Joint Identification of Multiple Tracked Targets

DOMINIC E. SCHAUB

This paper derives a rigorously Bayesian technique for estimating the identities of a plurality of targets that are well separated or tracked using the (joint) probabilistic data association filter. In contrast to the single-target classification problem, the joint identification of multiple targets is characterized by statistical dependencies between track-to-identity assignments that render track-level estimation of identity suboptimal. The present method rigorously accounts for these dependencies and allows arbitrary feature and kinematic measurements generated by individual targets to be used in finding the statistically-optimal track-to-identity assignment probabilities. The problem is decomposed into global combinatorial identity deconfliction and local target tracking and classification that is based on a unified measure-theoretic filtering framework. The computational complexity of this technique is shown to be dominated by calculation of the permanent of a non-negative matrix, which may be found exactly in exponential time or approximated in polynomial time using Markov chain Monte Carlo methods. Strategies for improving numerical performance are given for cases where certain subsets of targets are indistinguishable or unobservable. This work is relevant to applications in tactical settings, surveillance, including video tracking, air/land/maritime situational awareness, and automated intelligence collection.

Manuscript received February 25, 2015; revised February 7, 2016; released for publication June 21, 2016.

Refereeing of this contribution was handled by Dr. Paolo Braca.

Author's address: Defence Research and Development Canada, Atlantic Research Centre, 9 Grove St., P.O. Box 1012, Dartmouth NS B2Y 3Z7 Canada (E-mail: dominic.schaub@drdc-rddc.gc.ca).

This work was supported by DRDC Applied Research Project 06eo, *Situational Information for Enabling Development of Northern Awareness* (SEDNA).

1557-6418/17/$17.00 © 2017 JAIF

1. INTRODUCTION

Multisensor tracking is a branch of information fusion that appears in many domains that require situational awareness, such as air traffic control, video surveillance, and missile defence. The fundamental task in tracking involves estimating target kinematic states (such as position and/or velocity of an aircraft) from a time series of measurements generated by a suite of one or more sensors (e.g. radar). While several approaches have been developed for fusing measurement data, those based on (or approximating) Bayesian filtering [1]–[3] have enjoyed the widest adoption owing to their rigorous treatment of sensor error and target dynamics. In the Bayesian framework, a probability distribution over target state space is maintained and updated with new measurements using Bayes' theorem. In particular, where sensor noise and target dynamics admit a priori statistical characterizations, Bayesian methods offer a rigorously mathematical framework for computing optimal statistical estimates.

Over the course of several decades, enormous advances in sensors, communications systems, and computing power have enabled the development of a considerable number of tracking methods, from the simple, linear Kalman filter [4], [5] (which solves the single-target state-estimation problem under Gaussian conditions) to the sophisticated multi-hypothesis tracker [6] (which performs the task of associating multiple observations to multiple tracks and is used in conjunction with a collection of individual filters) and the probability hypothesis density (PHD) filter [7], [8] (which tightly integrates data association and filtering). Research has also broadened to include the related problems of target identification and classification, which naturally extend the mathematics of target tracking. In principle, these undertakings may be regarded as specializations of the same fundamental problem, as identification amounts to a constrained form of classification that assigns a given class to at most a single target.

As with tracking, identification and classification are marked by significant differences between their single- and multiple-target specializations. Single-target joint tracking and classification (JTC) has been extensively studied as a rigorously Bayesian problem, with detailed theoretical derivations provided in [9]–[11]. Particular attention has been given to exploiting kinematic data to assist classification, including [12], which demonstrated improvements in classification performance by applying a second-order uncertainty model to the mapping between the feature and target class spaces. Similarly, [13] developed a framework of multiple-model particle and mixture Kalman filters subject to kinematic constraints and subsequently considered its application to discriminating between commercial and military aircraft using radar contacts. In [14], it was shown that maximum entropy techniques can significantly improve accuracy in Bayesian classification characterized by epistemic uncertainty in the prior (which, when unknown, is typically taken as uniform). Specific applications have also been considered, including a decision-theoretical problem of identifying aircraft using radar measurements [15] and a joint tracking and classification framework for radar and electronic-support-measure observations [16]. It is worth highlighting that Bayesian single-target joint tracking and classification is typically computationally feasible for modest-dimensional problems, where the 'curse of dimensionality' (in the non-parametric case) is usually mild and may be overcome through particle filtering or carefully implemented fixed-grid discretizations.

In the multitarget context, identification has been theoretically analyzed as an extension to Finite Set Statistics (FISST) using the framework of labelled random sets [17]. Several approximation schemes have also been developed (albeit lacking in rigorous tracking formalism), including methods based on information theory [18], [19] and Sinkhorn rescaling [20]–[22]. Accounts of relating specific tracking implementations with the higher-level identification problem have been given in [23], which developed (using a series of approximations) a multitarget Kalman filter that utilizes target identity information. In particular, a group-theoretical Fourier method [24]–[26] has emerged as a general framework for approximate reasoning over combinatorial matchings. In this approach, distributions defined over the set of permutations are replaced with equivalent (Fourier-transformed) distributions over the irreducible representations of the related symmetric group. Under favourable conditions, the transformed quantities may be approximated with a small number of terms, thereby avoiding the factorial space and time complexities ordinarily encountered in combinatorial problems.

The foregoing methods of multitarget identification are either intractable (implementation of the random-set formulation entails further numerical simplifications) or dependent on complex approximations that impede the analysis of error in computed estimates. These limitations are a consequence not only of the inherent complexity of multitarget tracking, but of the fact that multitarget identification and constrained classification—where at most a fixed number of targets may belong to a given class[1]—, is itself non-local and combinatorial, as estimating the class of a target depends partly on observations made at distant tracks (e.g. a target at a given track is unlikely to be a particular identity when there is strong evidence for its presence elsewhere). Interestingly, the underlying mathematical structure—known as the assignment problem [27], [28]—is of considerable generality, appearing in several diverse contexts such as economics [29], [30], operations research [31], [32], and the joint probabilistic data association (JPDA) filter [33].

The present work develops a special case of rigorously Bayesian multitarget identification and classification wherein it becomes computationally feasible to calculate nearly exact estimates of identity and kinematic states. The necessary conditions are met whenever measurement-to-track associations are unambiguous, i.e., where tracks are well separated or, alternatively, where (J)PDA is used to resolve the data association problem. The primary objectives motivating this work lie with computing the marginal track-to-identity assignments in the form of posterior probabilities that a given identity—such as a vessel with a particular registration number—is present at a track of interest (Fig. 1) and, in addition, finding the optimal posterior target-state-space densities from which kinematic estimates may be calculated. It is shown that the algorithmic bottleneck stems from calculation of the matrix permanent, a standard function in combinatorics and one that may be efficiently computed using rapidly-mixing Markov chain Monte Carlo (MCMC) methods [34]–[36]. As these techniques yield approximations whose residual errors decrease exponentially with iteration number, optimal Bayesian estimates of any quantity may therefore be found to machine precision in polynomial time. It is further shown that the Ryser method [37], [38] (an analytical permanent algorithm) can be extended to exploit the mathematical structure present when targets form classes of indistinguishable members, a result that enables the *exact* constrained classification of very large numbers of targets to be performed efficiently.

The remainder of this paper is organized as follows. Section 2-A derives a general Bayesian framework that unifies the problems of tracking, classification, and identification under a single statistical system that fully accounts for their complex mathematical interdependence and makes optimal use of arbitrary feature and kinematic observations originating from individual tracks. In §2-B, the framework is equipped with simplifying conditions, including the requirement for unambiguous measurement-to-track associations, which in turn give way to an efficient factorization of the joint target density. This factorization naturally partitions the framework into the connected problems of local tracking and classification and global combinatorial deconfliction of identity, thereby significantly improving the associated time and space complexities. Section 2-C proceeds to show that the central task of deconfliction amounts to computing the permanent of a non-negative matrix, and in §2-D, the framework is extended to applications where data association is performed with (J)PDA. Section 3 then discusses algorithms for computing the matrix permanent, including a Ryser method that is modified to exploit the presence of unobservable

[1]For example, a given task group may include a known number of UAVs and helicopters. In this case, the fixed number of each aircraft imposes a global constraint on the multitarget classification problem.

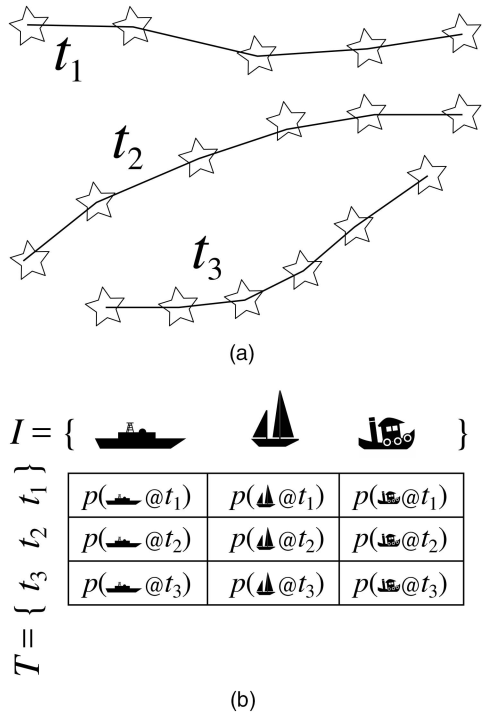

Fig. 1. Simplified overview of the problem studied in this paper. Given a set T of tracks with associated observations (a) and a second set I of unique identities, it is desired to find the matrix of posterior marginal track-to-identity assignment probabilities (b).

or indistinguishable targets to improve runtime (§3-A and §3-B) and various approximation methods, including Markov chain Monte Carlo techniques and loopy belief propagation (§3-C). Finally, a series of numerical examples are considered in §4, demonstrating how this work may be applied to benefit situational awareness.

2. BAYESIAN FORMULATION

A. Overview

This section derives a measure-theoretic Bayesian framework for multitarget filtering and classification in cases where target dynamics and measurements satisfy standard Markov conditions. Development proceeds by augmenting the conventional analytical representation of Bayesian multitarget tracking [8] with a singular-measure extension that allows seamless integration of classification based on static attributes. In what follows, an 'identity' refers to a physical entity of uncertain location that possesses a known signature, which can be represented by a combination of dynamic and static features.[2] Consistent with conventional tracking nomenclature, 'target' denotes the entity present at a given track, which is localized to a particular region in space but is of uncertain identity. The sets of all tracks and identities in existence are designated T and I, respectively (Fig. 1). Furthermore, these sets are well ordered, allowing each of their members to be referenced by a unique integer index.

The present framework is simplified when each identity is assignable to a distinct track. To accommodate circumstances where observed tracks are produced by a series of false alarms (or where there are fewer observed tracks than identities), it may be mathematically convenient to enlarge T by including additional tracks whose targets are never observed. Without loss of generality, the total set of tracks may thus be defined by the union

$$T = T_{\mathrm{O}} \cup T_{\mathrm{H}} \tag{1}$$

where T_{O} are those tracks with observed targets (to which one or more measurements are associated), and T_{H} is a (possibly empty) set of hidden tracks lacking measurements. The set T_{H} can be made identical in size to I to allow for the possibility that every track in T_{O} was generated by false alarms.

Each identity $i \in I$ assumes values in its associated state space ($\mathbb{S}_i$). For mathematical simplicity, an arbitrary identity state $\mathbf{x}_i \in \mathbb{S}_i$ is made to encapsulate both dynamic attributes (e.g. velocity) and static properties (e.g. length) in a manner that allows the state space to be decomposed as the Cartesian product

$$\mathbb{S}_i = \mathbb{S}_{i,\mathrm{d}} \times \mathbb{S}_{i,\mathrm{s}} \tag{2}$$

where $\mathbb{S}_{i,\mathrm{d}}$ and $\mathbb{S}_{i,\mathrm{s}}$ are the dynamic and static components, respectively. Both $\mathbb{S}_{i,\mathrm{d}}$ and $\mathbb{S}_{i,\mathrm{s}}$ must be equipped with σ-algebras, allowing $\mathbb{S}_i$ to be assigned the measure

$$\mu_i(W) = \lambda_i(\mathrm{pr}_{\mathrm{d}} W) \times \delta_i(\mathrm{pr}_{\mathrm{s}} W) \tag{3}$$

where $W \subseteq \mathbb{S}_i$ is an arbitrary measurable set, $\mathrm{pr}_{(\cdot)}$ are projection operators, and $\lambda_i(\cdot)$ and $\delta_i(\cdot)$ are measures on $\mathbb{S}_{i,\mathrm{d}}$ and $\mathbb{S}_{i,\mathrm{s}}$, respectively. The latter is the (singular) Dirac measure that encodes the given identity's static attribute information. For example, where there exist two identities of length 1 m and 2 m, their respective Dirac measures on their 'length' spaces would be $\delta_{1\,\mathrm{m}}(\cdot)$ and $\delta_{2\,\mathrm{m}}(\cdot)$. For an observation of target length given by[3] $\mathcal{N}(\mu = 1\ \mathrm{m}, \sigma^2 = 1\ \mathrm{m}^2)$, the likelihoods associated with each of the identities may be found by applying their Dirac measures to the measurement's pdf, and are thus $(2\pi)^{-1/2}$ and $e^{-1}(2\pi)^{-1/2}$ for the 1-m and 2-m identities, respectively. However, rather than extracting feature information directly from individual observations, it may be convenient to maintain a running product of static-feature measurements from previous updates—these products are simply individual measurements pdfs (pmfs) that are combined by successive applications of

[2]For example, each registered vessel on a lake would constitute a unique identity. Note that for a given suite of sensors, identities may not necessarily be distinguishable by way of their observed signatures, as occurs when identical models of watercraft are tracked with radar. In practical applications, the set of identities may be realized as a comprehensive database of physical assets (and their signatures) that could be encountered by the tracking system.

[3]Obviously, the normal distribution implies that lengths could assume negative (non-physical) values.

Bayes' rule.[4] Invocation of the identity-specific Dirac measures may then be deferred to the time at which the posterior probabilities are computed. By transferring the classification operations to the measure, static features are placed on equal footing as the dynamic components, allowing Bayesian feature measurement updates to be performed in manner identical to their kinematic counterparts. Note that this formulation simplifies the subsequent presentation but is theoretically equivalent to that derived with intermediate feature spaces.

Multitarget tracking and classification may be formulated with Cartesian products of individual identity spaces that are constructed as

$$\mathbb{S}_J = \prod_{i\in J} \mathbb{S}_i, \qquad J \subseteq I, \tag{4}$$

where the product order is given by the total order on I, and the subscripts are suppressed when $J = I$. The $\mathbb{S}_J$ of (4) are similarly assigned the family of product measures

$$\mu_J(W) = \prod_{i\in J} \mu_i(\mathrm{pr}_i W). \tag{5}$$

Uncertainty in the system's dynamic state may be represented with a random variable X on $\mathbb{S}$ whose probability density function $p_X(\mathbf{x})$ is uniquely defined up to an equivalence class determined by the measures of (3). In particular, each equivalence class contains a member that assumes strictly constant values on each of the static attribute spaces and thus satisfies

$$p_{X_i}(\mathbf{x}_i) = p_{X_i}(\mathbf{x}_i'), \qquad \forall \mathbf{x}_i, \mathbf{x}_i' \in \mathbb{S} : \mathrm{pr}_{\mathrm{d}}\{\mathbf{x}_i\} = \mathrm{pr}_{\mathrm{d}}\{\mathbf{x}_i'\} \tag{6}$$

which uses the fact that all functions on the ith static attribute space that evaluate to the same quantity for the argument selected by $\delta_i(\cdot)$ are equivalent under the associated measure (e.g. $f_1(x) = 1/\sqrt{2}$, $f_2(x) = |\sin(x)|$, and $f_3(x) = |\cos(x)|$ are equivalent under $\delta_{\pi/4}(\cdot)$). Priors on static attribute spaces may therefore be specified as constant functions. Marginalization of this density with respect to $J \subset I$ may be defined as

$$p_{X_{I\setminus J}}(\mathbf{x}_{I\setminus J}) = \int_{\mathbb{S}_J} p_X(\mathbf{x}) d\mu_J(\mathbf{x}_J). \tag{7}$$

The system state is updated with a sequence of discrete-time measurements $\mathfrak{z}^k$, which are samples of the time series of random variables $Z^k \in \mathbb{S}_Z$ that describe the measurement processes. The collection of measurements up to and including the kth time step is given by

$$\mathfrak{z}^{1:k} = \{\mathfrak{z}^{k'}\}_{k'=1}^{k}. \tag{8}$$

For each time step there exists a likelihood function $f : (\mathbb{S}_Z, \mathbb{S}) \to [0,\infty)$ that carries information on the measured static and/or dynamic attributes. When conditioned on the multitarget state and evaluated for a given measurement $Z^k = \mathfrak{z}^k$, this function may be written as

$$f_{Z^k|X^k}(\mathfrak{z}^k \mid \mathbf{x}). \tag{9}$$

Between measurements, the dynamic components of the state are projected forward in time by a Markov transition density $f : (\mathbb{S}, \mathbb{S}) \to [0,\infty)$ that encodes the known aspects of the identities' dynamics (e.g. acceleration, etc.). It is similarly conditioned on the kth state and is denoted by

$$f_{X^{k+1}|X^k}(\mathbf{x} \mid \mathbf{x}'). \tag{10}$$

Bayesian filtering consists of alternately updating the state with new measurements using Bayes' rule (usually increasing the Fisher information) and transitioning the state forward in time (usually decreasing the information). The measurement and transition steps may be written as

$$p_{X^k|Z^{1:k}}(\mathbf{x} \mid \mathfrak{z}^{1:k}) = \frac{f_{Z^k|X^k}(\mathfrak{z}^k \mid \mathbf{x}) p_{X^k|Z^{1:k-1}}(\mathbf{x} \mid \mathfrak{z}^{1:k-1})}{\int_{\mathbb{S}} f_{Z^k|X^k}(\mathfrak{z}^k \mid \mathbf{x}') p_{X^k|Z^{1:k-1}}(\mathbf{x}' \mid \mathfrak{z}^{1:k-1}) d\mu(\mathbf{x}')} \tag{BF.1}$$

and

$$p_{X^{k+1}|Z^{1:k}}(\mathbf{x} \mid \mathfrak{z}^{1:k}) = \int_{\mathbb{S}} f_{X^{k+1}|X^k}(\mathbf{x} \mid \mathbf{x}') p_{X^k|Z^{1:k}}(\mathbf{x}' \mid \mathfrak{z}^{1:k}) d\mu(\mathbf{x}') \tag{BF.2}$$

respectively. By virtue of the product measures of (3), these equations simultaneously subsume both tracking and classification (two specializations that occur in the absence of $\mathbb{S}_{\mathrm{s},i}$ and $\mathbb{S}_{\mathrm{d},i}$, respectively).

B. Tracking and Identification of Well-Separated Targets

The general framework of the previous section is challenging to implement numerically. For example, a fixed-grid discretization of a probability density function exhibits time and space (storage) complexities[5] that scale exponentially with the dimension of $\mathbb{S}$ (which is proportional to cardinality of I), precluding even approximate evaluation in all but the most elementary cases. The present work exploits the fact that in many instances, targets are far apart and non interacting. In

[4] Philosophically, this approach treats fixed features (such as length) as time-indexed probability distributions (akin to the distributions describing kinematics) subject to standard Bayesian recursive filtering. By choosing the prior feature density to be the Dirac delta function (or Kronecker delta function where the prior is a pmf) and defining the associated Markov transitions to have no effect, the feature posterior will remain the Dirac delta function indefinitely, as no measurement function can increase its information further. However, the delta function's magnitude will be scaled in accordance with how 'close' the feature measurements are to the feature encoded in the delta function. The Dirac measure is substituted for the Dirac delta function to avoid the mathematical sophistication required to use the latter.

[5] An overview of computational complexity theory may be found in [39].

this case, measurement-to-track associations are unambiguous, and the problem inherits several simplifying characteristics ordinarily associated with single-target tracking. Those simplifications form the basis the remaining analysis and are described below.

1) Individual identity state spaces are derived from a single common space. That is,

$$\mathbb{S}_i = \mathcal{S} \qquad \forall i \in I. \tag{11}$$

In practice, a sufficiently large $\mathcal{S}$ can be defined to encompass any set of original identity state spaces.

2) To enforce a clear separation of targets, the probability density functions maintained at each track[6] are confined to a unique set $r_t^k \subset \mathcal{S}$ indexed by track and time. These target regions need not be explicitly defined, provided that a disjoint construction ($r_t^k \cap r_{t'}^k = \emptyset$ for all $t \neq t'$) is guaranteed to exist—an assumption that is reasonable when targets are well separated. For example, where the tth target is tracked in two dimensions using Kalman filtering, the positional component of r_t^k may be defined as the area circumscribed by a few standard deviations from the means of the normally-distributed posteriors.[7] In what follows, the collection of all disjoint regions at the kth time step is given as

$$R^k = \{r_t^k\}_{t\in T}, \tag{12}$$

from which it may be seen that each r_t^k corresponds to a distinct t, and $|R^k| = |T|$. As each identity may be found in one of the regions, the following holds

$$\Pr\left\{ X_i^k \in \bigcup_{t\in T} r_t^k \right\} = 1, \qquad \forall i \in I, \tag{13}$$

where X_i^k (shorthand for $X_{\{i\}}^k$) is the ith-identity marginalization of X^k. In the present work, each track (region) is restricted to contain at most one target, yielding

$$0 \le \sum_{\substack{J\subseteq I\\ |J|\ \text{fixed}}} \Pr\{X_J^k \in (r_t^k)^{|J|} \wedge X_{I\setminus J}^k \in \complement(r_t^k)^{|I\setminus J|}\}$$
$$\le \begin{Bmatrix} 1 & |J| = 1 \\ 0 & |J| > 1 \end{Bmatrix} \qquad \forall r_t^k \in R^k, \tag{14}$$

where $\complement$, $|\cdot|$, and $(r_t^k)^{|\cdot|}$ denote set complementation, cardinality, and Cartesian products of r_t^k, respectively. As an immediate consequence of (13) and (14), each i-indexed component of any random sample of X^k must always fall upon a distinct r_t^k. Thus, as foreshadowed in §2-A, the number of disjoint regions (and hence the number of tracks) must match or exceed the number of identities. When the rightmost inequality in (14) is strict, the affected region may not actually contain a target, and the probability of existence of the corresponding track is therefore less than one.

The restrictions imposed by (13) and (14) may be expressed in terms of a time-indexed union of admissible subsets A_σ^k of the multitarget state space $\mathbb{S}$. Each subset corresponds to a single injective function σ between I and T and is defined as the Cartesian product

$$A_\sigma^k = \prod_{i\in I} r_{\sigma(i)}^k. \tag{15}$$

The collection of σ forms the set of all identity-track permutations $P(I,T)$. Thus, the admissible subset of $A^k \subset \mathbb{S}$ is given by the union

$$A^k = \bigcup_{\sigma\in P(I,T)} A_\sigma^k. \tag{16}$$

In turn, the A_σ^k may be used to construct the indicator functions $\mathbf{1}_{A_\sigma^k}(\mathbf{x})$ and $\mathbf{1}_{A^k}(\mathbf{x})$ defined on $\mathbb{S}$. Moreover, every $\mathbf{1}_{A_\sigma^k}(\mathbf{x})$ can be factored into identity-specific components as

$$\mathbf{1}_{A_\sigma^k}(\mathbf{x}) = \prod_{i\in I} \mathbf{1}_{r_{\sigma(i)}^k}(\mathbf{x}_i). \tag{17}$$

By virtue of the fact that $A_{\sigma_1}^k \cap A_{\sigma_2}^k = \emptyset$ for $\sigma_1 \neq \sigma_2$, the indicator functions also satisfy

$$\mathbf{1}_{A^k}(\mathbf{x}) = \sum_{\sigma\in P(I,T)} \mathbf{1}_{A_\sigma^k}(\mathbf{x}). \tag{18}$$

3) The joint density function admits a quasi statistically independent factorization as the product of a scaled indicator function and a set of normalized i-indexed probability density functions $p_i^k(\mathbf{x}_i)$ on $\mathcal{S}$ as

$$p_{X^k}(\mathbf{x}) = C^k \mathbf{1}_{A^k}(\mathbf{x}) \prod_{i\in I} p_i^k(\mathbf{x}_i) \tag{19}$$

where $\mathbf{1}_{A^k}(\mathbf{x})$ zeroes regions of the joint probability density that violate (14), and C^k is a positive constant that renormalizes the product, which is partially zeroed by the indicator function. Note that $p_i^k(\mathbf{x}_i)$ is distinct from $p_{X_i^k}(\mathbf{x}_i)$, which is the $I\setminus\{i\}$-marginalization found with (7). The $p_i(\mathbf{x}_i)$ may themselves be decomposed as sums of region (track)-specific normalized density functions to yield

$$p_{X^k}(\mathbf{x}) = C^k D^k \mathbf{1}_{A^k}(\mathbf{x}) \prod_{i\in I} \sum_{t\in T} (b_t^k)^{-1} a_{t,i}^k p_{t,i}^k(\mathbf{x}_i)$$
$$\operatorname{supp}(p_{t,i}^k) \subseteq r_t^k \tag{20}$$

where $\operatorname{supp}(\cdot)$ denotes function support, and the $a_{t,i}^k$, b_t^k, and D^k are non-negative coefficients that are consistent with C^k and $p_i^k(\mathbf{x}_i)$ in (19). To ensure the existence of $(b_t^k)^{-1}$, the b_t^k are further required to be positive. The C^k, D^k, and b_t^k are extraneous (they may be subsumed in the $a_{t,i}^k$) but simplify subsequent

[6]These are explicitly defined in condition 3. For the moment, they may be regarded as probability density functions produced by single-target trackers.

[7]In actual fact, the normal distribution has unbounded support, and the separation condition therefore fails to hold. This is may be overcome by treating the Kalman filter as an approximation that operates on truncated Gaussian functions.

analysis. Also note the relation

$$\mathbf{1}_{r_t^k}(\mathbf{x}_i) p_{t',i}^k(\mathbf{x}_i) = \begin{cases} p_{t',i}^k(\mathbf{x}_i), & t = t' \\ 0, & t \neq t'. \end{cases} \tag{21}$$

The factorizations of (19) and (20) allow the full joint distribution to be efficiently represented by series of constants that grows only linearly with the number identities and tracks, thereby immensely reducing the problem's computational complexity.

4) Target motion is assumed independent, allowing the Markov transition densities to be factored as

$$\begin{aligned} f_{X^{k+1}|X^k}(\mathbf{x} \mid \mathbf{x}') &= \prod_{i \in I} f_{X_i^{k+1}|X_i^k}(\mathbf{x}_i \mid \mathbf{x}_i') \\ &= \prod_{i \in I} \sum_{t \in T} f_{t,X_i^{k+1}|X_i^k}(\mathbf{x}_i \mid \mathbf{x}_i') \end{aligned} \tag{22}$$

where $\mathrm{supp}(f_{t,X_i^{k+1}|X_i^k})$ is a subset of the Cartesian product $r_t^{k+1} \times r_t^k$.

5) Observations are assumed to be of type produced in single-target Bayesian filtering. Each successive time step therefore corresponds to only a single observation at one track, where the history of observation-to-track associations is given by the time-indexed observation vector $\mathbf{o}(k)$ that maps each time step k to a single track t. To update the multitarget prior, a global measurement likelihood function must be constructed from a single-target, single-track observation function $f_{Z^k|X_S^k}(\mathfrak{z}^k \mid \mathbf{x} \in \mathcal{S})$ as follows

$$f_{Z^k|X^k}(\mathfrak{z}^k \mid \mathbf{x}) = \mathbf{1}_{A^k}(\mathbf{x}) \left(\sum_{i \in I} p_{\mathrm{d},i}^k f_{Z^k|X_S^k}(\mathfrak{z}^k \mid \mathbf{x}_i) + p_{\mathrm{f}}^k \right),$$
$$\mathrm{supp}(f_{Z^k|X_S^k}) \subseteq r_{\mathbf{o}(k)}, \tag{23}$$

where $t = \mathbf{o}(k)$ is the track that produced the observation, $p_{\mathrm{f}}^k \geq 0$ is the sensor false-alarm probability, $p_{\mathrm{d},i}^k \geq 0$ is the identity detection probability, and the leading indicator term is included for mathematical convenience (it is idempotent, and owing to its appearance in the prior, has no effect when used in the Bayesian update). Expressed as a sum, the measurement function does not obviously preserve the quasi statistical independence in (19). However, when $p_{\mathrm{f}}^k > 0$, the properties of the indicator function may be used to rewrite (23) as

$$\begin{aligned} &f_{Z^k|X^k}(\mathfrak{z}^k \mid \mathbf{x}) = \\ &\quad (p_{\mathrm{f}}^k)^{1-|I|} \mathbf{1}_{A^k}(\mathbf{x}) \prod_{i \in I} (p_{\mathrm{d},i}^k f_{Z^k|X_S^k}(\mathfrak{z}^k \mid \mathbf{x}_i) + p_{\mathrm{f}}^k), \end{aligned} \tag{24}$$

which is easily shown to leave the structure of (19) intact. While this reformulation introduces an $(|I|-1)$th-order singularity with respect to p_{f}^k, it is removable in the regions of $\mathbb{S}$ that are not zeroed by the indicator function, as it cancels with a powers of p_{f}^k greater than or equal to $|I|-1$. Thus, while (24) is undefined for $p_{\mathrm{f}}^k = 0$, the right-hand sides of (23) and (24) share the same pointwise limit $p_{\mathrm{f}}^k \to 0^+$, a property that may be used in cases where false-alarm probabilities are zero (a formal limiting procedure is derived in Appendix B). Finally, (24) may be algebraically recast as

$$\begin{aligned} f_{Z^k|X^k}(\mathfrak{z}^k \mid \mathbf{x}) &= p_{\mathrm{f}}^k \mathbf{1}_{A^k}(\mathbf{x}) \\ &\quad \cdot \prod_{i \in I} (\mathbf{1}_{r_t^k}(\mathbf{x}_i)(p_{\mathrm{f}}^k)^{-1} g_{i,\mathfrak{z}^k}(\mathbf{x}_i) + \mathbf{1}_{\complement r_t^k}(\mathbf{x}_i)), \end{aligned} \tag{25}$$

where

$$g_{i,\mathfrak{z}^k}(\mathbf{x}_i) = p_{\mathrm{d},i}^k f_{Z^k|X_S^k}(\mathfrak{z}^k \mid \mathbf{x}_i) + p_{\mathrm{f}}^k. \tag{26}$$

By induction, conditions 4–5 may be proved consistent with condition 3 for all time instants, provided that (19) holds at $k = 0$. Of particular significance is the fact that, while the X_i^k are not statistically independent, their joint probability density function $p_{X^k}(\mathbf{x})$ nonetheless admits an efficient factorization that is readily updated with new information from local tracks. The Bayesian filtering problem therefore amounts to iteratively computing $a_{t,i}^k$, b_t^k, D^k, and $p_{t,i}^k(\mathbf{x}_i)$ from previous values. In accordance with the derivation in Appendix A, the kth time-step quantities may be related to those at $k-1$ by

$$\begin{aligned} p_{t,i}^k(\mathbf{x}_i) &= (K_i^k)^{-1} g_{i,\mathfrak{z}^k}(\mathbf{x}_i) \\ &\quad \cdot \int_{r_t^k} f_{X_i^k|X_i^{k-1}}(\mathbf{x}_i \mid \mathbf{x}_i') p_{t,i}^{k-1}(\mathbf{x}_i') d\mu(\mathbf{x}_i') \\ a_{t,i}^k &= K_i^k a_{t,i}^{k-1} \\ b_t^k &= p_{\mathrm{f}}^k b_t^{k-1} \end{aligned} \tag{27}$$

for $t = \mathbf{o}(k)$, and

$$\begin{aligned} p_{t,i}^k(\mathbf{x}_i) &= \int_{r_t^k} f_{X_i^k|X_i^{k-1}}(\mathbf{x}_i \mid \mathbf{x}_i') p_{t,i}^{k-1}(\mathbf{x}_i') d\mu(\mathbf{x}_i') \\ a_{t,i}^k &= a_{t,i}^{k-1} \\ b_t^k &= b_t^{k-1} \end{aligned} \tag{28}$$

for $t \neq \mathbf{o}(k)$, with K_i and D^k given by

$$\begin{aligned} K_i^k &= \int_{r_t^k \times r_t^{k-1}} g_{i,\mathfrak{z}^k}(\mathbf{x}_i'') f_{X_i^k|X_i^{k-1}}(\mathbf{x}_i'' \mid \mathbf{x}_i') \\ &\quad \cdot p_{t,i}^{k-1}(\mathbf{x}_i') d\mu(\mathbf{x}_i'') d\mu(\mathbf{x}_i') \\ D^k &= p_{\mathrm{f}}^k D^{k-1} = \prod_{t \in T} b_t^k. \end{aligned} \tag{29}$$

Equations 27 and 28 illustrate that updating the coefficients may be carried out on a per-track basis. Thus, the problem may be formulated as a collection of individual single-target tracking problems whose identity information exhibit inter-track statistical dependence. Each track t is associated with a single region r_t^k, and a separate Bayesian filter is run for each observed track/identity pair for a total of $|T_O| \cdot |I|$ individual filters.

For a given track, this entails selecting a prior comprising $a^1_{t,i}$, b^1_t, D^1, and $p^1_{t,i}(\mathbf{x})$ and then recursively computing the filtering equations of (27) and (28) for each $i \in I$. Note that calculation of the normalization constant C^k is not required at each time step and may therefore be deferred to the final evaluation of the posterior.

Where observations occur simultaneously, measurements may be assigned consecutive time indexes by defining the zero-time-difference Markov transitions as identity maps. Alternatively, simultaneous measurements may be folded into a single time step. When concurrent observations originate from the same track, a composite function may be constructed as the product of individual measurement functions. In the general case—where observations originate from multiple tracks—$\mathbf{o}(k)$ becomes multivalued, and $\mathfrak{Z}^k$, $p^k_{\mathrm{d},i}$, p^k_{f}, and K^k_i must be relabeled as $\mathfrak{Z}^k_t$, $p^k_{\mathrm{d},t,i}$, $p^k_{\mathrm{f},t}$, and $K^k_{t,i}$, respectively.

C. Combinatorial Evaluation of Target Identity (Identity Deconfliction)

Information about the track-to-identity assignment probabilities may be summarized by the matrix $\mathbf{P}^k$, whose elements $\mathbf{P}^k_{t,i}$ are the marginal probabilities that identity i is located at track t. The kth time-step $\mathbf{P}^k_{t,i}$ are calculated from the $a^k_{t,i}$ and b^k_t coefficients (maintained by the individual trackers), which may be collected in the $|T| \times |T|$-dimensional matrix[8]

$$\mathbf{M}^k_{t,j} = \left\{ \begin{matrix} a^k_{t,i}, & j \leq |I| \\ b^k_t, & j > |I| \end{matrix} \right\}, \tag{30}$$

where the possible inclusion and repetition of b^k_t ensures that $\mathbf{M}^k$ is square. In accordance with Appendix B-A, each $\mathbf{P}^k_{t,i}$ may then be recovered from ith-identity marginalization of the multitarget joint density function

$$\begin{aligned} \mathbf{P}^k_{t,i} &= \int_{r^k_t} p_{X^k_i | Z^{1:k}}(\mathbf{x}'_i \mid \mathfrak{Z}^{1:k}) d\mu(\mathbf{x}'_i) \\ &= \mathbf{M}^k_{t,i} \frac{\mathrm{Per}(\mathbf{M}^k(t;i))}{\mathrm{Per}(\mathbf{M}^k)}, \end{aligned} \tag{31}$$

where $\mathbf{M}^k(t,i)$ is the matrix formed by deleting the tth row and ith column from $\mathbf{M}^k$ and $\mathrm{Per}(\cdot)$ denotes the matrix permanent [37], [38] (a combinatorial sum indexed by the $|T|$th-order symmetric group $S_{|T|}$) defined as

$$\mathrm{Per}(\mathbf{M}^k) = \sum_{\sigma \in S_{|T|}} \prod_{j=1}^{|T|} \mathbf{M}^k_{\sigma(j),j}. \tag{32}$$

Note that the updates to $a^k_{t,i}$, b^k_t, D^k, and $p^k_{t,i}(\mathbf{x})$ are entirely independent of (31), whose evaluation may be deferred to an arbitrary time step, and furthermore, only those $\mathbf{P}^k_{t,i}$ of interest need be computed. As a result of the limiting procedure described in Appendix B-B, the p^k_{f} associated with zero false-alarm measurements may be set to zero when computing $g_{i,\mathfrak{Z}^k}$ and b^k_t. Consequently, those tracks for which at least one measurement was completely certain (zero false alarm probability) will possess a b^k_t equal to zero.

The integration of identification into tracking also yields improvements to track-level estimates. In particular, the coefficients of $\mathbf{P}^k$ (which contain the marginal track-to-identity assignment probabilities) can be used to scale the individual density functions of (20) in the expansion

$$p^k_t(\mathbf{x}) = \sum_{i \in I} \mathbf{P}^k_{t,i} p^k_{t,i}(\mathbf{x}), \tag{33}$$

where the resulting probability density function $p^k_t(\mathbf{x})$ is the weighted average of the i-indexed posteriors generated by the trackers of the filter bank at track t. Estimation may be performed on $p^k_t(\mathbf{x})$ to compute track-level quantities of dynamic (e.g. kinematic) or static attributes. Interestingly, while the use of multiple models for transition and measurements functions clearly benefits tracking in a range of applications, the global combinatorial deconfliction step (which generates $\mathbf{P}^k$) improves the performance of local tracking—and any associated estimates—still further by incorporating non-local information into the track-level posterior (e.g. if a unique, fast-moving entity i is observed with high certainty at a remote track $t' \neq t$, the weight $\mathbf{P}^k_{t,i}$ in the sum of (33) will be reduced, and $p^k_t(\mathbf{x})$ will be commensurately improved).

Finally, groups of identities that possess indistinguishable feature measures $\delta(\cdot)$ and Markov transition densities (in (3) and (10), respectively) form equivalence classes that give rise the problem of constrained classification.[9] The foregoing analytical framework remains unchanged, though it should be noted that assigning a common prior to members of a group of indistinguishable identities will result in identical marginal track-to-identity probabilities ($\mathbf{P}^k_{t,\cdot}$) across members of that class.

D. Tracking and Identification using (J)PDAF

The preceding framework readily extends to circumstances where tracking is performed using the (joint) probabilistic data association filter [40], [41], albeit at the loss of mathematical optimality.[10] In this case, condition 2 of §2-B no longer holds (the marginalized probability density functions of different targets may overlap), and consequently, the joint state cannot be written as (19) or (20), which rely on indicator functions to zero the inadmissible regions of the multitarget space. This may be remedied by rewriting the joint density

[8]The t and i that index the elements of $\mathbf{M}^k$ are shorthand for the integer indices (over $\{1,\ldots,|T|\}$ and $\{1,\ldots,|I|\}$) induced by the total orders on T and I, respectively.

[9]Identification is the trivial case that occurs when a given class contains only a single member.

[10]Note that while JPDA employs similar combinatorial framework for determining *measurement-to-track associations*, the calculations involved are distinct from the those of the identification problem, which seek to find the *track-to-identity assignments*.

function as an explicit summation over identity-track permutations given by

$$p_{X^k}(\mathbf{x}) = C^k D^k \sum_{\sigma\in P(I,T)} \prod_{i\in I} (b_t^k)^{-1} a^k_{\sigma(i),i} p^k_{\sigma(i),i}(\mathbf{x}_i) \tag{34}$$

With this modification, the track-level processing and global identity deconfliction of the previous sections may be retained without further change. However, while the measurement updates may proceed as before, *determination of the updates themselves* (e.g. by nearest neighbour or maximum likelihood in PDA—or combinatorial methods in JPDA) should be performed using the previous timestep's posterior weights to the scale the track-level density functions (i.e., by using $\mathbf{P}^k_{t,i} p^k_{t,i}(\mathbf{x}_i)$ rather than $(b_t^k)^{-1} a^k_{t,i} p^k_{t,i}(\mathbf{x}_i)$). Thus, an optimally-implemented (J)PDA filter requires that the identity deconfliction step be carried out prior to each new measurement.

3. EVALUATION OF THE MATRIX PERMANENT

A. Exact Methods

Although the direct evaluation of (32) is readily implemented and numerically stable, its time complexity (for an n-dimensional matrix) of $\mathcal{O}(n \cdot n!)$ is prohibitive in most practical applications. An improvement in running time is realized by evaluating the sum by way of an inclusion-exclusion decomposition

$$\mathrm{Per}(\mathbf{M}^k) = (-1)^n \sum_{Y\in\mathcal{P}(\mathbb{N}_n)} (-1)^{|Y|} \prod_{q=1}^{n} \sum_{y\in Y} \mathbf{M}^k_{q,y}, \tag{35}$$

where $\mathcal{P}(\mathbb{N}_n)$ is the power set of the first n positive integers. Known as the Ryser formula [37], [38], equation (35) is the most efficient known exact method for finding the permanent of an arbitrary matrix. Its time complexity is $\mathcal{O}(n^2 2^n)$, which may be improved to $\mathcal{O}(n2^n)$ by evaluating the trailing sum using a Gray-code order [38].

The running time is further reduced when $\mathbf{M}^k$ possesses groups of repeated columns and/or rows. The permanent may then be computed from a smaller $\hat{n} \times \hat{m}$-dimensional matrix $\bar{\mathbf{M}}^k$ that comprises only the unique columns and rows of $\mathbf{M}^k$. As shown in Appendix C, (35) may be recast as

$$\mathrm{Per}(\mathbf{M}^k) = \sum_{\mathbf{d}\in\prod_{j=1}^{\hat{n}} \mathbb{N}_{n_j}} (-1)^{n+\|\mathbf{d}\|_1} \left[\prod_{j=1}^{\hat{n}} \binom{d_j}{n_j}\right] \prod_{q=1}^{\hat{m}} (\mathbf{d}\cdot\bar{\mathbf{M}}^k_q)^{m_q}. \tag{36}$$

where $\|\cdot\|_1$ is the L^1 norm, $\binom{d_j}{n_j}$ are binomial coefficients, $\prod_{j=1}^{\hat{n}} \mathbb{N}_{n_j}$ is a Cartesian product of sets of positive integers, and n_j and m_q are the number of identical members in the jth repeated-column and qth repeated-row groups, respectively. The corresponding time complexity is given by

$$\mathcal{O}\left(\min\left(\hat{m}\prod_{j=1}^{\hat{n}}(n_j+1), \hat{n}\prod_{q=1}^{\hat{m}}(m_q+1)\right)\right), \tag{37}$$

where the minimization operation results from the permanent's invariance under matrix transposition.

Repeated columns will be present in $\mathbf{M}^k$ whenever identities aggregate into equivalence classes and/or there are repeated b_t^k (i.e., $|T| > |I|$). Similarly, $\mathbf{M}^k$ will possess repeated rows in applications where there are two or more hidden tracks. For certain observability/equivalence-class conditions, the optimized Ryser formula of (36) may therefore render computable an otherwise intractable identity-deconfliction step. In particular, (37) demonstrates that the computability of the modified Ryser method is quite favourable when either the number of observed targets or the number of target classes is moderate in size. In this regime, the computation time exhibits little dependence on the absolute number of identities. Finally, note that computation of (36) and (35) may be readily (and efficiently) parallelized in a manner that preserves the Gray-code evaluation sequence.

B. Numerical Considerations Concerning Exact Methods

Although the Ryser formula and its derivatives yield exact results for infinite-precision arithmetic, they entail computing the sum of terms of alternating sign, some of which may be of considerable magnitude. Therefore, implementation of these methods demands careful selection of numerical libraries and their parameters to ensure that the minimum requirements for arithmetic precision (e.g., the number of bits in the mantissas of floating point numbers) are satisfied. In many cases, data types based on native floating point implementations are inadequate, requiring the use of variable-precision libraries such as the GNU Multiple Precision (GMP) Arithmetic Library [42].

The smallest floating-point mantissa that safeguards numerical accuracy follows directly from the maximum possible ratio between intermediate summation terms and the permanent itself. Although the former is easily bounded from above by inspection of (35) and (36), non-trivial lower bounds for arbitrary positive matrices are generally not available. However, in view of the fact that a tight lower bound of e^{-n} exists for the permanent of doubly stochastic[11] matrices [43], it is beneficial analyze the factorization

$$\mathbf{M}^k_{\mathrm{DS}} = \mathbf{D}^k_1 \mathbf{M}^k \mathbf{D}^k_2, \tag{38}$$

where $\mathbf{M}^k_{\mathrm{DS}}$ is doubly stochastic, and $\mathbf{D}_1$ and $\mathbf{D}_2$ are invertible diagonal matrices possessing only non-negative

[11] A doubly stochastic matrix possesses unit row and column sums.

entries. When each element of $\mathbf{M}^k$ is greater[12] than some $\alpha > 0$, the existence of this decomposition is guaranteed by the Sinkhorn theorem[13] [46]. Using elementary properties of the matrix permanent, (38) is easily shown to yield

$$\mathrm{Per}(\mathbf{M}_{\mathrm{DS}}^k) = \prod_{j=1}^{n} (\mathbf{D}_1^k)_{j,j} \prod_{j=1}^{n} (\mathbf{D}_2^k)_{j,j} \mathrm{Per}(\mathbf{M}^k), \tag{39}$$

allowing calculation of $\mathrm{Per}\mathbf{M}^k$ to be replaced with that of $\mathrm{Per}\mathbf{M}_{\mathrm{DS}}^k$. As the row sums of a doubly stochastic matrix are necessarily unity, an upper bound for the intermediate terms in both (35) and (36) may be taken as 2^n. The resulting ratio is thus $(2e)^n$, and the required size of the mantissa is given by

$$N_{\mathrm{M}} = n(1 + \log_2 e) + N_+, \tag{40}$$

where N_+ denotes the number of additional bits determined by the number of significant digits required in the final result and the anticipated accumulation of round-off error.

The variable cost of floating-point operations will affect the computational requirements of the (modified) Ryser algorithm. Provided that the binomial coefficients and their inverses are precomputed, the sum of (36) may be calculated in a generalized Gray-code sequence using only additions, subtractions, and multiplications, the last of which dominates the asymptotic complexity (the same holds unmodified Ryser algorithm). Thus, (37) may be revised by scaling it with the prefactor $N_{\mathrm{M}} \log N_{\mathrm{M}} 2^{\mathcal{O}(\log^* N_{\mathrm{M}})}$, which is the asymptotic complexity of integer multiplication using the Fürer algorithm[14] [47]. As it is unlikely that the original data or the rescaling step will outstrip the limitations of native data types, enhanced precision should only be necessary when computing the permanent itself.

Computing the factorization of (38) may itself add to the total computational complexity of the (modified) Ryser method. For a residual L^∞ error of ϵ (over the row and column sums in $\mathbf{M}_{\mathrm{DS}}^k$), the diagonal matrices may be computed approximately in $\mathcal{O}(n^4 \log(n/\epsilon)\log(1/\alpha))$ time using the RAS algorithm [48] or, alternatively, by iterative Sinkhorn scaling [46] at the cost of a slightly less favourable runtime. When $\epsilon < \alpha$, the permanent will deviate from that of its fully doubly-stochastic counterpart by at most a factor of $|(\alpha - \epsilon)/\alpha|^n$. To ensure that the commensurate change of accuracy requirements does not exceed one bit, the relation $0.5 < |(\alpha \pm \epsilon)/\alpha|^n < 2$ must be satisfied. By solving for ϵ at the extreme points, expanding the solutions using a pair of Taylor series with respect to $1/n$, and bounding from above the sums with those of the associated infinite geometric series, it may be found that $\epsilon < (\alpha \ln 2)/n$ must hold asymptotically. The complexity of the RAS algorithm therefore becomes, $\mathcal{O}(n^4 \log(n/\alpha)\log(1/\alpha))$, which may be improved by processing the matrix that comprises only the unique columns and rows of $\mathbf{M}^k$.

Finally, it is observed that $\mathbf{M}_{\mathrm{DS}}^k$ may itself be used as an approximation to the matrix of marginal track-to-identity assignment probabilities $\mathbf{P}^k$ (§2-C). Like $\mathbf{P}^k$, $\mathbf{M}_{\mathrm{DS}}^k$ possess unit row and column sums and may thus be regarded as a collection of quasi probabilities (though at the cost of an error that is difficult to characterize). This approach was used the basis of approximating the multitarget identification problem in [20]–[22].

C. Methods of Approximation

The challenge of the matrix permanent calculation rests with its membership in the #P-hard complexity class [49], whose problems are at least as computationally demanding as those of NP-hard. While the development of exact, polynomial-time algorithms thus seems unlikely, random-approximation methods for non-negative matrices appear considerably more promising and have drawn significant research interest. In particular, a fully polynomial-time randomized approximation scheme (FPRAS) for 0-1 and non-negative matrices was developed in [34]. This work, which constructed a Markov chain Monte-Carlo algorithm for sampling perfect matchings from the associated bipartite graph, exhibits a complexity[15] of $\mathcal{O}^*(n^{26})$, a result that was subsequently improved to $\mathcal{O}^*(n^{10})$ [35] and then $\mathcal{O}^*(n^7)$ [50], [51]. Several other methods have been developed for matrices with additional properties. In particular, [52] uses the self-reducibility of the permanent to derive an algorithm whose complexity is $\mathcal{O}(n^4 \log^4 n)$ for matrices satisfying certain density requirements.

Loopy belief propagation has served as another basis for approximating the permanent of non-negative matrices [53]–[57]. This approach appears to exhibit encouraging accuracy and computational efficiency in real-world applications, but is presently understood in terms of heuristics that lack the theoretical performance guarantees enjoyed by the Monte-Carlo methods. Nonetheless, belief propagation remains a promising area of research, and significant ongoing attention has been directed towards establishing the theoretical underpinnings necessary for conducting rigorous analysis of accuracy and running time.

Finally, it should be noted that latent structure in the coefficients of $\mathbf{M}_{\mathrm{DS}}^k$ (such as sparsity) may afford further savings in computation time. The manner in which the aforementioned algorithms may be optimized under this condition is expected to depend significantly on the data

[12] In accordance with Cromwell's rule [44], never allowing the Bayesian update step to produce a zero element in $\mathbf{M}^k$ is generally good practice, and consequently, elements of $\mathbf{M}^k$ are assumed to be positive.

[13] Actually, the strict positivity requirement can be relaxed for matrices that satisfy certain indecomposability conditions [45] [46].

[14] This is the fastest known integer multiplication algorithm and may be easily extended to floating point arithmetic.

[15] $\mathcal{O}^*(\cdot)$ denotes a complexity wherein logarithmic factors have been suppressed.

itself (and therefore the specific application). For this reason, investigations using empirical data from various problem domains may be valuable avenues of further research.

4. EXAMPLE PROBLEMS

The mathematical framework developed in the previous sections is illustrated by way of three examples (each a single run) that track and identify (or classify) a group of well-separated simulated identities moving independently in two dimensions.

A. Simulation Setup

1) Preliminaries:

Each identity's actual (time-indexed) state ($\mathfrak{X}_i^k$) comprises position and velocity ($\mathrm{pr}_\mathrm{d}\mathfrak{X}_i^k \in \mathbb{S}_{i,\mathrm{d}} = \mathbb{R}^2 \times \mathbb{R}^2$) and a single, unchanging attribute ($\mathrm{pr}_\mathrm{s}\mathfrak{X}_i^k \in \mathbb{S}_{i,\mathrm{s}} = \mathbb{R}$) as dynamic and static components, respectively. The static attribute may be regarded as a feature that is observed in some of the example problems. The single-identity state space is therefore $\mathcal{S} = \mathbb{S}_{\cdot,\mathrm{d}} \times \mathbb{S}_{\cdot,\mathrm{s}} = \mathbb{R}^4 \times \mathbb{R}$, which is equipped with the product measure $\lambda \times \delta_{A_i}$ defined by the ordinary Lebesgue measure λ and the Dirac measure δ_{A_i}, where $A_i \in \mathbb{R}$ is the value of the identity's static attribute.

To simplify the remaining analysis, the detection and false alarm probabilities are set to 1 and 0, respectively, and a common number of tracks and identities is used. Thus, $|T_\mathrm{O} \cup T_\mathrm{H}| = |I|$, and each observed track will correspond to a single identity (and vice versa). The actual mapping between identities and observed targets is described by the permutation matrix

$$\mathbf{U}_{t,i} = \begin{Bmatrix} 1 & \text{If identity } i \text{ is actually at track } t \\ 0 & \text{Otherwise} \end{Bmatrix}, \quad (41)$$

which, along with the collection of time-indexed identity states $\mathfrak{X}_i^k$, forms the reference 'ground truth' over the course of the simulations. Note that $\mathbf{U}$ remains fixed over the course of each simulation.

2) Simulation of Observations and Target Motion:

Observations of the system, during which the positions and features (but not velocity) of all targets are simultaneously measured, are made at intervals of one second (where the relabeling given at the end of §2-B is used). Targets are assumed to be well separated, and each observed track $t \in T_\mathrm{O}$ is unambiguously associated with a time series of measurements $\mathfrak{Z}_t^k$ (indexed by $k = \{1,2,\ldots\}$) that were chosen to carry noisy position and feature information as

$$\mathfrak{Z}_t^k = \mathbf{H}\mathfrak{X}_i^k + \begin{bmatrix} \mathfrak{M}_\mathrm{p} \\ \mathfrak{M}_\mathrm{f} \end{bmatrix}, \quad (42)$$

where

$$\mathbf{H} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}, \quad (43)$$

i is the actual identity at track t satisfying the 'ground-truth' mapping $\mathbf{U}_{t,i} = 1$, $\mathbf{I}$ is the 2×2 identity matrix, and the random samples $\mathfrak{M}_\mathrm{p}$ and $\mathfrak{M}_\mathrm{f}$ are drawn from the normally distributed bivariate position- and univariate feature-measurement noises M_p and M_f, respectively.

For a given identity i, transitions between an successive states (i.e., $\mathfrak{X}_i^k$ to $\mathfrak{X}_i^{k+1}$) were based on the white-noise acceleration model [58]. For a unit time step, consecutive state vectors are related by

$$\mathfrak{X}_i^{k+1} = \mathbf{F}\mathfrak{X}_i^k + \begin{bmatrix} \mathfrak{P}_i/2 \\ \mathfrak{P}_i \\ 0 \end{bmatrix}, \quad (44)$$

where

$$\mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 1 \end{bmatrix}, \quad (45)$$

and $\mathfrak{P}_i$ is a random sample of the bivariate normally distributed process noise P_i that is characteristic to the ith-identity (P_i is also stationary with respect to k). In general, the process noise (i.e., kinematics) is made to vary among identities such that $P_i \not\sim P_{i'}$ for $i \neq i'$, with the exception that $P_i \sim P_{i'}$ when i and i' are members of a common class of indistinguishable identities.

3) Implementation of Single-Target Filters:

Filtering at each observed track $t \in T_\mathrm{O}$ is carried out in accordance with §2-B using a bank of $|I|$ filters that are perfectly matched to the actual measurement and transition processes given in (42) and (44). The Markov transition and measurement steps for the ith identity in a given track's filter bank are therefore modeled as

$$X_i^{k+1} = \mathbf{F}X_i^k + \begin{bmatrix} P_i/2 \\ P_i \\ \mathbf{0} \end{bmatrix} \quad (46)$$

and

$$Z_i^k = \mathbf{H}X_i^k + \begin{bmatrix} M_\mathrm{p} \\ M_\mathrm{f} \end{bmatrix}, \quad (47)$$

respectively.

Owing to the fact that P_i, M_p, and M_f are normal, the normality of any prior X_i^1 will be preserved in subsequent X_i^k, allowing the tracking to be implemented using linear Kalman filters. The density function defined in (20), which corresponds to the ith identity at the tth track, is thus given by

$$p_{t,i}^k(\mathbf{x}_i) = \frac{1}{\sqrt{(2\pi)^5 \mathrm{Det}(\Sigma_{t,i}^k)}} e^{-\frac{1}{2}(\mathbf{x}_i - \mu_{t,i}^k)^\mathrm{T}(\Sigma_{t,i}^k)^{-1}(\mathbf{x}_i - \mu_{t,i}^k)}, \quad (48)$$

where $\mu_{t,i}^k$ and $\Sigma_{t,i}^k$ are the density function's mean vector and covariance matrix, respectively. A collection of $p_{t,i}^k(\mathbf{x}_i)$ spanning every $i \in I$ is maintained by a separate filter bank for each observed track $t \in T_\mathrm{O}$ (Fig. 2). The

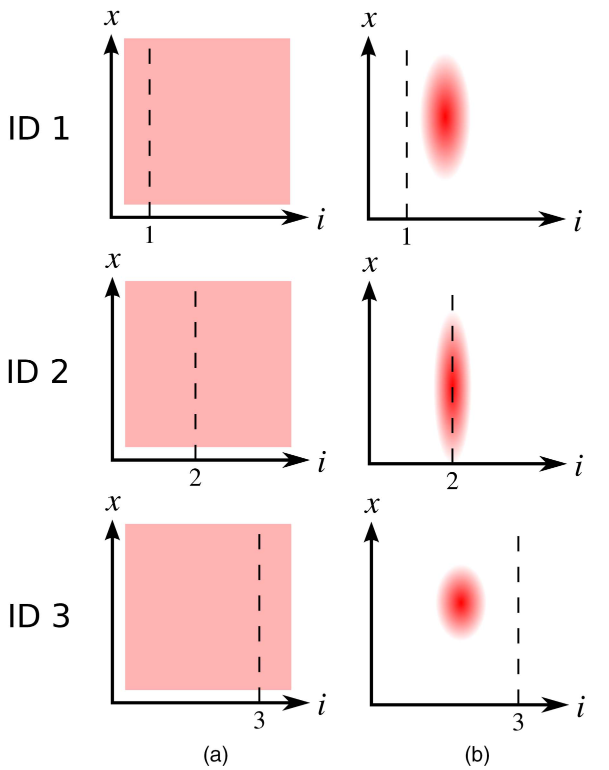

Fig. 2. Illustration of a filter bank's density functions ($p^k_{t,i}(\mathbf{x}_i)$), shown over three identities ($i = 1,2,3$) at some observed track $t \in T_{\mathrm{O}}$ and timesteps (a) $k = 0$ (which is the prior and taken to be uniform) and (b) some $k > 0$. Note that the actual functions span additional kinematic dimensions not shown. The dashed lines represent the Dirac measures of (3) that are unique to each identity space and 'select' the probability associated with a particular identity ($i = 1,2,3$). Therefore, ID 2 represents the most probable identity in the filter bank of (b). As in JTC, there exist variations between the filter bank's individual densities that result from employing different Markov and measurement models with distinct identities (classes).

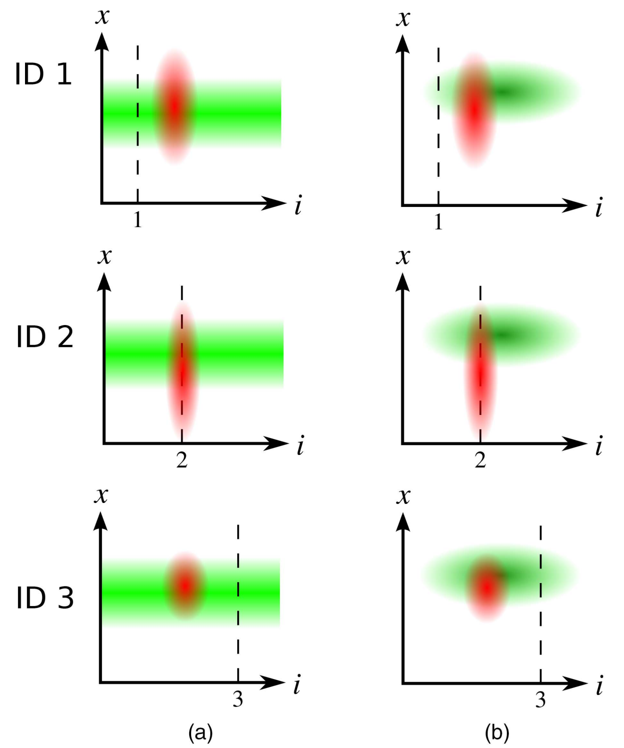

Fig. 3. Example relationship between the density functions ($p^k_{t,i}(\mathbf{x}_i)$, shown in red) of the filter bank at observed track $t \in T_{\mathrm{O}}$ and a given Gaussian measurement likelihood function ($g_{i,\mathfrak{z}^k_t}(\mathbf{x}_i)$, shown in green), where the latter contains (a) only kinematic information and (b) both kinematic and identity information. In this example, those measurement functions also carrying identity information correspond to a noisy feature observation centred near ID 2. The update steps of $p^k_{t,i}(\mathbf{x}_i)$ in (27) and (28) are equivalent to marching the $p^k_{t,i}(\mathbf{x}_i)$ forward in time (in accordance with (49)) and, if the observation was associated to the filter bank's track, multiplying the results with the $g_{i,\mathfrak{z}^{k+1}_t}(\mathbf{x}_i)$. The updated densities are renormalized alongside the computation of the b^{k+1}_t and $a^{k+1}_{t,i}$ coefficients. Once again, the dashed lines represent the identity-specific Dirac measures, and the actual functions span additional kinematic dimensions not shown.

identity-specific, track-level transition and likelihood functions (defined in (22) and (26)) are realized as

$$f_{t,X^{k+1}_i|X^k_i}(\mathbf{x}_i \mid \mathbf{x}'_i) = \frac{1}{\sqrt{(2\pi)^5 \mathrm{Det}(\Sigma^k_{P_i})}} e^{-\frac{1}{2}(\mathbf{x}_i - \mathbf{F}\mathbf{x}'_i)^{\mathrm{T}}(\Sigma_{P_i})^{-1}(\mathbf{x}_i - \mathbf{F}\mathbf{x}'_i)} \tag{49}$$

and

$$g_{i,\mathfrak{z}^k_t}(\mathbf{x}_i) = p^k_{\mathrm{f},t} + \frac{p^k_{\mathrm{d},t,i}}{\sqrt{(2\pi)^3 \mathrm{Det}(\Sigma^k_{M_i})}} e^{-\frac{1}{2}(\mathfrak{z}^k_t - \mathbf{H}\mathbf{x}_i)^{\mathrm{T}}(\Sigma_{M_i})^{-1}(\mathfrak{z}^k_t - \mathbf{H}\mathbf{x}_i)}, \tag{50}$$

with covariance matrices $\Sigma^k_{P_i}$ and $\Sigma^k_{M_i}$, respectively (recall that $p^k_{\mathrm{f},t}$ is zero). As each term in the first equation of (29) is Gaussian, the $K^k_{t,i}$ may be found analytically as

$$K^k_{t,i} = E \cdot e^{\frac{1}{2}(\mu'^{\mathrm{T}}\Sigma'^{-1}\mu' - \mu^{k\,\mathrm{T}}_{t,i}((\mathbf{F}\Sigma^k_{t,i}\mathbf{F}^{\mathrm{T}})^{-1} + \Sigma^{-1}_{P_i})\mu^k_{t,i} - \mathfrak{z}^{k\,\mathrm{T}}_t \Sigma^{-1}_{M_i}\mathfrak{z}^k_t)} \tag{51}$$

where

$$\begin{aligned} E &= \sqrt{\frac{\mathrm{Det}(\Sigma')}{(2\pi)^3 \mathrm{Det}(\Sigma_{M_i}) \mathrm{Det}((\Sigma^k_{t,i})^{-1} + \mathbf{F}^{\mathrm{T}}\Sigma^{-1}_{P_i}\mathbf{F})}} \\ \Sigma' &= ((\Sigma^k_{t,i})^{-1} + \mathbf{F}^{\mathrm{T}}\Sigma^{-1}_{P_i}\mathbf{F} + \mathbf{H}^{\mathrm{T}}\Sigma^{-1}_{M_i}\mathbf{H})^{-1} \\ \mu' &= \Sigma'[((\Sigma^k_{t,i})^{-1}\mathbf{F}^{-1} + \mathbf{F}^{\mathrm{T}}\Sigma^{-1}_{P_i})\mu^k_{t,i} + \mathbf{H}^{\mathrm{T}}\Sigma^{-1}_{M_i}\mathfrak{z}^k_i]. \end{aligned} \tag{52}$$

Bayesian updates (which involve computing the $k+1$ counterparts to the $p^k_{t,i}(\mathbf{x}_i)$, b^k_t, and $a^k_{t,i}$) are performed in accordance with (27) and (28) and are shown graphically in Fig. 3. At a given timestep k, the relationships between $\mathfrak{X}^k_i$ (actual state of identity i), $p^k_{t,i}(\mathbf{x}_i)$ (ith identity density at track t), $g_{i,\mathfrak{z}^k_t}(\mathbf{x}_i)$ (ith identity measurement likelihood function at track t), and $\mathbf{U}$ ('ground-truth' matrix that encodes the actual track-to-identity configuration) are illustrated in Fig. 4.

As a separate filter bank was maintained at each observed track, and a total of $|T_{\mathrm{O}}| \cdot |I|$ individual filters

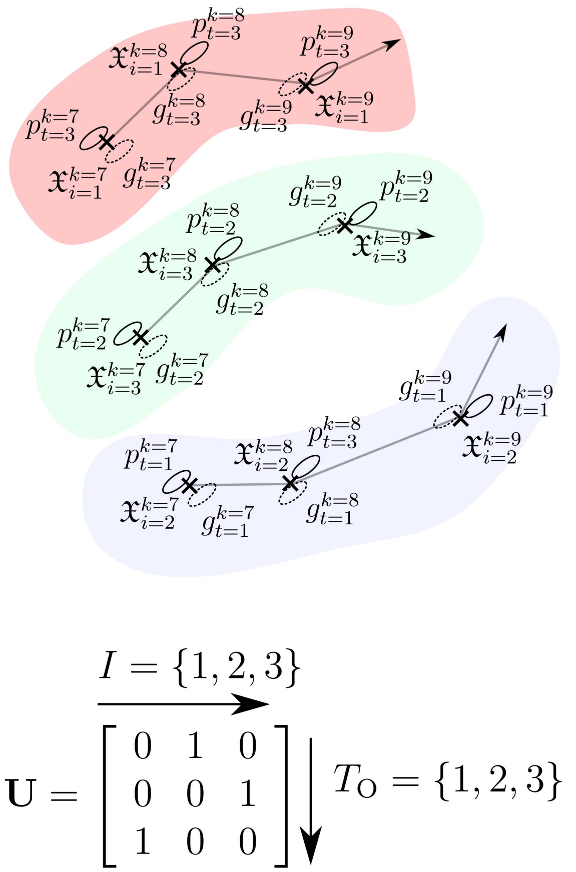

Fig. 4. Illustration of the relationship between the actual identity positions (crosses), projected position densities (solid ellipses), and measurement densities (dashed ellipses), which are labelled as $\mathfrak{X}_i^k$, p_t^k, and g_t^k, respectively, for the (arbitrarily chosen) timesteps $k = 7, 8, 9$. The latter are shorthand for the sets of functions $p_{t,i}^k(\mathbf{x}_i)$ (equation 48) and $g_{i,\mathfrak{z}_t^k}(\mathbf{x}_i)$ (equation 50), respectively. Similarly, each of their associated ellipses represents a set of ellipses indexed over I (recall that a bank of $|I|$ filters is run at every track, where each filter is programmed with the kinematic/static feature information of single $i \in I$). Remaining unknown to the estimation algorithms, the matrix $\mathbf{U}$ (equation 41)—shown here as the realization induced by an arbitrary order on I and T_{O}—contains the 'ground-truth' associations between identities and tracks that define a given simulation. Note that the velocity and static attribute components of p_t^k and $\mathfrak{X}_i^k$ are not shown.

were employed. In every example, targets were assigned a zero initial velocity, and a uniform prior was used in the remaining dimensions by setting $p_{t,i}^1(\mathbf{x}) = \mathbf{1}_{r_t}(\mathbf{x})$ and $\mathbf{M}_{t,i}^1 = 1$ for all t and i. The $\mathbf{P}_{t,i}^k$ were calculated with a parallel implementation of the speed-improved Ryser formula of §3-A using the GMP Arithmetic Library. In accordance with §3-B, the minimum significand size was found to be $n(1 + \log_2 e) \approx 244$, to which another 56 bits was added to furnish a working precision of approximately 16 significant digits.[16] Prior to calculating the permanent, small values of $a_{t,i}^k$ were rounded up to 10^{-100}, and the $\mathbf{M}^k$ were converted to doubly stochastic form using Sinkhorn scaling. The maximum-weight matchings were found with an $\mathcal{O}(n^3)$ implementation of the Hungarian method [59], and all processing was performed on an HP Z820 workstation.

When each identity is uniquely resolvable, the matrix of track-to-identity assignment probabilities $\mathbf{P}^k$ will asymptotically converge to $\mathbf{U}$ over the subset of observed tracks (i.e., $\mathbf{P}_{t,i}^k \to \mathbf{U}_{t,i}$ with increasing k for $t \in T_{\mathrm{O}}$), given adequate measurements and an initial prior that satisfies Cromwell's rule. However, when groups of targets are indistinguishable (e.g. common attributes and/or motion models), this is no longer holds. Assuming that identities of a given equivalence class are assigned equal initial values in each row of $\mathbf{M}^1$, the asymptotic matrix of track-to-identity assignment probabilities becomes

$$\mathbf{V}_{t,i} = \left\{ \begin{array}{cc} 1/|\mathcal{I}(i)| & \text{If any } i' \in \mathcal{I}(i) \text{ is at } t \\ 0 & \text{Otherwise} \end{array} \right\}, \tag{53}$$

where $\mathcal{I}(i) \subseteq I$ is the set of identities[17] that are indistinguishable from i. In cases where $|\mathcal{I}(i)| = 1$, adequate measurements will perfectly locate i, with the special case $\mathbf{V} = \mathbf{U}$ when this holds for all $i \in I$.

4) Comparisons to Approximate Methods:

The Bayesian algorithm developed in the previous sections was evaluated by computing the true posterior track-to-identity assignment probabilities $\mathbf{P}_{t,i}^k$ and comparing the results with the asymptotically-optimal assignments encoded in $\mathbf{V}^k$. Further comparisons were carried out against the track-to-identity assignments computed by approximation methods discussed in §1. The corresponding approximate assignment matrices (containing quasi posterior probabilities) are:

1) The local (track-level) weights given by

$$\tilde{\mathbf{P}}_{\mathrm{TL}}^k = \mathbf{D}^k \mathbf{M}^k, \tag{54}$$

where the diagonal matrix $\mathbf{D}^k$ normalizes the rows of $\mathbf{M}^k$ (i.e., $(\mathbf{D}_{t,t}^k)^{-1} = \sum_{i \in I} \mathbf{M}_{t,i}^k$). The elements in the tth row of $\mathbf{M}^k$ are identity/class coefficients maintained by the filter bank running at the tth track, and the scaling effected by $\mathbf{D}_{t,t}^k$ is the post-Bayesian-update normalization that arises by treating the tth track as a single-target JTC problem. The elements of $\tilde{\mathbf{P}}_{\mathrm{TL}}^k$ are thus the track-to-identity assignment probabilities computed by separate instances of single-target JTC (e.g. [9]–[11]) applied to each track.

2) The diagonally-scaled track-level weights given by

$$\tilde{\mathbf{P}}_{\mathrm{DS}}^k = \mathbf{D}_1^k \mathbf{M}^k \mathbf{D}_2^k, \tag{55}$$

where, as discussed in §3-B, $\mathbf{D}_1^k$ and $\mathbf{D}_2^k$ are positive diagonal matrices that yield a doubly-stochastic $\tilde{\mathbf{P}}_{\mathrm{DS}}^k$. This approximation was used in [20]–[22].

[16] Repeating the simulations with 10,000-bit mantissas yielded no significant change in the final results.

[17] Note that $i \in \mathcal{I}(i)$ and $\mathcal{I}(i') = \mathcal{I}(i)$ for all $i' \in \mathcal{I}(i)$.

TABLE I
Parameters of the example problem in §4-B.1

Description	Parameter	Value
Covariance (Process Noise)	Σ_i^{P}	$\lfloor i/2 \rfloor \begin{bmatrix} 0.25\mathbf{I} & 0.5\mathbf{I} & \mathbf{0} \\ 0.5\mathbf{I} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{bmatrix}$
Mean (Process Noise)	μ_i^{P}	$\mathbf{0}$
Covariance (Meas. Noise)	$\Sigma_{t,i}^{\mathrm{M}}$	$\begin{bmatrix} 4\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \infty\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 4+\infty((t-1) \mod 2) \end{bmatrix}$
Mean (Measurement Noise)	$\mu_{t,i}^{\mathrm{M}}$	$\mathbf{0}$
Static Attribute	A_i	i

Note that i, t, $\mathbf{I}$, and $\lfloor\cdot\rfloor$ are the identity number (1–100), track number (1–6), 2×2 identity matrix, and floor function, respectively. By definition, the process noise covariance is shared by pairs of successive identities, while the measurement noise covariance matrix renders static attributes unobservable in every second track. This may be seen by noting that the static attribute variance given by $4+\infty((t-1) \mod 2)$ evaluates to 4 and ∞ for odd and even t, respectively, where the latter case is completely non-informative.

3) The maximum-weight matching $\tilde{\mathbf{P}}_{\mathrm{MW}}^k$ defined as

$$(\tilde{\mathbf{P}}_{\mathrm{MW}}^k)_{t,i} = \begin{Bmatrix} 1, & i = \sigma_{\mathrm{MW}}^k(t) \\ 0, & i \neq \sigma_{\mathrm{MW}}^k(t) \end{Bmatrix}, \tag{56}$$

where

$$\sigma_{\mathrm{MW}}^k = \operatorname*{arg\,max}_{\sigma\in S_{|T|}} \prod_{t\in T} \mathbf{M}_{t,\sigma(t)}^k \tag{57}$$

and $S_{|T|}$ is the symmetric group of degree $|T|$. The matrix $\tilde{\mathbf{P}}_{\mathrm{MW}}^k$ serves as a (hard) maximum likelihood track-to-identity estimate over the set of all possible assignments and was used in [18], [19].

Discrepancies between the (quasi) probability distributions were quantified as the maximum Kullback-Liebler (KL) divergence over all observed tracks

$$\max_{t\in T} D_{\mathrm{KL}}(\alpha_t \| \beta_t) = \max_{t\in T} \sum_{i\in I} \ln\left(\frac{\alpha_{t,i}}{\beta_{t,i}}\right)\alpha_{t,i}, \qquad t \in T_{\mathrm{O}} \tag{58}$$

for the ordered pairs of matrix rows $(\alpha_t, \beta_t) = (\mathbf{V}_t, \mathbf{P}_t^k)$, $(\mathbf{V}_t, a\tilde{\mathbf{P}}_{\mathrm{TL}}^k)$, $(\mathbf{V}_t, \tilde{\mathbf{P}}_{\mathrm{DS}}^k)$, and $(\mathbf{V}_t, \tilde{\mathbf{P}}_{\mathrm{MW}}^k)$. Note that the summand in (58) yields the information gain (in nats) realized by substituting α_t for β_t as the set of identity assignment probabilities at track t.

B. Simulation Results

1) 100 Identities, 6 Observed Tracks, Incomplete Measurements:

In this example, $\mathcal{S} = \mathbb{S}_{\cdot,\mathrm{d}} \times \mathbb{S}_{\cdot,\mathrm{s}} = \mathbb{R}^4 \times \mathbb{R}$, $|I| = 100$, $|T_{\mathrm{O}}| = 6$, and $\mathbf{U} = \mathbf{I}$ (the identity matrix). The kinematics, static attributes, and measurement parameters of the identities are given in Table I, where tracks indexed 0–5 were made observable (by virtue of the fact that $\mathbf{U} = \mathbf{I}$, only the identites indexed 0–5 were actually observed). By definition, identities are endowed with unique static attributes but possess pairwise-common kinematic parameters (distinct kinematic characteristics are assigned to groups of two consecutively numbered identities). Furthermore, as the static attributes are only made observable for every second track, half of the single-target tracks should display persistent local identity ambiguities that fail to resolve with additional measurements. However, as the static attribute of one member in each pair of tracks is observable, *global* identity deconfliction will asymptotically resolve each track's identity exactly.

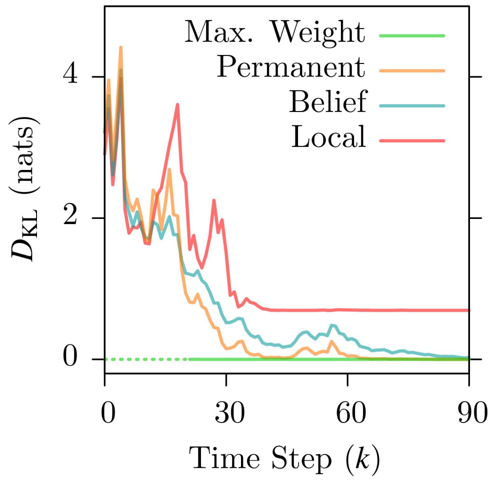

Fig. 5. Maximum Kullback-Liebler divergence (D_{KL}) over T_{O} (the set of observed tracks) as a function of time step for the problem of Table I. The maximum weight divergence is undefined for $k < 22$ (as denoted by a broken line). Note that only the local (track-level) assignment fails to converge, demonstrating that identity resolution can require non-local information.

The results of the simulation (which required 150 ms per timestep) are given in Fig. 5, which shows that the maximum KL divergences generally decrease with time step. Each of the global methods (permanent, belief matrix, and maximum-weight matching) converges to the correct identity-track permutation, as evidenced by respective maximum divergences that tend to zero. However, the maximum KL divergence computed using the local track probabilities converges to about 0.7 nats, reflecting the fact that the worst-performing local identifications—which occur for those tracks lacking observations of the static attribute—assign probabilities of ~ 0.5 to two identities that share the same kinematic properties. As expected, the statistically optimal matrix permanent algorithm outperforms the other soft assignment methods (which produce quasi probabilities). Finally, note that while the maximum KL divergence corresponding to the (hard) maximum-weight matching abruptly transitions from undefined to zero at $k = 22$, the soft assignments display more gradual convergence, behaviour that is broadly consistent with the differences between the respective classes of algorithms.

TABLE II
Parameters of the example problem in §4-B.2

Description	Parameter	Value
Covariance (Process Noise)	Σ_i^{P}	$\lfloor i/2 \rfloor \begin{bmatrix} 0.25\mathbf{I} & 0.5\mathbf{I} & \mathbf{0} \\ 0.5\mathbf{I} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 0 \end{bmatrix}$
Mean (Process Noise)	μ_i^{P}	$\mathbf{0}$
Covariance (Measurement Noise)	$\Sigma_{t,i}^{\mathrm{M}}$	$\begin{bmatrix} 4\mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \infty\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 4 \end{bmatrix}$
Mean (Measurement Noise)	$\mu_{t,i}^{\mathrm{M}}$	$\mathbf{0}$
Static Attribute	A_i	i

With the exception of the measurement noise covariance, which is defined in a manner that makes static attribute information visible in every track, these parameters are identical to those of the first example (Table I).

TABLE III
Parameters of the example problem in §4-B.3

Description	Parameter	Value
Covariance (Process Noise)	Σ_i^{P}	$i \begin{bmatrix} 0.25\mathbf{I} & 0.5\mathbf{I} \\ 0.5\mathbf{I} & \mathbf{I} \end{bmatrix}$
Mean (Process Noise)	μ_i^{P}	$\mathbf{0}$
Covariance (Measurement Noise)	$\Sigma_{t,i}^{\mathrm{M}}$	$\begin{bmatrix} 4\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \infty\mathbf{I} \end{bmatrix}$
Mean (Measurement Noise)	$\mu_{t,i}^{\mathrm{M}}$	$\mathbf{0}$
No. identities in Equivalence Class 1	$n(i = 1)$	60
No. identities in Equivalence Class 2	$n(i = 2)$	30
No. identities in Equivalence Class 3	$n(i = 3)$	10

The subscript i indexes the equivalence class. Note the absence of static attribute information in this example.

2) 100 Identities, 6 Observed Tracks, Complete Measurements:

This problem is a variation on the previous example, differing only by the measurement noise covariance (Table II), which now extends static attribute visibility to all observed tracks. In this case, both the local and global track-to-identity assignments should asymptotically converge to the actual track-to-identity configurations. Nonetheless, the global assignments are expected to exhibit improved pre-asymptotic characteristics, which may be germane to applications that require interim track-to-identity estimates (or simply do not run to convergence).

The KL divergences of this simulation (which also required 150 ms per timestep) are shown in Fig. 6. As expected, each of the maximum KL divergences tends to zero with increasing time step. However, between $k = 0$ and $k = 30$, there is significant discrepancy between the local and global identity assignments, supporting the assertion that global identity deconfliction improves target identification, even when identities are locally resolvable. As in the previous example, the permanent-based algorithm exhibits the fastest soft convergence, and the maximum-weight matching finds the correct assignment (in this case for $k \geq 9$). Finally, note the significant improvement in convergence rates as compared to that of §4-B.1, which may be ascribed to the information gained from doubling attribute measurements.

3) 100 Identities, 3 Identity Equivalence Classes:

In this example, $\mathcal{S} = \mathbb{S}_{\cdot,\mathrm{d}} = \mathbb{R}^4$, $|I| = 100$, $|T_{\mathrm{O}}| = 100$, and $\mathbf{U} = \mathbf{I}$. No static properties are visible, and identity dynamics are divided into three equivalence classes given in Table III (each identity's Markov process is described by one of three motion models). The results of this simulation (which required 270 ms per timestep) are given in Fig. 7, which shows that global deconfliction significantly improves the resolution of equivalence class after $k \sim 40$. Interestingly, the matrix-permanent algorithm performs only modestly better than the belief-matrix method, although both methods significantly outperform local assignment. Finally, in this example, convergence of the maximum-weight method appears substantially more uneven, finding the correct identity-track assignment at $k = 44$, then reverting to incorrect assignments at $k = 52$ and $k = 65$ before finally settling on correct assignment for $k \geq 66$.

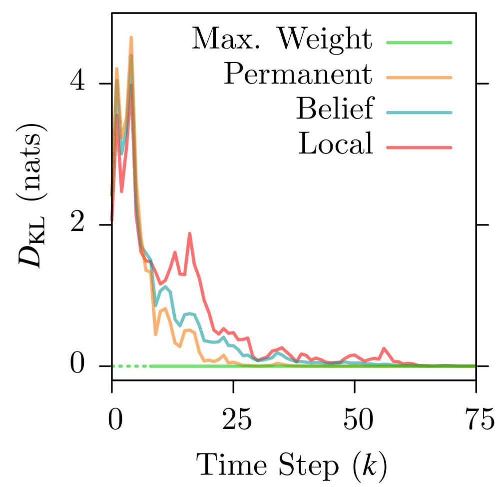

Fig. 6. Maximum Kullback-Liebler divergence (D_{KL}) over T_{O} as a function of time step for the problem of Table II. Once again, the maximum weight divergence is undefined for $k < 9$ (as denoted by a broken line).

5. CONCLUSION

This paper derived a rigorously Bayesian method for finding the optimal track-to-identity assignments for a group of targets that are well-separated or tracked using (J)PDA. Identification of targets is performed jointly across tracks to correctly account for the complex sta-

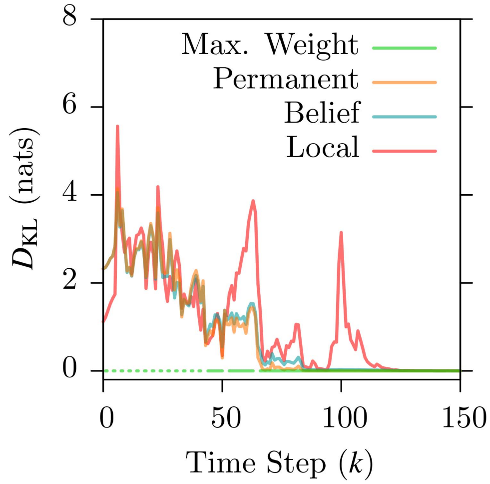

Fig. 7. Maximum Kullback-Leibler divergence (D_{KL}) over T_{O} as a function of time step for the problem of Table III. The maximum weight divergence is undefined for $k < 44$ and again for $k = 52$ and $k = 65$ (as denoted by broken lines).

tistical dependencies between track-to-identity assignments. The number of tracks need not equal the number of identities, and arbitrary feature and kinematic measurements may be used, provided that their corresponding sensor models can be characterized statistically. The problem naturally decomposes into local single-target tracking and classification and global combinatorial identity deconfliction, where the former is based on a unified measure-theoretic framework that treats tracking and classification on equal footing, and the latter reduces to computing the permanent of a non-negative matrix. While the computational complexity of the matrix permanent poses challenging implementation issues, Markov chain Monte Carlo methods may be used to find approximations in polynomial time. Furthermore, the existence of groups of targets that are indistinguishable, unobservable, or both allows the Ryser formula to be modified in a manner that improves computation speed. Reducing the complexity of approximating non-negative matrix permanents is an area of significant contemporary research, and advances in this field will directly benefit the performance of the algorithm described in this work.

ACKNOWLEDGEMENTS

The author expresses gratitude to D. J. Peters, G. R. Mellema, A. W. Isenor, and M. Farrell (DRDC, Atlantic Research Centre) for constructive comments and discussions during the preparation of this manuscript.

APPENDIX A MARKOV TRANSITION AND BAYESIAN UPDATE

Under conditions 1–5 of §2-B, the Markov transition step of the general filtering problem (BF.2) may be expanded as

$$
\begin{aligned}
&p_{X^k|Z^{k-1:1}}(\mathbf{x} \mid \mathfrak{z}^{k-1:1}) \\
&\quad = \int_{\mathbb{S}} f_{X^k|X^{k-1}}(\mathbf{x} \mid \mathbf{x}') p_{X^{k-1}|Z^{k-1:1}}(\mathbf{x}' \mid \mathfrak{z}^{k-1:1}) d\mu(\mathbf{x}') \\
&\quad = \int_{\mathbb{S}} \left[\prod_{i\in I}\sum_{t\in T} f_{t,X_i^k|X_i^{k-1}}(\mathbf{x}_i \mid \mathbf{x}_i')\right] \Big[C^{k-1}D^{k-1} \\
&\qquad \cdot \mathbf{1}_{A^{k-1}}(\mathbf{x}) \prod_{i\in I}\sum_{t\in T}(b_t^{k-1})^{-1} a_{t,i}^{k-1} p_{t,i}^{k-1}(\mathbf{x}_i')\Big] \, d\mu(\mathbf{x}').
\end{aligned}
\tag{59}
$$

Collecting product terms and noting the support of the $f_{t,X_i^k|X_i^{k-1}}(\mathbf{x}_i \mid \mathbf{x}_i')$ and $p_{t,i}^{k-1}(\mathbf{x}_i)$—given in (22) and (20), respectively—yields the simplification

$$
\begin{aligned}
p_{X^k|Z^{k-1:1}}(\mathbf{x} \mid \mathfrak{z}^{k-1:1}) &= C^{k-1}D^{k-1} \\
&\quad \cdot \mathbf{1}_{A^k}(\mathbf{x}) \prod_{i\in I}\sum_{t\in T}(b_t^{k-1})^{-1} a_{t,i}^{k-1} p_{t,i}^{k-0.5}(\mathbf{x}_i),
\end{aligned}
\tag{60}
$$

where

$$
p_{t,i}^{k-0.5}(\mathbf{x}_i) = \int_{r_t^{k-1}} f_{X_i^k|X_i^k}(\mathbf{x}_i \mid \mathbf{x}_i') p_{t,i}^{k-1}(\mathbf{x}_i') d\mu(\mathbf{x}_i') \tag{61}
$$

is the forward-in-time projection of $p_{t,i}^k(\mathbf{x}_i)$. The Bayesian update step may be similarly found by substituting (20) and (25) into the right-hand side of (BF.1) to produce

$$
\begin{aligned}
&p_{X^k|Z^{k:1}}(\mathbf{x} \mid \mathfrak{z}^{k:1}) \\
&\qquad \propto f_{Z^k|X^k}(\mathfrak{z}^k \mid \mathbf{x}) p_{X^k|Z^{k-1:1}}(\mathbf{x} \mid \mathfrak{z}^{k-1:1}) \\
&\quad = \Big[p_{\mathrm{f}}^k \mathbf{1}_{A^k}(\mathbf{x}) \prod_{i\in I}(\mathbf{1}_{r_t^k}(\mathbf{x}_i)(p_{\mathrm{f}}^k)^{-1} g_{i,\mathfrak{z}^k}(\mathbf{x}_i) \\
&\qquad\quad + \mathbf{1}_{\complement r_t^k}(\mathbf{x}_i))\Big] \\
&\qquad\quad \cdot \Big[C^{k-1}D^{k-1}\mathbf{1}_{A^k}(\mathbf{x}) \prod_{i\in I}\sum_{t\in T}(b_t^{k-1})^{-1} \\
&\qquad\quad \cdot a_{t,i}^{k-1} p_{t,i}^{k-0.5}(\mathbf{x}_i)\Big].
\end{aligned}
\tag{62}
$$

Using (21), equation (62) may be rewritten as

$$
\begin{aligned}
&p_{X^k|Z^{1:k}}(\mathbf{x} \mid \mathfrak{z}^{1:k}) = \\
&\quad C^{k-1}(D^{k-1}p_{\mathrm{f}}^k)\mathbf{1}_{A^k}(\mathbf{x}) \\
&\quad \cdot \prod_{i\in I}\Bigg[\left(\sum_{t\in T\setminus\{\mathbf{o}(k)\}} (b_t^{k-1})^{-1} a_{t,i}^{k-1} p_{t,i}^{k-1}(\mathbf{x}_i)\right) + \\
&\quad (b_{\mathbf{o}(k)}^{k-1} p_{\mathrm{f}}^k)^{-1}\left(a_{\mathbf{o}(k),i}^{k-1}\int_{r_{\mathbf{o}(k)}^k} p_{\mathbf{o}(k),i}^{k-0.5}(\mathbf{x}_i') g_{i,\mathfrak{z}^k}(\mathbf{x}_i) d\mu(\mathbf{x}_i')\right) \\
&\quad \cdot \left(\frac{p_{\mathbf{o}(k),i}^{k-0.5}(\mathbf{x}_i) g_{i,\mathfrak{z}^k}(\mathbf{x}_i)}{\int_{r_{\mathbf{o}(k)}^k} p_{\mathbf{o}(k),i}^{k-0.5}(\mathbf{x}_i') g_{i,\mathfrak{z}^k}(\mathbf{x}_i') d\mu(\mathbf{x}_i')}\right)\Bigg].
\end{aligned}
\tag{63}
$$

The terms $p_{t,i}^k(\mathbf{x}_i)$, b_t^k, and $a_{t,i}^k$ are thus related to their counterparts at the previous time step by (27) and (28) for $t = \mathbf{o}(k)$ and $t \neq \mathbf{o}(k)$, respectively. In particular, note that only the coefficients associated with the measured track $t = \mathbf{o}(k)$ are updated.

APPENDIX B EVALUATION OF POSTERIOR ASSIGNMENT PROBABILITIES

A. Track-to-Identity Probabilities as a Ratio of Permanents

The posterior track-to-identity assignment probabilities $\mathbf{P}_{t,i}^k$ are found by integrating the ith-identity marginalization of state's joint density function over over the region of $\mathbb{S}$ associated with the tth track. This may be expanded with (20) as

$$\begin{aligned}\mathbf{P}_{t,i}^k &= \int_{r_t^k} p_{X_i^k}(\mathbf{x}_i')d\mu(\mathbf{x}_i') \\ &= \int_{r_t^k}\int_{\mathbb{S}_{I\setminus\{i\}}} p_{X^k}(\mathbf{x}')d\mu(\mathbf{x}_i')d\mu(\mathbf{x}_{\mathbb{S}_{I\setminus\{i\}}}') \\ &= \int_{r_t^k}\int_{\mathbb{S}_{I\setminus\{i\}}} C^k D^k \mathbf{1}_{A^k}(\mathbf{x}') \\ &\quad\cdot \prod_{i'\in I}\sum_{t'\in T}(b_{t'}^k)^{-1} a_{t',i'}^k p_{t',i'}^k(\mathbf{x}_{i'}')d\mu(\mathbf{x}'). \end{aligned} \tag{64}$$

Noting the decompositions of the indicator function given in (17) and (18), equation (64) becomes

$$\begin{aligned}\mathbf{P}_{t,i}^k &= C^k D^k \int_{r_t^k}\int_{\mathbb{S}_{I\setminus\{i\}}} \sum_{\sigma\in P(I,T)}\prod_{i''\in I}\mathbf{1}_{\sigma(i'')}(\mathbf{x}_{i''}') \\ &\quad\cdot \prod_{i'\in I}\sum_{t'\in T}(b_{t'}^k)^{-1} a_{t',i'}^k p_{t,i}^k(\mathbf{x}_{i'}')d\mu(\mathbf{x}') \\ &= C^k D^k \int_{r_t^k}\int_{\mathbb{S}_{I\setminus\{i\}}} \sum_{\sigma\in P(I,T)}\prod_{i'\in I}\sum_{t'\in T}(b_{t'}^k)^{-1} \\ &\quad\cdot a_{t',i'}^k \mathbf{1}_{\sigma(i')}(\mathbf{x}_i') p_{t',i'}^k(\mathbf{x}_{i'}')d\mu(\mathbf{x}'). \end{aligned} \tag{65}$$

Using (21), this simplifies to

$$\begin{aligned}\mathbf{P}_{t,i}^k &= C^k D^k \int_{r_t^k}\int_{\mathbb{S}_{I\setminus\{i\}}} \sum_{\sigma\in P(I,T)}\prod_{i'\in I}(b_{\sigma(i')}^k)^{-1} \\ &\quad\cdot a_{\sigma(i'),i'}^k p_{\sigma(i'),i'}^k(\mathbf{x}_{i'}')d\mu(\mathbf{x}'). \end{aligned} \tag{66}$$

As the probability densities in (66) have unit-valued integrals, integration with respect to $\mathbf{x}_{\mathbb{S}_{I\setminus\{i\}}}$ yields

$$\begin{aligned}\mathbf{P}_{t,i}^k &= C^k D^k \int_{r_t^k}\sum_{\sigma\in P(I,T)} p_{\sigma(i),i}^k(\mathbf{x}_i')\prod_{i'\in I}(b_{\sigma(i')}^k)^{-1} \\ &\quad\cdot a_{\sigma(i'),i'}^k d\mu(\mathbf{x}_i'). \end{aligned} \tag{67}$$

The final integral zeros all terms not indexed by $\sigma(i) = t$, giving

$$\mathbf{P}_{t,i}^k = C^k D^k \sum_{\substack{\sigma\in P(I,T)\\ \sigma(i)=t}}\prod_{i'\in I}(b_{\sigma(i')}^k)^{-1} a_{\sigma(i'),i'}^k. \tag{68}$$

The constant C^k may be found by solving

$$\begin{aligned} 1 &= \int_{\mathbb{S}} p_X(\mathbf{x}')d\mu(\mathbf{x}') \\ &= \int_{\mathbb{S}} C^k D^k \mathbf{1}_{A^k}(\mathbf{x}') \\ &\quad\cdot \prod_{i'\in I}\sum_{t'\in T}(b_{t'}^k)^{-1} a_{t',i'}^k p_{t',i'}^k(\mathbf{x}_{i'}')d\mu(\mathbf{x}') \end{aligned} \tag{69}$$

for $(C^k)^{-1}$. With the exception of the final integral, this process mirrors the steps of (64)–(68), giving

$$C^k = D^k \sum_{\sigma\in P(I,T)}\prod_{i'\in I}(b_{\sigma(i')}^k)^{-1} a_{\sigma(i'),i'}^k. \tag{70}$$

Noting that T and I are endowed with total orders, $t\in T$ and $i\in I$ may be used to index the elements of a $|T|\times|I|$-dimensional matrix defined as

$$\mathbf{A}_{t,i}^k = a_{t,i}^k. \tag{71}$$

Similarly, a $|T|\times|T|$ diagonal matrix may be defined as

$$\mathbf{B}_{t,j}^k = \mathrm{diag}(b_t^k) = \left\{\begin{matrix} b_t^k, & t = j \\ 0, & t\neq j\end{matrix}\right\}, \tag{72}$$

where j has been used in lieu of i, given that $i \leq |I| \leq |T|$ (as described in §2-A). Equation (68) thus becomes

$$\mathbf{P}_{t,i}^k = C^k D^k \sum_{\substack{\sigma\in P(I,T)\\ \sigma(i)=t}}\prod_{i'\in I}((\mathbf{B}^k)^{-1}\mathbf{A}^k)_{\sigma(i'),i'}. \tag{73}$$

A new $|T|\times|T|$-dimensional augmented matrix may be constructed as

$$[(\mathbf{B}^k)^{-1}\mathbf{A}^k \mid \mathbf{J}], \tag{74}$$

where $\mathbf{J}$ is an $|T|\times(|T|-|I|)$-dimensional matrix of ones. Note that this matrix is empty when $|T|-|I| = 0$. Reformulated with respect to $[(\mathbf{B}^k)^{-1}\mathbf{A}^k \mid \mathbf{J}]$, equation (73) becomes

$$\mathbf{P}_{t,i}^k = \frac{C^k D^k}{(|T|-|I|)!}\sum_{\substack{\sigma\in S_{|T|}\\ \sigma(i)=t}}\prod_{j=1}^{|T|}[(\mathbf{B}^k)^{-1}\mathbf{A}^k \mid \mathbf{J}]_{\sigma(j),j}, \tag{75}$$

where $S_{|T|} = P(\mathbb{N}_{|T|},\mathbb{N}_{|T|})$ is the symmetric group of degree $|T|$, and $(|T|-|I|)!$ is the number of terms in $S_{|T|}$ that are associated with a single term in (the smaller) $P(I,T)$. Using the second equation in (29) and noting that D^k may be brought under the sum product by scaling the rows of $[\mathbf{B}^k\mathbf{A}^k \mid \mathbf{J}]$ by $\mathbf{B}^k$, equation (69) may be rewritten as

$$\begin{aligned}\mathbf{P}_{t,i}^k &= \frac{C^k}{(|T|-|I|)!}\sum_{\substack{\sigma\in S_{|T|}\\ \sigma(i)=t}}\prod_{j=1}^{|T|}[\mathbf{A}^k \mid \mathbf{B}^k\mathbf{J}]_{\sigma(j),j} \\ &= \frac{C^k}{(|T|-|I|)!}\sum_{\substack{\sigma\in S_{|T|}\\ \sigma(i)=t}}\prod_{j=1}^{|T|}\mathbf{M}_{\sigma(j),j}^k, \end{aligned} \tag{76}$$

where $\mathbf{M}^k$ is defined as

$$\mathbf{M}^k = [\mathbf{A}^k \mid \mathbf{B}^k\mathbf{J}]. \tag{77}$$

Finally, (76) may be simplified to

$$\mathbf{P}^k_{t,i} = \frac{C^k}{(|T|-|I|)!}\mathbf{M}^k_{t,i}\sum_{\sigma\in S_{|T|-1}}\prod_{j=1}^{|T|-1}\mathbf{M}^k_{\sigma(j),j}(t;i), \tag{78}$$

where $\mathbf{M}^k(t;i)$ is the matrix formed from $\mathbf{M}^k$ by deleting the tth row and ith column. Equation (70) may be similarly expressed as

$$C^k = \frac{1}{(|T|-|I|)!}\sum_{\sigma\in S_{|T|}}\prod_{j=1}^{|T|}\mathbf{M}^k_{\sigma(j),j}. \tag{79}$$

Each of the combinatorial sums in (78) and (79) conforms to a matrix permanent defined in (32), and marginal probability in (78) thus becomes the ratio

$$\mathbf{P}^k_{t,i} = \mathbf{M}^k_{t,i}\frac{\mathrm{Per}(\mathbf{M}^k(t;i))}{\mathrm{Per}(\mathbf{M}^k)}. \tag{80}$$

While the resulting matrix of marginal probabilities $\mathbf{P}^k$ is at least left stochastic, it becomes doubly stochastic when every track is certain to exist, a condition that results when each identity can be found at some track and vice versa. In this case, $|T| = |I|$, $\mathbf{J}$ is empty (the b_t coefficients are consequently irrelevant), and each track must therefore exist and correspond to some identity—regardless of the system's false alarm probabilities. When this is undesired, a set of hidden tracks may be defined in accordance with (1) of §2-A. Finally, note that groups of identities that are indistinguishable with respect to the suite of sensors form equivalence classes, the presence of which is manifested by repeated columns in $\mathbf{M}^k$.

B. Limiting Process for Zero False-Alarm Probabilities

The zero false-alarm probabilities that were temporarily reassigned non-zero values (Condition 5 in §2-B) may now be taken to zero in the limit. In what follows, the $p_{\mathrm{f}}^{k'\le k}$ in $\mathbf{M}^k$ of (78) are sequentially brought to zero for each $p_{\mathrm{f}}^{k'}$ for which a limit is required. This process commences by recursively expanding (BF.1) and (BF.2) over all $k'' \le k$ with the measurement function defined in (24). For any $p_{\mathrm{f}}^{k'}$ undergoing the limiting step, this yields

$$\begin{aligned}&\lim_{p_{\mathrm{f}}^{k'}\to 0^+} p_{X^k|Z^{1:k}}(\mathbf{x}^{k'} \mid \mathfrak{Z}^{1:k}) = \\ &\lim_{p_{\mathrm{f}}^{k'}\to 0^+}\cdots\int_{\mathbb{S}}\mathbf{1}_{A^k}(\mathbf{x}^{k'})\prod_{i\in I}\\ &(\mathbf{1}_{r_t^{k'}}(\mathbf{x}_i^{k'})(p_{\mathrm{f}}^{k'})^{-1}g_{i,\mathfrak{Z}^{k'}}(\mathbf{x}_i^{k'}) + \mathbf{1}_{\mathsf{C}r_t^{k'}}(\mathbf{x}_i^{k'}))\\ &\qquad\cdots d\mu(\mathbf{x}^{1:k-1}),\end{aligned} \tag{81}$$

where $1 \le k' \le k$ and the ellipses contain the measurement functions for $k'' \ne k'$ and the Markov transition densities and their associated integrals for all $k'' \le k$. Employing the equivalence between (23) and (24) results in the reformulation

$$\begin{aligned}\lim_{p_{\mathrm{f}}^{k'}\to 0^+} p_{X^k|Z^{1:k}}(\mathbf{x} \mid \mathfrak{Z}^{1:k}) &= \lim_{p_{\mathrm{f}}^{k'}\to 0^+}\left(\cdots\int_{\mathbb{S}}\mathbf{1}_{A^{k'}}(\mathbf{x}^{k'})\right.\\ &\left.\cdot\left(\sum_{i\in I}p^{k'}_{\mathrm{d},i}f_{Z^{k'}|X_S^{k'}}(\mathfrak{Z}^{k'} \mid \mathbf{x}_i^{k'}) + p_{\mathrm{f}}^{k'}\right)\cdots d\mu(\mathbf{x}^{1:k-1})\right).\end{aligned} \tag{82}$$

Noting that the summand in (82) has no dependence on $p_{\mathrm{f}}^{k'}$, the dominated convergence theorem [60] may be trivially applied to the constant $p_{\mathrm{f}}^{k'}$ term to give

$$\begin{aligned}&\lim_{p_{\mathrm{f}}^{k'}\to 0^+} p_{X^k|Z^{1:k}}(\mathbf{x} \mid \mathfrak{Z}^{1:k}) = \\ &\left(\cdots\int_{\mathbb{S}}\mathbf{1}_{A^{k'}}(\mathbf{x})\left(\sum_{i\in I}p^{k'}_{\mathrm{d},i}f_{Z^{k'}|X_S^{k'}}(\mathfrak{Z}^{k'} \mid \mathbf{x}_i^{k'})\right.\right.\\ &\left.\left.+\lim_{p_{\mathrm{f}}^{k'}\to 0^+}p_{\mathrm{f}}^{k'}\right)\cdots d\mu(\mathbf{x}^{1:k-1})\right)\\ &=\left(\cdots\int_{\mathbb{S}}\mathbf{1}_{A^{k'}}(\mathbf{x})\left(\sum_{i\in I}p^{k'}_{\mathrm{d},i}f_{Z^{k'}|X_S^{k'}}(\mathfrak{Z}^{k'} \mid \mathbf{x}_i^{k'})\right.\right.\\ &\left.\left.+0\right)\cdots d\mu(\mathbf{x}^{1:k-1})\right),\end{aligned} \tag{83}$$

and the relation

$$\lim_{p_{\mathrm{f}}^{k'}\to 0^+} p_{X^k|Z^{1:k}}(\mathbf{x} \mid \mathfrak{Z}^{1:k}) = p_{X^k|Z^{1:k}}(\mathbf{x} \mid \mathfrak{Z}^{1:k})|_{p_{\mathrm{f}}^{k'}=0} \tag{84}$$

thus holds. Equation (64) may now be written as

$$\begin{aligned}\mathbf{P}^k_{t,i}|_{p_{\mathrm{f}}^{k'}=0} &= \int_{r_t^k}p_{X_i^k|Z^{1:k}}(\mathbf{x}_i \mid \mathfrak{Z}^{1:k})|_{p_{\mathrm{f}}^{k'}=0}d\mu(\mathbf{x}_i)\\ &=\int_{r_t^k}\lim_{p_{\mathrm{f}}^{k'}\to 0^+}p_{X_i^k|Z^{1:k}}(\mathbf{x}_i \mid \mathfrak{Z}^{1:k})d\mu(\mathbf{x}_i).\end{aligned} \tag{85}$$

As the $p_{X_i^k|Z^{1:k}}(\mathbf{x}_i \mid \mathfrak{Z}^{1:k})$ may be assumed bounded, and because the limit in (84) is satisfied pointwise, the dominated convergence theorem may be applied a second time to yield

$$\mathbf{P}^k_{t,i}|_{p_{\mathrm{f}}^{k'}=0} = \lim_{p_{\mathrm{f}}^{k'}\to 0^+}\int_{r_t^k}p_{X_i^k|Z^{1:k}}(\mathbf{x}_i \mid \mathfrak{Z}^{1:k})d\mu(\mathbf{x}_i). \tag{86}$$

By way of the algebraic and integration steps between (64) and (78), the permanent ratio becomes

$$\mathbf{P}^k_{t,i}|_{p_{\mathrm{f}}^{k'}=0} = \lim_{p_{\mathrm{f}}^{k'}\to 0^+}\mathbf{M}^k_{t,i}\frac{\mathrm{Per}(\mathbf{M}^k(t;i))}{\mathrm{Per}(\mathbf{M}^k)}. \tag{87}$$

Provided that $\lim_{p_{\mathrm{f}}^{k'}\to 0^+}\mathbf{M}^k$ possesses at least one non-zero permutation, the denominator is positive, and the limit of (87) may distributed under the permanent operations through successive applications of the algebraic limit theorem [61]

$$\mathbf{P}^k_{t,i}|_{p_{\mathrm{f}}^{k'}=0} = (\lim_{p_{\mathrm{f}}^{k'}\to 0^+}\mathbf{M}^k_{t,i})\frac{\mathrm{Per}(\lim_{p_{\mathrm{f}}^{k'}\to 0^+}\mathbf{M}^k(t;i))}{\mathrm{Per}(\lim_{p_{\mathrm{f}}^{k'}\to 0^+}\mathbf{M}^k)}. \tag{88}$$

The limiting process thus amounts to calculating

$$\lim_{p_{\mathrm{f}}^{k'}\to 0^+} \mathbf{M}^k = \left[\lim_{p_{\mathrm{f}}^{k'}\to 0^+} \mathbf{A}^k \mid \lim_{p_{\mathrm{f}}^{k'}\to 0^+} \mathbf{B}^k\mathbf{J}\right]$$

$$= \left[\lim_{p_{\mathrm{f}}^{k'}\to 0^+} \mathbf{A}^k \mid \mathbf{B}^k\mathbf{J}|_{p_{\mathrm{f}}^{k'}=0}\right], \qquad (89)$$

where the second equality results from the fact that $\mathbf{B}^k\mathbf{J}$ submatrix contains only linear terms of $p_{\mathrm{f}}^{k'}$. The remaining limit in (89) may be found by calculating the limits to the individual $a_{t,i}^k$. Noting the definition of $\mathbf{o}(\cdot)$ in Condition 5 in §2-B and denoting its preimage by $\mathbf{o}^{-1}(\cdot)$, the recursion of (27) gives

$$\lim_{p_{\mathrm{f}}^{k'}\to 0^+} a_{t,i}^k = \lim_{p_{\mathrm{f}}^{k'}\to 0^+} a_{t,i}^1 \prod_{\substack{k''\in\mathbf{o}^{-1}(t)\\ k''\leq k}} K_i^{k''}, \qquad (90)$$

which, using (29), may be expanded as

$$\lim_{p_{\mathrm{f}}^{k'}\to 0^+} a_{t,i}^k =$$

$$\lim_{p_{\mathrm{f}}^{k'}\to 0^+} a_{t,i}^1 \prod_{\substack{k''\in\mathbf{o}^{-1}(t)\\ k''\leq k}} \left(\int_{\mathcal{S}^2} g_{i,3^{k''}}(\mathbf{x}_i'')\right.$$

$$\left.\cdot f_{X_i^{k''}|X_i^{k''-1}}(\mathbf{x}_i'' \mid \mathbf{x}_i') p_{t,i}^{k''-1}(\mathbf{x}_i') d\mu(\mathbf{x}_i'') d\mu(\mathbf{x}_i')\right). \qquad (91)$$

The $g_{i,3^{k''}}$ may be substituted with the right-hand side of (26), and after applying a process based on the steps of (81)–(84), equation (91) simplifies to

$$a_{t,i}^k|_{p_{\mathrm{f}}^{k'}=0} = a_{t,i}^1 \prod_{\substack{k''\in\mathbf{o}^{-1}(t)\\ k''\leq k}} K_i^{k''}|_{p_{\mathrm{f}}^{k'}=0}. \qquad (92)$$

Finally,

$$\mathbf{P}_{t,i}^k|_{p_{\mathrm{f}}^{k'}=0} = (\mathbf{M}_{t,i}^k|_{p_{\mathrm{f}}^{k'}=0}) \frac{\mathrm{Per}(\mathbf{M}^k(t;i)|_{p_{\mathrm{f}}^{k'}=0})}{\mathrm{Per}(\mathbf{M}^k|_{p_{\mathrm{f}}^{k'}=0})}, \qquad (93)$$

allowing $p_{\mathrm{f}}^{k'}$ to be set to zero in (26) and (27).

A zero row $\mathbf{B}_t^k$ signifies that at least one measurement with a zero false-alarm probability was encountered over the history of track t. The target corresponding to such a track is therefore certain to exist, and consequently, all track-identity permutations of non-zero weight include this track. Therefore, the $\mathbf{B}^k$ matrix may be seen as a tabulation of evidence supporting the existence of individual tracks by maintaining running products of track-level false alarm probabilities.

The denominator of (93) may become zero over the course of performing the limiting steps on the series of vanishing $p_{\mathrm{f}}^{k'}$, a condition that arises by an over-designation of zero false-alarm probabilities that renders the number of certain tracks larger than the number of identities. This scenario occurs when one or more measurement likelihood functions are defined in a manner that is incongruent with the properties of the statistical system. While the difficulty associated with undefined denominators may be superficially remedied by altogether avoiding zero false-alarm probabilities, a preferable solution entails addressing the underlying deficiencies in the affected measurement functions.

APPENDIX C RYSER FORMULA FOR MATRICES WITH DUPLICATE ROWS AND COLUMNS

Noting that members of the power set $Y \in \mathcal{P}(\mathbb{N}_n)$ may be brought into one-one correspondence with those of the set of n-dimensional 0-1 vectors ($\mathbf{c}_Y \in 2^n \equiv \prod_n\{0,1\}$), (35) may be rewritten as

$$\mathrm{Per}(\mathbf{M}) = (-1)^n \sum_{\mathbf{c}_Y\in 2^n} (-1)^{\|\mathbf{c}_Y\|_1} \prod_{q=1}^{n} \sum_{y\in Y} \mathbf{M}_{q,y}, \qquad (94)$$

where $\|\mathbf{c}_Y\|_1$ is the L^1 norm of $\mathbf{c}_Y$ (the sum of the entries in $\mathbf{c}_Y$), which is equivalent to the cardinality of Y. A further simplification results by expressing the trailing sum as a dot product

$$\mathrm{Per}(\mathbf{M}) = (-1)^n \sum_{\mathbf{c}_Y\in 2^n} (-1)^{\|\mathbf{c}_Y\|_1} \prod_{q=1}^{n} \mathbf{c}_Y \cdot \mathbf{M}_q, \qquad (95)$$

where $\mathbf{M}_q$ is the qth row of $\mathbf{M}$. In general, the set 2^n may be decomposed as the Cartesian product

$$2^n = 2^{n_1} \times 2^{n_2} \times \cdots \times 2^{n_{\hat{n}}} \qquad (96)$$

for any set of positive integers n_j that satisfies $n = \sum_{j=1}^{\hat{n}} n_j$. Under such a decomposition, $\|\mathbf{c}_Y\|_1 = \|c_1\|_1 + \|c_2\|_1 + \cdots + \|c_{\hat{n}}\|_1$, and $\mathbf{c}_Y = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_{\hat{n}})$. Consequently, (95) becomes

$$\mathrm{Per}(\mathbf{M}) = (-1)^n \sum_{\mathbf{c}_1\in 2^{n_1}} \sum_{\mathbf{c}_2\in 2^{n_2}} \cdots \sum_{\mathbf{c}_{\hat{n}}\in 2^{n_{\hat{n}}}}$$

$$(-1)^{\|c_1\|_1+\|c_2\|_1+\cdots+\|c_{\hat{n}}\|_1} \prod_{q=1}^{n} \mathbf{c}_Y \cdot \mathbf{M}_q. \qquad (97)$$

Each c_j may be associated with a distinct group of repeated columns in $\mathbf{M}$, provided that each group is formed by a single, contiguous submatrix.[18] In this case, the jth group of unique columns interacts only with $\mathbf{c}_j$ in the dot product of (97). By assigning the number of column groups in $\mathbf{M}$ to $\hat{n}$ and setting n_j to the multiplicity of the jth group, the permanent may be written as

$$\mathrm{Per}(\mathbf{M}) = (-1)^n \sum_{\mathbf{c}_1\in 2^{n_1}} \sum_{\mathbf{c}_2\in 2^{n_2}} \cdots \sum_{\mathbf{c}_{\hat{n}}\in 2^{n_{\hat{n}}}}$$

$$(-1)^{\|c_1\|_1+\|c_2\|_1+\cdots+\|c_{\hat{n}}\|_1}$$

$$\cdot \prod_{q=1}^{n} (\|\mathbf{c}_1\|_1, \|\mathbf{c}_2\|_1, \ldots, \|\mathbf{c}_{\hat{n}}\|_1) \cdot \bar{\mathbf{M}}_q', \qquad (98)$$

[18] This implies that each column belonging to the jth column group appears to the left of every column in the $(j+1)$th group. A matrix can always be brought into this form, as the permanent is invariant under column permutations.

where $\bar{\mathbf{M}}'_q$ is the qth row of the smaller matrix $\bar{\mathbf{M}}'$, which comprises the unique columns of $\mathbf{M}$. As the summands in (98) depend only on the L^1 norm of the $\mathbf{c}_j$, the sum may be rewritten as

$$\mathrm{Per}(\mathbf{M}) = (-1)^n \sum_{c_1=1}^{n_1} \binom{c_1}{n_1} \sum_{c_2=1}^{n_2} \binom{c_2}{n_2} \cdots \sum_{c_3=1}^{n_{\hat{n}}} \binom{c_{\hat{n}}}{n_{\hat{n}}} \cdot (-1)^{c_1+c_2+\cdots+c_{\hat{n}}} \prod_{q=1}^{n} (c_1, c_2, \ldots, c_{\hat{n}}) \cdot \bar{\mathbf{M}}'_q, \tag{99}$$

where the binomial coefficient $\binom{c_j}{n_j}$ gives the number of $\mathbf{c}_j$ vectors with an L^1 norm of c_j. Finally, the sums and binomial coefficients may be collected by defining $\mathbf{d} = (c_1, c_2, \ldots, c_{\hat{n}})$ to give

$$\mathrm{Per}(\mathbf{M}) = \sum_{\mathbf{d} \in \prod \mathbb{N}_{n_j}} (-1)^{n+\|\mathbf{d}\|_1} \left[\prod_{j=1}^{\hat{n}} \binom{d_j}{n_j} \right] \prod_{q=1}^{n} \mathbf{d} \cdot \bar{\mathbf{M}}'_q. \tag{100}$$

Inspection of (99)—which is functionally equivalent to (100)—reveals this reduced form to have a more favourable complexity of $\mathcal{O}(n \prod_{j=1}^{\hat{n}} (n_j + 1))$ when the sum are evaluated using a generalized Gray code (i.e., consecutive summation terms differ by only a single d_j in $\mathbf{d}$ such that $(d_j + 1 \mod n_j)$ yields the d_j of the subsequent term).

Finally, when the matrix possess repeated rows, the dot products $\mathbf{d} \cdot \bar{\mathbf{M}}'_{q_1}$ and $\mathbf{d} \cdot \bar{\mathbf{M}}'_{q_2}$ are equal when q_1 and q_2 belong to a common row group. The identical terms may be collected and replaced with a single term exponentiated to the size of the row group. Equation (100) then becomes

$$\mathrm{Per}(\mathbf{M}) = \sum_{\mathbf{d} \in \prod \mathbb{N}_{n_j}} (-1)^{n+\|\mathbf{d}\|_1} \cdot \left[\prod_{j=1}^{\hat{n}} \binom{d_j}{n_j} \right] \prod_{q=1}^{\hat{m}} (\mathbf{d} \cdot \bar{\mathbf{M}}_q)^{m_q}. \tag{101}$$

where $\hat{m}$ is the number of unique rows, m_q is the size of the qth group of duplicate rows, and $\bar{\mathbf{M}}_q$ is the qth row of matrix $\bar{\mathbf{M}}$ comprising only unique rows (and columns). The time complexity of this simplification is $\mathcal{O}(\hat{m} \prod_j (n_j + 1))$. Finally, as the permanent is invariant under matrix transposition, the minimum complexity becomes equation (37).

REFERENCES

[1] S. Challa, M. R. Morelande, D. Mušicki, and R. J. Evans
Fundamentals of Object Tracking.
Cambridge: Cambridge University Press, 2011.

[2] S. Challa, R. J. Evans, and D. Musicki
Target Tracking—A Bayesian Perspective,
in *Fourteenth International Conference on Digital Signal Processing*, vol. 1. IEEE, 2002, pp. 437–440.

[3] I. R. Goodman, R. P. Mahler, and H. T. Nguyen
Mathematics of Data Fusion.
Dordrecht, The Netherlands: Kluwer Academic Publishers, 1997.

[4] R. E. Kalman
A New Approach to Linear Filtering and Prediction Problems,
Transactions of the ASME—Journal of Basic Engineering, vol. 82, no. Series D, pp. 35–45, 1960.

[5] C. K. Chui and G. Chen
Kalman Filtering: With Real-Time Applications.
Springer, 2009.

[6] S. S. Blackman
Multiple-Target Tracking with Radar Applications.
Dedham, MA: Artech House, 1986.

[7] M. Mallick, V. Krishnamurthy, and B.-N. Vo
Integrated Tracking, Classification, and Sensor Management: Theory and Applications.
Wiley-IEEE Press, 2012.

[8] R. P. Mahler
Statistical Multisource-Multitarget Information Fusion.
Norwood, MA: Artech House, 2007.

[9] B. Ristic, N. Gordon, and A. Bessell
On Target Classification using Kinematic Data,
Information Fusion, vol. 5, no. 1, pp. 15–21, 2004.

[10] M. L. Krieg
Joint Multi-Sensor Kinematic and Attribute Tracking using Bayesian Belief Networks,
in *Proceedings of the 2003 International Conference on Information Fusion*, 2003, pp. 17–24.

[11] ———
"Kinematic and Attribute Fusion Using a Bayesian Belief Network Framework,"
Defence Science and Technology Organisation, PO Box 1500 Edinburgh, South Australia 5111, Australia, Tech. Rep., 2006.

[12] W. Mei, G.-l. Shan, and Y.-f. Wang
A second-order uncertainty model for target classification using kinematic data,
Information Fusion, vol. 12, no. 2, pp. 105–110, 2011.

[13] D. Angelova and L. Mihaylova
Joint Target Tracking and Classification with Particle Filtering and Mixture Kalman Filtering using Kinematic Radar Information,
Digital Signal Processing, vol. 16, no. 2, pp. 180–204, 2006.

[14] F. Palmieri and D. Ciuonzo
Data Fusion with Entropic Priors,
Proceedings of 20th Workshop on Neural Networks, WIRN, Frontiers in Artificial Intelligence and Applications, vol. 226, pp. 107–114, 2011.

[15] J. Teti Jr, R. Gorman, and W. Berger
A Multifeature Decision Space Approach to Radar Target Identification,
Aerospace and Electronic Systems, IEEE Transactions on, vol. 32, no. 1, pp. 480–487, 1996.

[16] S. Challa and G. W. Pulford
Joint Target Tracking and Classification Using Radar and ESM Sensors,
IEEE Transactions on Aerospace and Electronic Systems, vol. 37, no. 3, pp. 1039–1055, 2001.

[17] B. T. Vo and B. N. Vo
Tracking, Identification, and Classification with Random Finite Sets,
in *SPIE Defense, Security, and Sensing*. International Society for Optics and Photonics, 2013, pp. 87 450D–87 450D.

[18] J. Shin, N. Lee, S. Thrun, and L. Guibas
Lazy Inference on Object Identities in Wireless Sensor Networks,
in *Proceedings of the 4th international symposium on Information processing in sensor networks*. IEEE Press, 2005, p. 23.

[19] B. Schumitsch, S. Thrun, G. Bradski, and K. Olukotun
The Information-Form Data Association Filter,
in *Advances in Neural Information Processing Systems*, 2005, pp. 1193–1200.

[20] J. Shin, L. J. Guibas, and F. Zhao
A Distributed Algorithm for Managing Multi-Target Identities in Wireless Ad-Hoc Sensor Networks,
in *Information Processing in Sensor Networks*. Springer, 2003, pp. 223–238.

[21] H. Balakrishnan, I. Hwang, and C. J. Tomlin
Polynomial Approximation Algorithms for Belief Matrix Maintenance in Identity Management,
in *43rd IEEE Conference on Decision and Control*, vol. 5. IEEE, 2004, pp. 4874–4879.

[22] I. Hwang, K. Roy, H. Balakrishnan, and C. Tomlin
A Distributed Multiple-Target Identity Management Algorithm in Sensor Networks,
in *43rd IEEE Conference on Decision and Control*, vol. 1. IEEE, 2004, pp. 728–734.

[23] B. Schumitsch, S. Thrun, L. J. Guibas, and K. Olukotun
The Identity Management Kalman Filter (IMKF),
in *Robotics: Science and Systems*. Citeseer, 2006.

[24] J. Huang, C. Guestrin, and L. Guibas
Fourier theoretic probabilistic inference over permutations,
Journal of Machine Learning Research, vol. 10, pp. 997–1070, 2009.

[25] ———
Efficient Inference for Distributions on Permutations,
in *Advances in Neural Information Processing Systems*, 2007, pp. 697–704.

[26] R. Kondor, A. Howard, and T. Jebara
Multi-Object Tracking with Representations of the Symmetric Group,
in *International Conference on Artificial Intelligence and Statistics*, 2007, pp. 211–218.

[27] A. S. Asratian
Bipartite Graphs and their Applications.
Cambridge University Press, 1998, no. 131.

[28] R. E. Burkard, M. Dell'Amico, and S. Martello
Assignment Problems, Revised Reprint.
Philadelphia: Society for Industrial and Applied Mathematics, 2009.

[29] A. B. Badiru and O. A. Omitaomu
Computational Economic Analysis for Engineering and Industry.
Boca Raton, FL: CRC Press, 2010.

[30] M. Fujita and J.-F. Thisse
Economics of Agglomeration: Cities, Industrial Location, and Globalization.
New York: Cambridge University Press, 2013.

[31] C. K. Mustafi
Operations Research methods and practice.
New Age International, 1996.

[32] H. S. Kasana and K. D. Kumar
Introductory Operations Research: Theory and Applications.
Berlin Heidelberg: Springer-Verlag, 2004.

[33] J. Collins and J. Uhlmann
Efficient Gating in Data Association with Multivariate Gaussian Distributed States,
IEEE Transactions on Aerospace and Electronic Systems, vol. 28, no. 3, pp. 909–916, July 1992.

[34] M. Jerrum, A. Sinclair, and E. Vigoda
A Polynomial-Time Approximation Algorithm for the Permanent of a Matrix with Non-Negative Entries,
in *Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing*. ACM, 2001, pp. 712–721.

[35] ———
A Polynomial-Time Approximation Algorithm for the Permanent of a Matrix with Nonnegative Entries,
Journal of the ACM (JACM), vol. 51, no. 4, pp. 671–697, 2004.

[36] I. Bezáková, D. Štefankovic, V. V. Vazirani, and E. Vigoda
"Approximating the Permanent in O (n7) Time,"
2004.

[37] E. W. Weisstein
CRC Concise Encyclopedia of Mathematics.
Boca Raton, FL: CRC Press, 2003.

[38] L. Hogben
Handbook of Linear Algebra.
Boca Raton, FL: Chapman and Hall/CRC, 2006.

[39] O. Goldreich
Computational Complexity: A Conceptual Perspective.
New York: Cambridge University Press, 2008.

[40] T. Fortmann and Y. Bar-Shalom
Tracking and Data Association.
San Diego: Academic Press, 1988.

[41] L. D. Stone, R. L. Streit, T. L. Corwin, and K. L. Bell
Bayesian Multiple Target Tracking.
Norwood, MA: Artech House, 2013.

[42] T. Granlund
GNU MP, 5th ed., Free Software Foundation, 2013.
[Online]. Available: https://gmplib.org/manual/.

[43] S. Friedland
A Lower Bound for the Permanent of a Doubly Stochastic Matrix,
Annals of Mathematics, pp. 167–176, 1979.

[44] Lindley, Dennis V
Understanding Uncertainty.
Hoboken, NJ: John Wiley & Sons, 2006.

[45] A. W. Marshall and I. Olkin
Scaling of Matrices to Achieve Specified Row and Column Sums,
Numerische Mathematik, vol. 12, no. 1, pp. 83–90, 1968.

[46] R. Sinkhorn
A Relationship Between Arbitrary Positive Matrices and Doubly Stochastic Matrices,
The Annals of Mathematical Statistics, vol. 35, no. 2, pp. 876–879, 1964.

[47] M. Fürer
Faster Integer Multiplication,
SIAM Journal on Computing, vol. 39, no. 3, pp. 979–1005, 2009.

[48] B. Kalantari and L. Khachiyan
On the Complexity of Nonnegative-Matrix Scaling,
Linear Algebra and its Applications, vol. 240, pp. 87–103, 1996.

[49] L. G. Valiant
The Complexity of Computing the Permanent,
Theoretical Computer Science, vol. 8, no. 2, pp. 189–201, 1979.

[50] I. Bezáková, D. Štefankovič, V. V. Vazirani, and E. Vigoda
Accelerating Simulated Annealing for the Permanent and Combinatorial Counting Problems,
in *Proceedings of the Seventeenth Annual ACM-SIAM Symposium on Discrete Algorithms*. ACM, 2006, pp. 900–907.

[51] I. Bezáková, D. Štefankovic, V. V. Vazirani, and E. Vigoda
Accelerating Simulated Annealing for the Permanent and Combinatorial Counting Problems,
SIAM Journal on Computing, vol. 37, no. 5, pp. 1429–1454, 2008.

[52] M. Huber and J. Law
Fast Approximation of the Permanent for Very Dense Problems,
in *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2008, pp. 681–689.

[53] B. Huang and T. Jebara
"Approximating the Permanent with Belief Propagation,"
arXiv preprint arXiv:0908.1769, 2009.

[54] P. O. Vontobel
The Bethe Permanent of a Non-Negative Matrix,
in *48th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2010, pp. 341–346.

[55] Y. Watanabe and M. Chertkov
Belief Propagation and Loop Calculus for the Permanent of a Non-Negative Matrix,
Journal of Physics A: Mathematical and Theoretical, vol. 43, no. 24, p. 242002, 2010.

[56] J. L. Williams and R. A. Lau
"Approximate Evaluation of Marginal Association Probabilities with Belief Propagation,"
arXiv preprint arXiv:1209.6299, 2012.

[57] M. Chertkov and A. B. Yedidia
"Computing the Permanent with Belief Propagation,"
arXiv preprint arXiv:1108.0065, 2011.

[58] Y. Bar-Shalom and X. R. Li
Estimation and Tracking: Principles, Techniques, and Software.
Boston: Artech House, 1993.

[59] K. L. Stern
"Hungarian Method C++ Source Code,"
Online, 2012. [Online]. Available: https://github.com/KevinStern/software-and-algorithms-cpp/blob/master/src/hungarian.h.

[60] M. Capinski and P. E. Kopp
Measure, Integral and Probability.
London: Springer, 2004.

[61] F. Lárusson
Lectures on Real Analysis.
New York: Cambridge University Press, 2012, vol. 21.

Dominic Schaub received his B.Sc. in Computer Engineering and Ph.D. in Electrical Engineering from the University of Manitoba in 2004 and 2011, respectively. In 2011 he joined Defence Research and Development Canada, Atlantic Research Centre.