

# A Complete Variational Tracker

**Authored by:**

Ryan D. Turner  
Steven Bottone  
Bhargav Avasarala

## **Abstract**

We introduce a novel probabilistic tracking algorithm that incorporates combinatorial data association constraints and model-based track management using variational Bayes. We use a Bethe entropy approximation to incorporate data association constraints that are often ignored in previous probabilistic tracking algorithms. Noteworthy aspects of our method include a model-based mechanism to replace heuristic logic typically used to initiate and destroy tracks, and an assignment posterior with linear computation cost in window length as opposed to the exponential scaling of previous MAP-based approaches. We demonstrate the applicability of our method on radar tracking and computer vision problems.

## **1 Paper Body**

We introduce a novel probabilistic tracking algorithm that incorporates combinatorial data association constraints and model-based track management using variational Bayes. We use a Bethe entropy approximation to incorporate data association constraints that are often ignored in previous probabilistic tracking algorithms. Noteworthy aspects of our method include a model-based mechanism to replace heuristic logic typically used to initiate and destroy tracks, and an assignment posterior with linear computation cost in window length as opposed to the exponential scaling of previous MAP-based approaches. We demonstrate the applicability of our method on radar tracking and computer vision problems. The field of tracking is broad and possesses many applications, particularly in radar/sonar [1], robotics [14], and computer vision [3]. Consider the following problem: A radar is tracking a flying object, referred to as a target, using measurements of range, bearing, and elevation; it may also have Doppler measurements of radial velocity. We would like to construct a track which estimates the trajectory of the object over time. The Kalman filter [16], or a more general state space model, is used to filter out measurement errors. The key difference between tracking and filtering is the presence of clutter (noise measurements) and missed detections of true objects. We must determine which

measurement to ‘‘plug in’’ to the filter before applying it; this is known as data association. Additionally complicating the situation is that we may be in a multi-target tracking scenario in which there are multiple objects to track and we do not know which measurement originated from which object. There is a large body of work on tracking algorithms given its standing as a long-posed and important problem. Algorithms vary primarily on their approach to data association. The dominant approach uses a sliding window MAP estimate of the measurement-to-track assignment, in particular the multiple hypothesis tracker (MHT) [1]. In the standard MHT, at every frame the algorithm finds the most likely matching of measurements to tracks, in the form of an assignment matrix, under a one-to-one constraint (see Figure 1). One track can only result in one measurement, and vice versa, which we refer to as framing constraints. As is typical in MAP estimation, once an assignment is determined, the filters are updated and the tracker proceeds as if these assignments were known to be correct. The one-to-one constraint makes MAP estimation a bipartite matching task where algorithms exist to solve it exactly in polynomial time in the number of tracks  $NT$  [15]. However, the multi-frame MHT finds the joint MAP assignment over multiple frames, in which case the assignment problem is known to be NP-hard, although good approximate solvers exist [20].

S1  
S2  
S3  
Assignment Matrices A1  
A2  
A3  
(all) Track States X1  
X2  
X3  
Measurements Z1  
Z2  
Z3  
Meta-states Track Swap  
track 3 (Cesna) track 2 (777)  
clutter (birds) track 1 (747)  
?  
Sk Ak  
?  
Xk Zk

Figure 1: Simple scenario with a track swap: filtered state estimates  $\hat{x}$ , associated measurements  $z$ , and clutter  $c$ ; and corresponding graphical model. Note that  $X_k$  is a matrix since it contains state vectors for all three tracks.

1

Despite the complexity of the MHT, it only finds a sliding window MAP estimate of measurement-to-track assignments. If a clutter measurement is by chance associated with a track for the duration of a window then the tracker will assume with certainty that the measurement originated from that track, and never reconsider despite all future evidence to the contrary. If multiple

clutter (or otherwise incorrect) measurements are associated with a track, then it may veer off into space and result in spurious tracks. Likewise, an endemic problem in tracking is the issue of track swaps, where two trajectories can cross and get mixed up as shown in Figure 1. Alternatives to the MAP approach include the probabilistic MHT (PMHT) [9, Ch. 4] and probabilistic data association (PDA). However, the PMHT drops the one-to-one constraint in data association and the PDA only allows for a single target. This led to the development of the joint PDA (JPDA) algorithm for multiple targets, which utilizes heuristic calculations of the assignment weights and does not scale to multiple frame assignment. Particle filter implementations of the JPDA have tried to alleviate these issues, but they have not been adopted into real-time systems due to their inefficiency and lack of robustness. The probability hypothesis density (PHD) filter [19] addresses many of these issues, but only estimates the intensity of objects and does not model full trajectories; this is undesirable since the identity of an object is required for many applications including the examples in this paper. Lazaro-Gredilla et al. [18] made the first attempt at a variational Bayes (VB) tracker. In their approach every trajectory follows a Gaussian process (GP); measurements are thus modeled by a mixture of GPs. We develop additional VB machinery to retain the framing constraints, which are dropped in Lazaro-Gredilla et al. [18] despite being viewed as important in many systems. Secondly, our algorithm utilizes a state space approach (e.g. Kalman filters) to model tracks, providing linear rather than cubic time complexity in track length. Hartikainen and Sarkka [11] showed by an equivalence that there is little loss of modeling flexibility by taking a state space approach over GPs. Most novel tracking algorithms neglect the critical issue of track management. Many tracking algorithms unrealistically assume that the number of tracks  $N_T$  is known a priori and fixed. Additional wrapper logic is placed around the trackers to initiate and destroy tracks. This logic involves many heuristics such as M-of-N logic [1, Ch. 3]. Our method replaces these heuristics in a model-based manner to make significant performance gains. We call our method a complete variational tracker as it simultaneously does inference for track management, data association, and state estimation. The outline of the paper is as follows: We first describe the full joint probability distribution of the tracking problem in Section 1. This includes how to solve the track management problem by augmenting tracks with an active/dormant state to address the issue of an unknown number of tracks. By studying the full joint we develop a new conjugate prior on assignment matrices in Section 2. Using this new formulation we develop a variational algorithm for estimating the measurement-to-track assignments and track states in Section 3. To retain the framing constraints and efficiently scale in tracks and measurements, we modify the variational lower bound in Section 4 using a Bethe entropy approximation. This results in a loopy belief propagation (BP) algorithm being used as a subroutine in our method. In Sections 5-6 we show the improvements our method makes on a difficult radar tracking example and a real data computer vision problem in sports. Our paper presents the following novel contributions: First, we develop the first efficient deterministic approximate inference algorithm for

solving the full tracking problem, which includes the framing constraints and track management. The most important observation is that the VB assignment posterior has an induced factorization over time with regard to assignment matrices. Therefore, the computational cost of our variational approach is linear in window length as opposed to the exponential cost of the MAP approach. The most astounding aspect is that by introducing a weaker approximation (VB factorization vs MAP) we lower the computational cost from exponential to linear; this is a truly rare and noteworthy example. Second, in the process, we develop new approximate inference methods on assignment matrices and a new conjugate assignment prior (CAP). We believe these methods have much larger applicability beyond our current tracking algorithm. Third, we develop a process to handle the track management problem in a model-based way.

1

#### Model Setup for the Tracking Problem

In this section we describe the full model used in the tracking problem and develop an unambiguous notation. At each time step  $k \in \{1, \dots, N_1\}$ , known as a frame, we observe  $N_Z(k) \in \{0, \dots, N_0\}$  measurements,  $N_Z(k)$  in a matrix  $Z_k = \{z_{j,k}\}_{j=1}^{N_Z(k)}$ , from both real targets and clutter (spurious measurements). In the 2

radar example  $z_{j,k} \in \mathbb{R}^3$  is a vector of position measurements in  $\mathbb{R}^3$ . In data association we estimate the assignment matrices  $A$ , where  $A_{ij} = 1$  if and only if track  $i$  is associated with measurement  $j$ . Recall that each track is associated with at most one measurement, and vice versa, implying:  $\sum_{i=1}^{N_T} A_{ij} = 1, j \in \{1, \dots, N_Z\}, j=0$

$i=0$

The zero indices of  $A \in \{0, 1\}^{N_T \times (N_Z + 1)}$  are the ‘dummy row’ and ‘dummy column’ to represent the assignment of a measurement to clutter and the assignment of a track to a missed detection. Distribution on Assignments Although not explicitly stated in the literature, a careful examination of the cost functions used in the MAP optimization in MHT yields a particular and intuitive prior on the assignment matrices. The number of tracks  $N_T$  is assumed known a priori and  $N_Z$  is random. The corresponding generative process on assignment matrices is as follows: 1) Start with a one-to-one mapping from measurements to tracks:  $A \in \{0, 1\}^{N_T \times N_Z}$ . 2) Each track is observed with probability  $PD \in [0, 1]$ . Only keep the columns of detected tracks:  $A \leftarrow A(:, d)$ ,  $d_i \sim \text{Bernoulli}(PD(i))$ . 3) Sample a Poisson number of clutter measurements (columns):  $A \leftarrow [A, 0_{N_T \times N_c}]$ ,  $N_c \sim \text{Poisson}(\cdot)$ . 4) Use a random permutation vector  $\pi$  to make the measurement order arbitrary:  $A \leftarrow A(:, \pi)$ . 5) Append a dummy row and column on  $A$  to satisfy the summation constraints (1). This process gives the following normalized prior on assignments:  $\sum_{i=1}^{N_T} \sum_{j=1}^{N_Z} A_{ij} = N_c \exp(\cdot) / N_Z! PD(i)^{d_i} (1 - PD(i))^{1-d_i}$ . (2)  $i=1$

Note that the detections  $d$ ,  $N_Z$ , and clutter measurement count  $N_c$  are deterministic functions of  $A$ . Track Model We utilize a state space formulation over  $K$  time steps. The latent states  $x_{1:K} \in \mathbb{X}^K$  follow a Markov process, while the measurements  $z_{1:K} \in \mathbb{Z}^K$  are iid conditional on the track state:  $K \times K \times Y \times Y$   $p(z_{1:K}, x_{1:K}) = p(x_1) p(x_k - x_{k-1}) p(z_k - x_k)$ , (3)  $k=2$

$k=1$

where we have dropped the track and measurements indices  $i$  and  $j$ . Although more general models are possible, within this paper each track independently follows a linear system (i.e. Kalman filter):  $p(\mathbf{x}_k | \mathbf{x}_{k-1}) = N(\mathbf{x}_k - \mathbf{F}\mathbf{x}_{k-1}, \mathbf{Q})$ ,  $p(\mathbf{z}_k | \mathbf{x}_k) = N(\mathbf{z}_k - \mathbf{H}\mathbf{x}_k, \mathbf{R})$ . (4) **Track Meta-states** We address the track management problem by augmenting track states with a two-state Markov model with an active/dormant meta-state  $s_k$  in a 1-of- $N$  encoding:  $K \times Y \times P(s_{1:K}) = P(s_1) \prod_{k=2}^K P(s_k | s_{k-1})$ ,  $s_k \in \{0, 1\}^{NS}$ . (5)

This effectively allows us to handle an unknown number of tracks by making  $NT$  arbitrarily large;  $PD$  is now a function of  $s$  with a very small  $PD$  in the dormant state and a larger  $PD$  in the active state. Extensions with a larger number of states  $NS$  are easily implementable. We refer to the collection of track meta-states over all tracks at frame  $k$  as  $S_k := \{s_{i,k}\}_{i=1}^N$ ; likewise,  $X_k := \{x_{i,k}\}_{i=1}^N$ . **Full Model** We combine the assignment process and track models to get the full model joint:  $K \times Y \times p(Z_{1:K}, X_{1:K}, A_{1:K}, S_{1:K}) = \prod_{k=1}^K p(\mathbf{z}_k | \mathbf{x}_k, \mathbf{A}_k) p(\mathbf{x}_k | \mathbf{x}_{k-1}) P(S_k | S_{k-1}) P(\mathbf{A}_k | S_k)$  (6)

$$= \prod_{k=1}^K \sum_{Y} \sum_{NT} \sum_{NZ(k)} p(x_{i,k} | x_{i,k-1}) P(s_{i,k} | s_{i,k-1}) \sum_{j=1}^N \sum_{k} p_0(z_{j,k} | A_{0j}) \sum_{k} p(z_{j,k} | x_{i,k}, A_{kij} = 1) A_{ij},$$

where  $p_0$  is the clutter distribution, which is often a uniform distribution. The traditional goal in tracking is to compute  $p(X_k | Z_{1:k})$ , the exact computation of which is intractable due to the ‘combinatorial explosion’ in summing out the assignments  $A_{1:k}$ . The MHT MAP-based approach [1] for a sliding window  $w = k_2 - k_1 + 1$  tackles this with  $P(\mathbf{A}_{k_1:k_2} | Z_{1:k}) \propto \prod_{k=k_1}^{k_2} I\{\mathbf{A}_{k_1:k_2} = \mathbf{A}_{1:2}\}$ . Clearly an approximation is needed, but we show how to do much better than the MAP approach of the MHT. This motivates the next section where we derive a conjugate prior on the assignments  $A_{1:k}$ , which is useful for improving upon MAP; and we cast (2) as a special case of this distribution. 3

## 2 The Conjugate Assignment Prior

Given that we must compute the posterior  $P(\mathbf{A} | \mathbf{Z})$ , it is natural to ask what conjugate priors on  $\mathbf{A}$  are possible. Deriving approximate inference procedures is often greatly simplified if the prior on the parameters is conjugate to the complete data likelihood:  $p(\mathbf{Z}, \mathbf{X} | \mathbf{A})$  [2]. We follow the standard procedure for

deriving the conjugate prior for an exponential family (EF) complete likelihood:

$$\begin{aligned}
p(\mathbf{Z}, \mathbf{X}|\mathbf{A}) &= \prod_{j=1}^{NZ} p_0(z_j) \prod_{i=1}^{NT} p(z_j = x_i, A_{ij} = 1) A_{ij} \\
L_{ij} &:= \log p(z_j = x_i, A_{ij} = 1), \\
p(x_i) &= \prod_{i=1}^{NT} L_{i0} := 0, \\
p(x_i) &= \exp(\sum_{j=1}^{NZ} L_{ij} A_{ij}), \\
L_{0j} &:= \log p_0(z_j), \\
(7)
\end{aligned}$$

where we have introduced the matrix  $\mathbf{L} \in \mathbb{R}^{(NT+1) \times (NZ+1)}$  to represent log likelihood contributions from various assignments. Therefore, we have the following EF quantities [4, Ch. 2.4]: base measure  $\prod_{i=1}^{NT} p(x_i)$ , partition function  $g(\mathbf{A}) = 1$ , natural parameters  $\eta(\mathbf{A}) = \text{vec } \mathbf{A}$ , and sufficient statistics  $\mathbf{T}(\mathbf{Z}, \mathbf{X}) = \text{vec } \mathbf{L}$ . This implies the conjugate assignments prior (CAP) for  $\mathbf{P}(\mathbf{A}|\mathbf{Z}, \mathbf{X})$ :  $\mathbf{P}(\mathbf{A}|\mathbf{Z}, \mathbf{X}) := \mathbf{Z}(\eta(\mathbf{A})) \prod_{i=1}^{NT} \mathbb{I}\{\mathbf{A} \in \mathcal{A}\} \exp(\sum_{j=1}^{NZ} L_{0j} (\eta(\mathbf{A}))_j)$ ,  $\mathbf{Z}(\eta) := \exp(\sum_{j=1}^{NZ} \eta_j L_{0j})$ , (8)  $\mathbf{A} \in \mathcal{A}$

where  $\mathcal{A}$  is the set of all assignment matrices that obey the one-to-one constraints (1). Note that  $\eta$  is a function of the track meta-states  $\mathbf{S}$ . We recover the assignment prior of (2) in the form of the CAP distribution (8) via the following parameter settings, with  $\eta(\eta)$  denoting the logistic,

PD (i)  $\eta_{ij} = \log \frac{p_{ij}}{p_{i0}} = \log \frac{p_{ij}}{p_{i0}}$ ,  $i = 1:NT$ ,  $j = 1:NZ$ ,  $\eta_{0j} = \log \frac{p_{0j}}{p_{00}}$ . (9) Due to the symmetries in the prior of (9) we can analytically normalize (8) in this special case:  $\mathbf{Z}(\eta) = \mathbf{P}(\mathbf{A} | \mathbf{1}_{NT}, \mathbf{1}_{NZ} = 0) = \text{Poisson}(NZ \times \eta)$

$$\begin{aligned}
&\prod_{i=1}^{NT} \prod_{j=1}^{NZ} (1 + \eta_{ij}) \\
(10) \\
&\prod_{i=1}^{NT}
\end{aligned}$$

Given that the dummy row and columns of  $\eta$  are zero in (9), equation (10) is clearly the only way to get (8) to match (2) for the 0 assignment case. Although the conjugate prior (8) allows us to “compute” the posterior, “posterior = prior + L, computing  $E[\mathbf{A}]$  or  $\mathbf{Z}(\eta)$  remains difficult in general. This will cause problems in Section 3, but be ameliorated in Section 4 by a slight modification of the variational objective. One insight into the partition function  $\mathbf{Z}(\eta)$  is that if we slightly change the constraints in  $\mathcal{A}$  so that all the rows and columns must sum to one, i.e. we do not use a dummy row or column and  $\mathcal{A}$  becomes the set of permutation matrices, then  $\mathbf{Z}(\eta)$  is equal to the matrix permanent of  $\exp(\eta)$ , which is #P-complete to compute [24]. Although the matrix permanent

is #P-complete, accurate and computationally efficient approximations exist, some based on belief propagation [25; 17].

3

#### Variational Formulation

As explained in Section 1, exact inference on the full model in (6) is intractable, and as promised we show how to perform better inference than the existing solution of sliding window MAP. Our variational tracker enforces the factorization constraint that the posterior factorizes across assignment matrices and latent track states:  $p(A_{1:K}, X_{1:K}, S_{1:K} | Z_{1:K}) \approx q(A_{1:K}, X_{1:K}, S_{1:K}) = q(A_{1:K})q(X_{1:K}, S_{1:K})$ .

(11)

In some sense we can think of  $A$  as the “parameters” with  $X$  and  $S$  as the “latent variables” and use the common variational practice of factorizing these two groups of variables. This gives the variational lower bound  $L(q)$ :  $L(q) = \mathbb{E}_q [\log p(Z_{1:K}, X_{1:K}, A_{1:K}, S_{1:K})] + H[q(X_{1:K}, S_{1:K})] + H[q(A_{1:K})]$ , 1

In this section we drop the frame index  $k$  and implicitly condition on meta-states  $S_k$  for brevity.

4

(12)

where  $H[\cdot]$  denotes the Shannon entropy. From inspecting the VB lower bound (12) and (6) we arrive at the following induced factorizations without forcing further factorization upon (11):  $K \times Y$

$$\begin{aligned} q(A_{1:K}) &= \\ q(A_k) &, \\ q(X_{1:K}, S_{1:K}) &= \\ \prod_{i=1}^N \prod_{j=1}^Y & \\ q(x_{i,j})q(s_{i,j}) &. \end{aligned}$$

(13)

$i=1$

$k=1$

In other words, the approximate posterior on assignment matrices factorizes across time; and the approximate posterior on latent states factorizes across tracks. State Posterior Update Based on the induced factorizations in (13) we derive the updates for the track states  $x_{i,j}$  and meta-states  $s_{i,j}$  separately. Additionally, we derive the updates for each track separately. We begin with the variational updates for  $q(x_{i,j})$  using the standard VB update rules [4, c Ch. 10] and (6), denoting equality to an additive constant with  $\propto$ ,

$$\begin{aligned} \log q(x_{i,j}) &= \log p(x_{i,j}) + \\ K \sum_{k=1}^K \sum_{j=1}^Y & \mathbb{E}[A_{kij}] \log N(z_{j,k} - Hx_{i,k}, R) \end{aligned}$$

(14)

$k=1 \ j=1$

$$\propto q(x_{i,j}) \prod_{k=1}^K \prod_{j=1}^Y p(x_{i,j})$$

$K \sum_{j=1}^Y \sum_{k=1}^K$

$$N(z_{j,k} - Hx_{i,k}, R / \mathbb{E}[A_{kij}]) .$$

(15)

$k=1 \ j=1$

Using the standard product of Gaussians formula [6] this is proportional to  $q(x_i, ?) \propto p(x_i, ?)$

$$\prod_{k=1}^N \prod_{j=1}^N \frac{1}{\sqrt{|R/E[di,k]|}} \exp\left(-\frac{1}{2} \frac{(z_{j,k} - Hx_{i,k})^2}{R/E[di,k]}\right) \prod_{j=1}^N \frac{1}{\sqrt{|E[di,k]|}} \exp\left(-\frac{1}{2} \frac{(z_{j,k} - E[A_{kij}])^2}{E[di,k]}\right) \quad (16)$$

PNZ and recall that  $E[di,k] = 1 - E[A_{ki0}] = \sum_{j=1}^N E[A_{kij}]$ . The form of the posterior  $q(x_i, ?)$  is equivalent to a linear dynamical system with pseudo-measurements  $z$  covariance  $R/E[di,k]$ . Therefore,  $q(x_i, ?)$  is simply implemented using a Kalman smoother [22]. Meta-state Posterior Update We next consider the posterior on the track meta-states:

$$\log q(s_i, ?) = \log P(s_i, ?) + \sum_{k=1}^K \sum_{j=1}^N \log P(A_k - S_k) = \log P(s_i, ?) + \sum_{k=1}^K \sum_{j=1}^N \log q(s_i, ?) \propto P(s_i, ?) \prod_{k=1}^K \prod_{j=1}^N \exp(s_{i,k} \cdot i_{i,k}) \quad (17)$$

$$s_{i,k}(s) := E[di,k] \log(PD(s)) + (1 - E[di,k]) \log(1 - PD(s)) \quad (18)$$

$$\sum_{k=1}^K \sum_{j=1}^N \exp(s_{i,k} \cdot i_{i,k}) \quad (19)$$

where (18) follows from (2). If  $P(s_i, ?)$  follows a Markov chain then the form for  $q(s_i, ?)$  is the same as a hidden Markov model (HMM) with emission log likelihoods  $i_{i,k} \in [R]^{NS}$ . Therefore, the meta-state posterior  $q(s_i, ?)$  update is implemented using the forward-backward algorithm [21]. Like the MHT, our algorithm also works in an online fashion using a (much larger) sliding window. Assignment Matrix Update The reader can verify using (7)-(9) that the exact updates under the lower bound  $L(q)$  (12) yields a product of CAP distributions:  $q(A_{1:K}) =$

$$\prod_{k=1}^K \prod_{j=1}^N \text{CAP}(A_k - E_q(X_k) [L_k] + E_q(S_k) [i_{i,k}]) \quad (20)$$

This poses a challenging problem, as the state posterior updates of (16) and (19) require  $E_q(A_k) [A_k]$ ; since  $q(A_k)$  is a CAP distribution we know from Section 2 its expectation is difficult to compute.



### The Assignment Matrix Update Equations

In this section we modify the variational lower bound (12) to obtain a tractable algorithm. The resulting algorithm uses loopy belief propagation to compute  $\text{Eq}(\mathbf{A}_k) [\mathbf{A}_k]$  for use in (16) and (19). 5

We first note that the CAP distribution (8) is naturally represented as a factor graph:  $\text{NT NZ NT Y NZ Y Y Y S CAP}(\mathbf{A} - ?) \text{ ? fiR}(\mathbf{A}_i \text{ ?}) \text{ fjC}(\mathbf{A} \text{ ? } j) \text{ fij}(\mathbf{A}_{ij})$ ,  $i=1$

$$j=1$$

$$(21)$$

$$i=0 \ j=0$$

$\text{PNT PNZ } v_i = 1\}$  (C for column  $v_j = 1\}$  (R for row factors),  $\text{fjC}(v) := \prod_{i=0}^I \text{fiR}(v) := \prod_{j=0}^S \text{fij}(v) := \exp(\sum_{ij} v)$ . We use reparametrization methods (see [10]) to convert (21) to a pairwise factor graph, where derivation of the Bethe free energy is easier. The Bethe entropy is:  $\text{NZ X NT NT X NZ X X H}[q(c_j, \mathbf{A}_{ij})] \text{ H}[q(r_i, \mathbf{A}_{ij})] + \text{H}[q(\mathbf{A})] := \sum_{j=1} \sum_{i=0}$

$$i=1 \ j=0$$

$$?$$

$$\text{NT X}$$

$$\text{NZ H}[q(r_i)] ?$$

$$\text{NT X } i=1$$

$$\text{NT H}[q(c_j)] ?$$

$$j=1$$

$$i=1$$

$$=$$

$$\text{NZ X}$$

$$\text{H}[q(\mathbf{A}_i \text{ ?})] +$$

$$\text{NZ X}$$

$$\text{H}[q(\mathbf{A} \text{ ? } j)] ?$$

$$\text{NT X NZ X}$$

$$\text{H}[q(\mathbf{A}_{ij})]$$

$$(22)$$

$$i=1 \ j=1 \ \text{NZ NT X X}$$

$$\text{H}[q(\mathbf{A}_{ij})],$$

$$(23)$$

$$i=1 \ j=1$$

$$j=1$$

where the pairwise conversion used constrained auxiliary variables  $r_i := \mathbf{A}_i \text{ ?}$  and  $c_j := \mathbf{A} \text{ ? } j$ ; and used the implied relations  $\text{H}[q(r_i, \mathbf{A}_{ij})] = \text{H}[q(r_i)] + \text{H}[q(\mathbf{A}_{ij} - r_i)] = \text{H}[q(r_i)] = \text{H}[q(\mathbf{A}_i \text{ ?})]$ . We define an altered variational lower bound  $L^?(q)$ , which merely replaces the entropy  $\text{H}[q(\mathbf{A}_k)]$  c with  $\text{H}[q(\mathbf{A}_k)]$ .2 Note that  $L^?(q) = L(q)$  with respect to  $q(\mathbf{X}_{1:K}, \mathbf{S}_{1:K})$ , which implies the state posterior updates under the old bound  $L(q)$  in (16) and (19) remain unchanged with the new bound  $L^?(q)$ . To get the new update equations for  $q(\mathbf{A}_k)$  we examine  $L^?(q)$  in terms of  $q(\mathbf{A}_{1:K})$ :  $K \times c \text{ H}[q(\mathbf{A}_k)]$  (24)  $L^?(q) = \text{Eq}[\log p(\mathbf{Z}_{1:K} - \mathbf{X}_{1:K}, \mathbf{A}_{1:K})] + \text{Eq}[\log P(\mathbf{A}_{1:K} - \mathbf{S}_{1:K})] + \sum_{k=1}^c$

$$=$$



(a) Radar Example  
0.8  
S  
(b) SIAP Metrics  
C  
ARI  
NC?ARI  
0?1  
(c) Assignment Accuracy

Figure 2: Left: The output of the trackers on the radar example. We show the true trajectories (red  $\times$ ), 2D MHT (solid magenta), 3D MHT (solid green), and OMGP (cyan  $\times$ ). The state estimates for the VB tracker when active (black  $\times$ ) and dormant (black  $\circ$ ) are shown, where a  $\times$  90% threshold on the meta-state  $s$  is used to deem a track active for plotting. Center: SIAP metrics for  $N = 100$  realizations of the scenario on the left with 95% error bars. We show positional accuracy (i.e. RMSE) (PA, lower better), spurious tracks (S, lower better), and track completeness (C, higher better). The bars are in order: VB tracker (blue), 3D MHT (cyan), 2D MHT (yellow), and OMGP (red). The PA has been rescaled relative to OMGP so all metrics are in %. Right: Same as center but looking at assignment accuracy on ARI (higher better), no clutter (NC) ARI (higher better), and 0-1 loss (lower better) for classifying measurements as clutter.

5

#### Radar Tracking Example

We borrow the radar tracking example of the OMGP paper [18]. We have made the example more realistic by adding clutter  $\times = 8$  and missed detections  $PD = 0.5$ , which were omitted in [18]; and also used  $N = 100$  realizations to get confidence intervals on the results. We also compare with the 2D and 3D (i.e. multi-frame) MHT trackers as a baseline as they are the most widely used methods in practice. The OMGP requires the number of tracks  $N_T$  to be specified in advance, so we provided it with the true number of tracks, which should have given it an extra advantage. The trackers were evaluated using the SIAP metrics, which are the standard evaluation metrics in the field [7]. We also use the adjusted Rand index (ARI) [13] to compare the accuracy of the assignments made by the algorithms; the “no clutter” ARI (which ignores clutter) and the 0-1 loss for classifying measurements as clutter also serve as assignment metrics. In Figure 2(a) both OMGP and 2D MHT miss the real tracks and create spurious tracks from clutter measurements. The 3D MHT does better, but misses the western portion of track 3 and makes a swap between track 1 and 3 at their intersection. By contrast, the VB tracker gets the scenario almost perfect, except for a small bit of the southern portion of track 2. In that area, VB designates the track as dormant, acknowledging that the associated measurements are likely clutter. This replaces the notion of a “confirmed” track in the standard tracking literature with a model-based method, and demonstrates the advantages of using a principled and model-based paradigm for the track management problem. This is quantitatively shown over repeated trials in Fig-

ure 2(b) in terms of positional error; even more striking are illustrations of the near lack of spurious tracks in VB and much higher completeness than the competing methods. We also show that the assignments are much more accurate in Figure 2(c). To check the statistical significance of our results we used a paired t-test to compare the difference between VB and the second best method, the 3D MHT. Both the SIAP and assignment metrics all have  $p < 10^{-4}$ .

6

#### Real Data: Video Tracking in Sports

We use the VS-PETS 2003 soccer player data set as a real data example to validate our method. The data set is a 2500 frame video of players moving around a soccer field, with annotated ground truth; the variety of player interactions make it a challenging test case for multi-object tracking algorithms. To demonstrate the robustness of our tracker to correct a detector provided minimal training examples, we used multi-scale histogram of oriented gradients (HOG) features from 50 positive and 50 negative examples of soccer players to train a sliding window support vector machine (SVM) [23]. HOG features have been shown to work particularly well for pedestrian detection on the Caltech and INRIA data sets, and thus used for this example [8]. For each frame, the center of each bounding box is provided as the only input to our tracker. Despite modest detection rates from HOG-SVM, our tracker is still capable of separating clutter and dealing with missed detections.

7

#### Performance

0.8

0.6

0.4

0.2

0 ARI

(a) Soccer Tracking Problem

NC?ARI

0?1

(b) Soccer Assignment Metrics

Figure 3: Left: Example from soccer player tracking. We show the filtered state estimates of the MHT (magenta) and VB tracker (cyan) for the last 25 frames as well as the true positions (black). The green boxes show the detection of the HOG-SVM for the current frame. Right: Same as Figure 2(c) but for the soccer data. Methods in order: VB-DP (dark blue), VB (light blue), 3D MHT (green), 2D MHT (orange), and OMGP (red). Soccer data source: <http://www.cvg.rdg.ac.uk/slides/pets.html>.

We modeled player motion using (4) with  $F$  and  $Q$  derived from an NCV model [1, Ch. 1.5]. The parameters for the NCV,  $R$ ,  $PD$ ,  $\gamma$ , and the track meta-state parameters were trained by optimizing the variational lower bound  $L$  on the first 1000 frames, although the algorithm did not appear sensitive to these parameters. We additionally show an extension to the VB tracker with nonparametric clutter map learning; we learned the clutter map by passing the training measurements into a VB Dirichlet process (DP) mixture [5] with

their probability of being clutter under  $q(A)$  as weights. The resulting posterior predictive distribution served as  $p_0$  in the test phase; we refer to this method as the VB-DP tracker. We split the remainder of the data into 70 sequences of  $K = 20$  frames for a test set. Due to the nature of this example, we evaluate the batch accuracy of assigning boxes to the correct players. This demonstrates the utility of our algorithm for building a database of player images for later processing and other applications. In Figure 3(b) we show the ARI and related assignment metrics for VB-DP, VB, 2D MHT, 3D MHT, and OMGP. Note that the ARI only evaluates the accuracy of the MAP assignment estimate of VB; VB additionally provides uncertainty estimates on the assignments, unlike the MHT. VB manages to increase the no clutter ARI to  $0.95 \pm 0.01$  from  $0.86 \pm 0.01$  for 3D MHT; and decrease the 0-1 clutter loss to  $0.18 \pm 0.01$  from  $0.21 \pm 0.01$  for OMGP. Using the nonparametric clutter map lowered the 0-1 loss to  $0.016 \pm 0.005$  and increased the ARI to  $0.94 \pm 0.01$  (vs.  $0.76 \pm 0.01$  for the 2D and 3D MHT) as the VB-DP tracker knew certain areas, such as the post in the lower right, were more prone to clutter. As in the radar example the VB vs. MHT and VB vs. OMGP improvements are significant at  $p \leq 10^{-4}$ . The poor NC-ARI of OMGP is likely due to its lack of framing constraints, ignoring prior information on the assignments. Furthermore, in Figure 3(a) we plot filtered state estimates for the (non-DP) VB tracker; we again use the  $\approx 90\%$  meta-state threshold as a confirmed track. We see that the MHT is tricked by the various false detections from HOG-SVM and has spurious tracks across the field; the VB tracker introspectively knows when a track is unlikely to be real. While both the MHT and VB detect the referee in the upper right of the frame, the VB tracker quickly sets this track to dormant when he leaves the frame. The MHT temporarily extrapolates the track into the field before destroying it.

7

## Conclusions

The model-based manner of handling the track management problem shows clear advantages and may be the path forward for the field, which can clearly benefit from algorithms that eliminate arbitrary tuning parameters. Our method may be desirable even in tracking scenarios under which a full posterior does not confer advantages over a point estimate. We improve accuracy and reduce the exponential cost of the MAP approach to linear, which is a result of the induced factorizations of (13). We have also incorporated the often neglected framing constraints into our variational algorithm, which fits nicely with loopy belief propagation methods. Other areas, such as more sophisticated meta-state models, provide opportunities to extend this work into more applications of tracking and prove it as a general method and alternative to dominant approaches such as the MHT.

## 2 References

- [1] Bar-Shalom, Y., Willett, P., and Tian, X. (2011). Tracking and Data Fusion: A Handbook of Algorithms. YBS Publishing. [2] Beal, M. and Ghahramani,

Z. (2003). The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures. In *Bayesian Statistics*, volume 7, pages 453?464. [3] Benfold, B. and Reid, I. (2011). Stable multi-target tracking in real-time surveillance video. In *Computer Vision and Pattern Recognition (CVPR)*, 2011 IEEE Conference on, pages 3457?3464. IEEE. [4] Bishop, C. M. (2007). *Pattern Recognition and Machine Learning*. Springer. [5] Blei, D. M., Jordan, M. I., et al. (2006). Variational inference for Dirichlet process mixtures. *Bayesian analysis*, 1(1):121?143. [6] Bromiley, P. (2013). Products and convolutions of Gaussian probability density functions. Tina-Vision Memo 2003-003, University of Manchester. [7] Byrd, E. (2003). Single integrated air picture (SIAP) attributes version 2.0. Technical Report 2003-029, DTIC. [8] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition (CVPR)*, 2005 IEEE Conference on, pages 886?893. [9] Davey, S. J. (2003). Extensions to the probabilistic multi-hypothesis tracker for improved data association. PhD thesis, The University of Adelaide. [10] Eaton, F. and Ghahramani, Z. (2013). Model reductions for inference: Generality of pairwise, binary, and planar factor graphs. *Neural Computation*, 25(5):1213?1260. [11] Hartikainen, J. and S?arkk?a, S. (2010). Kalman filtering and smoothing solutions to temporal Gaussian process regression models. In *Machine Learning for Signal Processing (MLSP)*, pages 379?384. IEEE. [12] Heskes, T. (2003). Stable fixed points of loopy belief propagation are minima of the Bethe free energy. In *Advances in Neural Information Processing Systems 15*, pages 359?366. MIT Press. [13] Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of classification*, 2(1):193?218. [14] Jensfelt, P. and Kristensen, S. (2001). Active global localization for a mobile robot using multiple hypothesis tracking. *Robotics and Automation, IEEE Transactions on*, 17(5):748?760. [15] Jonker, R. and Volgenant, A. (1987). A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325?340. [16] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Transactions of the ASME ? Journal of Basic Engineering*, 82(Series D):35?45. [17] Lau, R. A. and Williams, J. L. (2011). Multidimensional assignment by dual decomposition. In *Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*, 2011 Seventh International Conference on, pages 437?442. IEEE. [18] L?azaro-Gredilla, M., Van Vaerenbergh, S., and Lawrence, N. D. (2012). Overlapping mixtures of Gaussian processes for the data association problem. *Pattern Recognition*, 45(4):1386?1395. [19] Mahler, R. (2003). Multitarget Bayes filtering via first-order multitarget moments. *Aerospace and Electronic Systems, IEEE Transactions on*, 39(4):1152?1178. [20] Poore, A. P., Rijavec, N., Barker, T. N., and Munger, M. L. (1993). Data association problems posed as multidimensional assignment problems: algorithm development. In *Optical Engineering and Photonics in Aerospace Sensing*, pages 172?182. International Society for Optics and Photonics. [21] Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257?286. [22] Rauch, H. E., Tung, F., and Striebel, C. T. (1965). Maximum likelihood estimates of linear dynamical systems. *AIAA*

Journal, 3(8):1445–1450. [23] Schölkopf, B. and Smola, A. J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. The MIT Press, Cambridge, MA, USA. [24] Valiant, L. G. (1979). The complexity of computing the permanent. Theoretical computer science, 8(2):189–201. [25] Watanabe, Y. and Chertkov, M. (2010). Belief propagation and loop calculus for the permanent of a non-negative matrix. Journal of Physics A: Mathematical and Theoretical, 43(24):242002. [26] Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2001). Bethe free energy, Kikuchi approximations, and belief propagation algorithms. In Advances in Neural Information Processing Systems 13.