT Press (1974).

[12] S. Blackman. Multiple Hypothesis Methods. *IEEE AESS Magazine*, **19**, 1, p. 41–52, 2004.

[13] W. Koch. Target Tracking. Chapter 8 in *S. Stergiopoulos (Ed.), Signal Processing Handbook.* CRC Press, 2001.

[14] D. Salmond. "Mixture Reduction Algorithms for Target Tracking in Clutter". In *SPIE SMT*, vol. 1305, pp. 434-445, (1990).

[15] Rong X. Li and V.P. Jilkov. "Survey of maneuvering target tracking. Part V. Multiple model methods". In *IEEE T-AES*, Vol. 41, Nr. 4, pp. 1255- 1321 (2005).

[16] D.L. Hall and J. Llinas (Eds.). *Handbook of Multisensor Data Fusion.* CRC Press, 2001.

[17] W. Koch. Fixed-Interval Retrodiction Approach to Bayesian IMM-MHT for Maneuvering Multiple Targets. *IEEE T-AES*, Vol. 36, No. 1, p. 2-14, 2002.

[18] Harville, D. A., *Matrix Algebra from a Statistician's Perspective*. Springer, 1997.