Tracking without tracks: Multiple-target tracking using a Fourier-domain representation

Jeremy Maitin-Shepard

1 Notation

- Let $\mathbb{J}_n$ be the set of positive integers from 1 to n.
- Let $J_n = \{j_m | m \in \mathbb{J}_n\}$ be a set of n unique elements corresponding to the n targets/objects.
- Let $I_t = \mathbb{J}_t$ correspond to the set of time steps from 1 to t.
- $x_t^{(j)}$ is the true state of target j at time t.
- $z_t = (w_t, y_t)$ is the observation at time t.
- s_t, the data/measurement association value at time t, is equal to the index $j \in J_n$ of the single target that generated the measurement z_t, or equal to t if the measurement at time t is spurious.
- The shorthand $s_{a:b}$ denotes the set of variables $\{s_m | a \leq m \leq b\}$.
- w_t is the position component of the observation at time t.
- y_t is the discrete component of the observation, which may provide identity information.

2 Exact model

The observations and true states of each target are modeled using a factorial state-space model, as shown in Figure 2.

3 Permutation representation of $s_{1:t}$

In order to exploit an efficient representation of a distribution over permutations, the entire sequence of data associations for the measurements at times 1 through t are represented as a permutation σ_t of the set $J_n \cup I_t$.

Define $s_t^{(j)} = \{i \in I_t | s_i = j\}$ to be the set of time steps no later than t at which target j is observed.

For all $j \in J_n$, define

$$\sigma_t(j) = \begin{cases} \max(s_t^{(j)}) & \text{if } s_t^{(j)} \neq \emptyset; \\ j & \text{otherwise.} \end{cases}$$

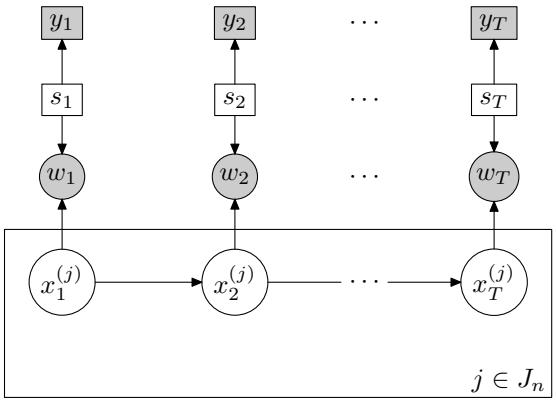

Figure 1: Graphical model description of the state-space model

Thus, $\sigma_t(j)$ is the most recent (as of time t) time step at which target j is observed, or is equal to j if target j was never observed.

For all $i \in I_t$, likewise define

$$\sigma_t(i) = \begin{cases} \max s_{i-1}^{(s_i)} & \text{if } s_{i-1}^{(s_i)} \neq \emptyset; \\ s_i & \text{otherwise.} \end{cases}$$

Thus, if z_i is spurious, i.e. $s_i = i$, then $s_{i-1}^{(s_i)} = \emptyset$, and $\sigma_t(i) = i$. Otherwise, $s_i \in J_n$, and thus $\sigma_t(i)$ is equal to the most recent time step before time step t at which the same target was observed as was observed at time step t, or is equal to the target s_i if s_i was never observed before time step t.

Due to this definition, if target j was observed at least k times as of time step t, $\sigma_t^k(j)$ is the kth most recent time step at which target j was observed. (Note that the superscript in $\sigma_t^k(j)$ denotes the standard exponentiation operator in the multiplicative group of permutations.) If target j was observed exactly $k-1$ times as of time step t, $\sigma_t^k(j) = j$. In general, if target j was observed exactly $m = |s_t^{(j)}|$ times as of time step t, $\sigma_t^k(j) = \sigma_t^{k \bmod m}(j)$.

Let $\sigma_t^{1:k}(j)$ denote the set $\{\sigma_t^m(j) | 1 \leq m \leq k\}$. Note that this set is equal to the subset of the k largest elements of $s_t^{(j)}$, or is equal to $s_t^{(j)} \cup \{j\}$ if $|s_t^{(j)}| < k$, and completely specifies the k most recent time steps at which target j was observed.

4 Exact independence

Note that due to the factorial structure of the model, given all of the measurements $z_{1:t}$ and all of the data associations, the conditional distributions of the states of each target are independent. As an equation, this is stated as

$$\Pr[x_t | z_{1:t}, s_{1:t}] = \prod_{j \in J_n} \Pr[x_t^{(j)} | z_{1:t}, s_{1:t}].$$

Furthermore, the conditional distribution of $x_t^{(j)}$ does not depend on the precise values of $s_{1:t}$, but rather only on the boolean values $s_{1:t} = j$, which is equivalently represented by $s_t^{(j)}$. As an equation, this is stated by

$$\Pr[x_t^{(j)} | z_{1:t}, s_{1:t}] = \Pr[x_t^{(j)} | z_{1:t}, s_t^{(j)}].$$

5 Assumed Density Model

Following the approach of assumed density filtering, it is assumed as an approximation that the joint distribution $\Pr(\sigma_t, x_t|z_{1:t})$ has a particular form, such that it can be represented compactly while still allowing efficient inference. As each measurement z_t is processed sequentially, this approximate representation is maintained and updated by performing one step of inference and then mapping the resultant distribution back to a distribution of the assumed form.

More specifically, the distributions $\Pr(\sigma_t|z_{1:t})$ and $\Pr(x_t|\sigma_t, z_{1:t})$ are represented separately. Representing $\Pr_{\sigma_t|z_{1:t}}(\sigma_t)$ exactly would require $O\left((n+t)!\right)$ space, and inference would have similarly intractable time complexity. Furthermore, the exact representation size grows with t.

5.1 Independence/decoupling assumption

It is assumed that given all of the measurements $z_{1:t}$ and, for each target j, the most recent time step at which it generated an observation, i.e. $\sigma_t(j)$, the conditional distributions of the states of each target are independent. As an equation, this assumption can be stated as

$$\Pr[x_t^{(j)}|z_{1:t}, s_t^{(j)}] \approx \Pr[x_t^{(j)}|z_{1:t}, \sigma_t(j)].$$

Note that $s_t^{(j)}$ is also equivalent to $\sigma_t^{1:t}(j)$. Intuitively, this assumption states that given all of the measurements, knowing about more than the most recent data association for a target provides no additional information about its current state.

5.2 Pruning assumption

Instead of representing the full joint distribution over $\sigma_t \in \mathbb{S}_{J_n \cup \mathbb{J}_t}$, portions of σ_t are effectively marginalized out, and only the marginal distribution over the resultant smaller permutation $\hat{\sigma}_t$ is represented. Specifically, $\sigma_t \in \mathbb{S}_{J_n \cup \mathbb{J}_t}$ is mapped to $\hat{\sigma}_t \in \mathbb{S}_{S_t}$, where $S_t \subset (S_{t-1} \cup \{t\}) \subset (\mathbb{J}_t \cup J_n)$, by the *pruning* operation π_{S_t} given by the recurrence

$$\hat{\sigma}_t(x) = (\pi_{S_t} \circ \sigma_t)(x) = \begin{cases} \sigma_t(x) & \text{if } \sigma_t(x) \in S_t; \\ (\pi_{S_t} \circ \sigma_t^2)(x) & \text{otherwise.} \end{cases}$$

Alternatively, viewing σ_t as the product of disjoint cycles, $\pi_{S_t}(\sigma_t)$ is simply equal to the same product of disjoint cycles except that the elements not in S_t are removed from the cycles.

5.3 Band-limited Fourier-domain representation

Instead of representing $\Pr_{\hat{\sigma}_t|z_{1:t}}(\hat{\sigma}_t|z_{1:t})$ exactly, it is represented approximately using a band-limited Fourier-domain representation[**?**] $\hat{\Pr}_{\hat{\sigma}_t|z_{1:t}}$, where $\hat{\Pr}_{\hat{\sigma}_t|z_{1:t}}$ is a function from a subset $\Lambda_{|S_t|}$ of the set of partitions of the number $|S_t|$. In the simplest case, $\Lambda_{|S_t|} = \{(|S_t|), (|S_t|-1, 1)\}$, which allows $\hat{\Pr}_{\hat{\sigma}_t|z_{1:t}}$ to represent information about first-order marginals of $\hat{\sigma}_t$. With this definition of Λ, $\hat{\Pr}_{\hat{\sigma}_t|z_{1:t}}$ has an $O(|S_t|^2)$ representation size, and typically $|S_t|$ is limited to be $O(n)$. Like the Fourier-domain representation of a function over $\mathbb{R}$, the Fourier-domain representation of a function of a group of permutations has the key property of linearity.

5.4 Compact representation of $\Pr(x_t^{(j)}|z_{1:t}, \sigma_t(j))$

Finally, it is assumed that $\Pr(x_t^{(j)}|z_{1:t}, \sigma_t(j)) \approx \Pr(x_t^{(j)}|z_{1:t}, \hat{\sigma}_t(j))$. Each of these $j \cdot |S_t|$ independent conditional distributions is then represented as a single Gaussian distribution, such that the total representation of $\Pr(x|\sigma_t, z_{1:t})$ is $O(j \cdot |S_t|)$ in size.

6 Analysis of assumptions

6.1 Exact measurement case

Consider the case that measurements are exact and completely reveal the state of the target, i.e. $x_t^{(j)}$ is deterministic given z_t and given that $s_t = j$. This case is obviously not one of great interest, but it does illustrate the extreme case of measurements being precise. This case might apply to measurements received as precise GPS coordinates or by a high resolution overhead camera with an accurate person detector. In this case,

$$\Pr[x_t^{(j)}|z_{1:t}, s_t^{(j)}] = \Pr[x_t^{(j)}|z_i, \sigma_t(j) = i],$$

and thus the assumption clearly holds.

6.1.1 Case where measurements are not all of the same type, and provide different partial information about the state

(TODO)

6.1.2 Case where measurements have non-uniform variance

(TODO)

7 Approximate inference

The procedure for processing a single measurement z_{t+1} and updating the approximate representation of $\Pr(x_t, \sigma_t|z_{1:t})$ to the approximate representation of $\Pr(x_{t+1}, \sigma_{t+1}|z_{1:t+1})$ is broken up into several steps.

7.1 Prediction step for $\Pr(x_t|z_{1:t}, \sigma_t)$

For all j and i, since $\Pr(x_{t+1}^{(j)}|z_{1:t}, \sigma_t(j) = i)$ is assumed to be Gaussian, $\Pr(x_{t+1}^{(j)}|z_{1:t}, \sigma_t(j) = i)$ is also Gaussian and can be computed efficiently using the Kalman filter prediction step that takes into account the state transition model for $x_t^{(j)}$.

7.2 Conditioning $x^{(j)}$

Suppose $s_t = j \neq t$ (i.e. $\sigma_t(j) = t$).

$$
\begin{aligned}
&\Pr(x_t^{(j)}|z_{1:t}, \sigma_t(j) = t) \\
&\propto \Pr(x_t^{(j)}, z_{1:t}, \sigma_t(j) = t) \\
&= \sum_{i=1}^{t-1} \Pr(x_t^{(j)}, z_{1:t}, \sigma_t(j) = t, \sigma_{t-1}(j) = i) \\
&= \sum_{i=1}^{t-1} \Pr(z_t|x_t^{(j)}, z_{1:t-1}, \sigma_t(j) = t, \sigma_{t-1}(j) = i) \Pr(x_t^{(j)}, z_{1:t-1}, \sigma_t(j) = t, \sigma_{t-1}(j) = i) \\
&= \sum_{i=1}^{t-1} \Pr(z_t|x_t^{(j)}, \sigma_t(j) = t) \Pr(x_t^{(j)}, z_{1:t-1}, \sigma_t(j) = t, \sigma_{t-1}(j) = i) \\
&= \sum_{i=1}^{t-1} \Pr(z_t|x_t^{(j)}, \sigma_t(j) = t) \Pr(\sigma_t(j) = t|x_t^{(j)}, z_{1:t-1}, \sigma_{t-1}(j) = i) \Pr(x_t^{(j)}, z_{1:t-1}, \sigma_{t-1}(j) = i) \\
&= \sum_{i=1}^{t-1} \Pr(z_t|x_t^{(j)}, \sigma_t(j) = t) \Pr(\sigma_t(j) = t) \Pr(x_t^{(j)}, z_{1:t-1}, \sigma_{t-1}(j) = i) \\
&\propto \sum_{i=1}^{t-1} \Pr(z_t|x_t^{(j)}, \sigma_t(j) = t) \Pr(x_t^{(j)}|z_{1:t-1}, \sigma_{t-1}(j) = i) \Pr(\sigma_{t-1}(j) = i|z_{1:t-1}).
\end{aligned}
$$

Thus, in this case conditioning reduces to repeated application of the Kalman filter conditioning rule, and depends only on first-order marginal probabilities of σ_{t-1}.

Suppose $s_t \neq j$ (i.e. $\sigma_t(j) = \sigma_{t-1}(j)$). Then $x_t^{(j)}$ is independent of z_t given $z_{1:t-1}$ and σ_t, and thus, $\Pr(x_t^{(j)}|z_{1:t}, \sigma_t(j)) = \Pr(x_t^{(j)}|z_{1:t-1}, \sigma_{t-1}(j))$; that is, the conditioning step in this case is trivial.

7.3 Prediction and conditioning for σ_t

This inference step depends heavily on application of the *branching rule*, which specifies a procedure for efficiently transforming a Fourier-domain representation of a function over $\mathbb{S}_m$ to a Fourier-domain representation of a function over either $\mathbb{S}_{m+1}$ or $\mathbb{S}_{m-1}$. The two specific instances of this rule that are used allow the following transformations:

1. simply restricting a function over $\mathbb{S}_m$ to $\mathbb{S}_{m-1}$ (where an element of $\mathbb{S}_{m-1}$ is viewed as an element of $\mathbb{S}_m$ that fixes m) in the Fourier domain;

2. extending the domain of a function over $\mathbb{S}_{m-1}$ to $\mathbb{S}_m$ according to the rule that elements of $\mathbb{S}_m$ that do not fix m map to the value 0.

Both of these transformations can be done in time linear in the size of the representation. Additionally, transforming in the Fourier-domain a function $f(\sigma)$ to the function $f(\sigma \circ \pi)$ or to the function $f(\pi \circ \sigma)$ for any permutation π simply involves a matrix multiplication, and for a first-order representation is $O(m^{3/2})$, where m is the size of the representation. This operation is $O(m)$ if π is simply an adjacent swap, due to sparsity of the resultant matrix.

$$\begin{aligned}\Pr(\sigma_{t+1}|z_{1:t+1}) &= \sum_j \Pr(\sigma_{t+1}, s_{t+1} = j|z_{1:t+1}) \\ &= \sum_j \Pr(\sigma_{t+1}, s_{t+1} = j, z_{t+1}|z_{1:t}) \cdot c \\ &\propto \sum_j \Pr(\sigma_{t+1}, s_{t+1} = j, z_{t+1}|z_{1:t})\end{aligned}$$

where $c = \Pr(z_{1:t})/P(z_{1:t+1})$.

If $j = t + 1$:

$$\Pr(\sigma_{t+1}, s_t = j, z_{t+1}|z_{1:t}) = \Pr(z_{t+1}|\sigma_{t+1}, s_t = j, z_{1:t}) \cdot \Pr(\sigma_{t+1}|s_{t+1} = t + 1, z_{1:t}) \cdot \Pr(s_{t+1} = t + 1).$$

Otherwise, $j \in J_n$:

$$\begin{aligned}\Pr(\sigma_{t+1}, s_t = j, z_{t+1}|z_{1:t}) &= \sum_i \Pr(\sigma_{t+1}, s_t = j, z_{t+1}, \sigma_{t+1}(t + 1) = i|z_{1:t}) \\ &= \sum_i \Pr(z_{t+1}|\sigma_{t+1}, s_t = j, \sigma_{t+1}(t + 1) = i, z_{1:t})\times \\ &\qquad \Pr(\sigma_{t+1}, \sigma_{t+1}(t + 1) = i|s_t = j, z_{1:t}) \cdot \Pr(s_t = j).\end{aligned}$$

Assuming a first-order representation, the prediction and condition of σ has time complexity $O(n^5)$.

7.4 Pruning

The prediction and conditioning step results in a permutation σ_{t+1} one element larger than σ_t. To avoid this increase in size, some element $i \in S_t \cup \{t + 1\} \setminus J_n$ is marginalized out of σ_{t+1}. The particular element i is chosen out of the $|S_t| + 1 - n$ possible choices to minimize

$$\mathrm{E}_{\sigma_{t+1}|z_{1:t+1}}[\mathrm{KL}(\Pr(x_{t+1}|z_{1:t+1}, \sigma_{t+1})||\Pr(x_{t+1}|z_{1:t+1}, \hat{\sigma}_{t+1}))],$$

where $\hat{\sigma}_{t+1} = \pi_{S_t \cup \{t+1\} \setminus \{i\}}(\sigma_{t+1})$.

$$\begin{aligned}&\mathrm{E}_{\sigma_{t+1}|z_{1:t+1}}[\mathrm{KL}(\Pr(x_{t+1}|z_{1:t+1}, \sigma_{t+1})||\Pr(x_{t+1}|z_{1:t+1}, \hat{\sigma}_{t+1}))] \\ &= \sum_{\sigma_{t+1}} \Pr(\sigma_{t+1}|z_{1:t+1}) \sum_j \mathrm{KL}(\Pr(x^{(j)}_{t+1}|z_{1:t+1}, \sigma_{t+1}(j))||\Pr(x^{(j)}_{t+1}|z_{1:t+1}, \hat{\sigma}_{t+1}(j))) \\ &= \sum_j \sum_{i'} \sum_{i''} \Pr(\sigma_{t+1}(j) = i', \hat{\sigma}_{t+1}(j) = i''|z_{1:t+1})\times \\ &\qquad \mathrm{KL}(\Pr(x^{(j)}_{t+1}|z_{1:t+1}, \sigma_{t+1}(j) = i')||\Pr(x^{(j)}_{t+1}|z_{1:t+1}, \hat{\sigma}_{t+1}(j) = i'')).\end{aligned}$$

8 Remaining Work

- Now that an efficient procedure for performing inference using this Fourier-domain representation has been developed, it will be implemented and tested both on simulated data as well as on actual data, likely camera data.
- This algorithm will be compared to alternative approaches such as Rao-Blackwellized particle filter method, where the sampling is over sequences of data associations.