2-Step Temporal Bayesian Networks (2TBN):
Filtering, Smoothing, and Beyond
Technical Report: TRCIM1030

Anqi Xu
anqixu(at)cim(dot)mcgill(dot)ca

School of Computer Science,
McGill University,
Montreal, Canada, H3A 0E9

Revision 1.1
Contents

1 Introduction 2
  1.1 Background .................................................. 3

2 Derivations of 2TBN Inference and Prediction Tasks 3
  2.1 Sequential Filtering for 2TBN ............................. 3
  2.2 Sequential Smoothing for 2TBN ............................ 5
  2.3 Prediction of Observed Evidence for 2TBN ............. 7

Revision History

<table>
<thead>
<tr>
<th>Revision</th>
<th>Date</th>
<th>Author(s)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>16/06/10</td>
<td>Anqi Xu</td>
<td>cleaned up formatting</td>
</tr>
<tr>
<td>1.0</td>
<td>14/12/09</td>
<td>Anqi Xu</td>
<td>derivations for 2TBN filtering, smoothing, and prediction</td>
</tr>
</tbody>
</table>
1 Introduction

Dynamic Bayesian Networks (DBN) are often used to model beliefs about a sequence of states for a time-varying latent random variable, in relationship to other observed factors. Common inference tasks on these networks include filtering, which aims to track the current state of the latent variable given past observations, and smoothing, where a batch of temporal observations is used to determine the belief over the latent variable at any time during the batch window.

This manuscript primarily supplements [1] by expanding on the derivations for sequential Bayesian filtering and smoothing formulations for dynamic Bayesian networks. It also expands the family of applicable networks for these sequential inference tasks to a broad class of 2-Step Temporal Bayesian Networks (2TBN) that are shaped as shown in Fig. 1(b). Furthermore, derivation for the prediction of observed evidence in 2TBN is provided, which is useful for evaluating the accuracy of these Bayesian models. Finally, algorithmic and pictorial examples for inference, prediction, and model training on 2TBN, using a histogram-based implementation, are provided for a toy robotics context [2]. The reader is encouraged to refer to [3, 4, 2] for comprehensive treatments on Dynamic Bayesian Networks.

![Figure 1: Model structures for Hidden Markov Models (HMM) and 2-Step Temporal Bayesian Networks (2TBN).](image)

(a) HMM - Unrolled Form  
(b) 2TBN - Unrolled Form  
(c) HMM - Recursive Form  
(d) 2TBN - Recursive Form
1.1 Background

Derivations for performing sequential filtering and smoothing on Hidden Markov Models (HMM) (as illustrated in Fig. 1(a)) are well established in the literature of temporal Bayesian networks [1]. This report reproduces these derivations in a verbose manner, and also generalizes them for the class of 2-step Temporal Bayesian Networks (2TBN) with structures that are shaped like Fig. 1(b).

This 2TBN structure is often seen in robotics, where at each time step the belief over the latent robot state, $X$, is estimated by propagating the belief on the previous robot state, $X'$, with the latest control inputs $U$. Subsequently, sensor data $Z$ are used to highlight regions in the belief space that are consistent with these observations while downvoting other less likely hypotheses.

The derivations in this report will make repeated use of Bayes’ rule, variable marginalization, and simplifications due to variable independence (including those resulting from the Markov assumption for Bayesian Networks [4]). Let $\text{Bayes}(A,B)$ denote the act of substitution between $p(A,B|C)$ and $p(A|B,C)$, following Bayes’ rule:

$$p(A,B|C) = p(A|B,C)p(B|C)$$  (1)

In addition, $\text{Marg}(A)$ refers to the marginalization of variable(s) $A$ into an arbitrary probability distribution $p(B|C)$:

$$p(B|C) = \int p(A,B|C) \, dA$$  (2)

Furthermore, $\text{Ind}(A \perp B|C)$ signifies the simplification of a conditional distribution $p(A|B,C)$, by eliminating conditionally independent variable(s) $B$ as follows:

$$p(A|B,C) = p(A|C) \quad \text{assuming } A \perp B|C$$  (3)

The above properties extend to non-conditional forms, i.e. cases where $C = \emptyset$. In addition, these properties are sufficiently general, where the letters $A, B, C$ may represent one or multiple variables.

2 Derivations of 2TBN Inference and Prediction Tasks

This section supplements [1] by presenting verbose derivations for sequential filtering, smoothing, and prediction tasks for 2-Step Temporal Bayesian Networks (2TBN) that are shaped like Fig. 1(b).

2.1 Sequential Filtering for 2TBN

Sequential Bayesian filtering (also known as Bayesian tracking) refers to the task of performing repeated inference on a dynamic Bayesian Network, such as the one in...
In this report, the filtering process aims to compute a probability distribution for the latest latent variable $X_n$, given the history of all past observations, at times $i = 1 : n$, pertaining to control inputs $\{U_i\}$ and observed evidence $\{Z_i\}$. Derivations for sequential filtering for 2TBN are prominent in the literature [2], and are included here for completeness.

In other words, at each moment of time $n$, the filtering process aims to compute a probability distribution for the latest latent variable $X_n$, given the history of all past observations, at times $i = 1 : n$, pertaining to control inputs $\{U_i\}$ and observed evidence $\{Z_i\}$. Derivations for sequential filtering for 2TBN are prominent in the literature [2], and are included here for completeness.

When following the derivations above, it can be useful to refer to the graphical form of the partially unrolled structure of a 2TBN, as shown in Fig. 2. In particular, one can pictorially verify the independence properties, especially those based on the Markov independence assumption.

From Eqn. 7, it follows that the desired filtering belief, $filter(X_n) := p(X_n|U_{1:n}, Z_{1:n})$, can be expressed in a recursive manner:

$$filter(X_n) = \int p(Z_n|X_n, X_{n-1}) p(X_n|X_{n-1}, U_n, U_{n-1}) \frac{filter(X_{n-1})}{p(Z_n|U_{1:n}, Z_{1:n-1})} dX_{n-1}$$

The above can be simplified by renaming $p(Z_n|X_n, X_{n-1})$ and $p(X_n|X_{n-1}, U_n, U_{n-1})$, as their corresponding time-invariant observed evidence and propagation conditional distributions, $observe(Z|X, X')$ and $propagate(X'|U, U')$. It is also interesting to note that the denominator term, $p(Z_n|U_{1:n}, Z_{1:n-1})$, corresponds to the marginal of the numerator terms in Eqn. 7. Consequently, the final form for sequential Bayesian filtering

$$p(X_n|U_{1:n}, Z_{1:n}) \triangleq \frac{\int p(Z_n|X_n, X_{n-1}) p(X_n|X_{n-1}, U_n, U_{n-1}) \frac{filter(X_{n-1})}{p(Z_n|U_{1:n}, Z_{1:n-1})} dX_{n-1}}{p(Z_n|U_{1:n}, Z_{1:n-1})}$$

$$p(Z_n|U_{1:n}, Z_{1:n-1})$$

1Technically speaking, all inference problems in this report are further conditioned by a known prior distribution over the latent variable, $X_0$. Popular instantiations include using an uniform prior, using a hand-engineered prior, or using the final posterior at the end of a previous inference sequence. Nevertheless, the conditioning on the prior distribution for $X_0$ is typically omitted in the literature [4], and is thus also omitted in this report.
can be expressed elegantly as it relates to the joint belief over consecutive latent states, \( \text{belief}(X_n, X_{n-1}) \), as follows:

\[
\text{belief}(X_n, X_{n-1}) := \text{observe}(Z|X, X') \cdot \text{propagate}(X|X', U, U') \cdot \text{filter}(X_{n-1}) \tag{6}
\]

\[
\text{filter}(X_n) = \frac{\int \text{belief}(X_n, X_{n-1}) \, dX_{n-1}}{\int\int \text{belief}(X_n, X_{n-1}) \, dX_{n-1} \, dX_n} \tag{7}
\]

Finally, the related inference task of MAP inference [4] can be addressed for 2TBN, by converting the sum-product form (or more correctly, the integral-product form) in Eqns. 6, 7 into the max-product or max-log-sum forms.

### 2.2 Sequential Smoothing for 2TBN

Assuming a finite time window up to time \( T \), sequential Bayesian smoothing refers to the task of inferring the beliefs \( p(X_n|U_1:T, Z_1:T) \), i.e., computing the probability distributions of the latent variable \( X \) at times \( n \in 1:T \), given the entire history of past, present, and future observations at times \( i = 1:T \) pertaining to control variables \( \{U_i\} \) and evidence variables \( \{Z_i\} \). The key distinction between filtering and smoothing is that the smoothed beliefs of the latent state at time \( n \) will have considered effects from observable control and evidence states in the future (i.e. for time steps \( j \in (n+1):T \)), in addition to those from the past and present.

It is well established that exact inference of an arbitrary distribution within a Bayesian Network can be computed efficiently, by relating it to the joint distribution of the network. In particular, using a Clique Tree representation or following non-loopy belief propagation [4], it has been shown that the joint distribution can be obtained in two linear passes: first by propagating the influence of factors downwards through the tree from root to leaves, and then propagating influence upwards back to the root. Therefore, it is not surprising to expect that smoothing for 2TBN can be similarly computed.
in an efficient manner, by first propagating forwards through the unrolled network, and
then propagating backwards.

The following derivation is based on the forward-filtering backward-smoothing
method [1] for computing the marginal distributions \( p(X_n|U_{1:T}, Z_{1:T}) \), \( n \in 1:T \). Once
again, it can be helpful to refer to the graphical structure of the network, shown in
Fig. 2, when following reasonings about the (conditional) independence of various
variables.

\[
p(X_n|U_{1:T}, Z_{1:T}) \overset{\text{Marg}(X_{n+1})}{=} \int p(X_n, X_{n+1}|U_{1:T}, Z_{1:T}) \, dX_{n+1}
\]

\[
\overset{\text{Bayes}(X_n, X_{n+1})}{=} \int p(X_n|X_{n+1}, U_{1:T}, Z_{1:T}) \, p(X_{n+1}|U_{1:T}, Z_{1:T}) \, dX_{n+1}
\]

\[
\overset{\text{Ind}(X_n, U_{1:T}, Z_{1:T})}{=} \int p(X_n|X_{n+1}, U_{1:T}, Z_{1:T}) \, p(X_{n+1}|U_{1:T}, Z_{1:T}) \, dX_{n+1}
\]

(8)

The first distribution inside the integral, \( p(X_n|X_{n+1}, U_{1:T+1}, Z_{1:T}) \), can be simplified
as follows:

\[
p(X_n|X_{n+1}, U_{1:T+1}, Z_{1:T}) \overset{\text{Bayes}(X_n, X_{n+1}, Z_{n+1})}{=} \frac{p(X_n, X_{n+1}, Z_{n+1}|U_{1:T+1}, Z_{1:T})}{p(X_{n+1}, Z_{n+1}|U_{1:T+1}, Z_{1:T})}
\]

\[
\overset{\text{Bayes}(Z_{n+1}, X_{n+1})}{=} \frac{p(Z_{n+1}|X_n, X_{n+1}, U_{1:T+1}, Z_{1:T}) \, p(X_{n+1}|U_{1:T+1}, Z_{1:T})}{p(X_{n+1}, Z_{n+1}|U_{1:T+1}, Z_{1:T})}
\]

\[
\overset{\text{Ind}(X_{n+1}, U_{1:T}, Z_{1:T})}{=} \frac{p(Z_{n+1}|X_n, X_{n+1}) \, p(X_{n+1}|U_{1:T+1}, Z_{1:T})}{p(X_{n+1}, Z_{n+1}|U_{1:T+1}, Z_{1:T})}
\]

\[
\overset{\text{Ind}(X_{n+1}, U_{n})}{=} \frac{p(Z_{n+1}|X_n, X_{n+1}) \, p(X_{n+1}|U_{1:T+1}, Z_{1:T})}{p(X_{n+1}, Z_{n+1}|U_{1:T+1}, Z_{1:T})}
\]

(9)

Substituting the result of Eqn. 9 back into Eqn. 8:

\[
p(X_n|U_{1:T}, Z_{1:T}) = \int p(X_n|X_{n+1}, U_{1:T+1}, Z_{1:T+1}) \, p(X_{n+1}|U_{1:T}, Z_{1:T}) \, dX_{n+1}
\]

\[
= \int \frac{p(Z_{n+1}|X_n, X_{n+1}) \, p(X_{n+1}|X_{n+1}, U_{n+1}, Z_{n+1}) \, p(X_{n+1}|U_{1:T}, Z_{1:T})}{p(X_{n+1}, Z_{n+1}|U_{1:T+1}, Z_{1:T})} \, dX_{n+1}
\]

(10)

Similar to the derivations for Bayesian filtering, the expressions \( p(Z_{n+1}|X_n, X_{n+1}) \)
and \( p(X_{n+1}|X_{n+1}, U_{n+1}, U_n) \) can be interpreted as the observed evidence and propaga-

tion likelihoods, observe \( (Z|X, X') \) and propagate \( (X'|X, U, U') \). One may also spot the

presence the joint belief over consecutive latent states belief \( (X_n, X_{n-1}) \). Furthermore,
by once again realizing that the denominator term, \( p(X_{n+1}, Z_{n+1}|U_{1:n+1}, Z_{1:n}) \), is the marginal of its numerator, the smoothed belief \( \text{smooth}(X_n):= p(X_n|U_{1:T}, Z_{1:T}) \) can be expressed in an elegant recursive form:

\[
\text{smooth}(X_n) = \int \int \text{belief}(X_n, X_{n-1}) \cdot \text{smooth}(X_{n+1}) dX_n \cdot \text{smooth}(X_{n+1}) dX_{n+1}
\]  

(11)

2.3 Prediction of Observed Evidence for 2TBN

The previous sections in this report discussed one of the primary purposes of a 2TBN model, namely to infer the filtered or smoothed beliefs for latent variable(s). When evaluating these models, it is often difficult to obtain ground truth for these variable(s), and thus evaluation must be carried out implicitly, by comparing each observed states of a specific evidence variable against a predictive belief computed from the model.

A common approach for evaluating 2TBN models by predicting observed evidences is through offline or batch processing of a dataset:

1. omit all instances of a target evidence variable, \( \{O_n|O_n \neq \emptyset, n \in 1:N\} \) \(^2\)
2. for each omitted instance, predict \( p(O_n|U_{1:n}, Z_{1:n}) \), i.e. the belief over a (presumed) unobserved evidence state \( O_n \) given all control inputs \( U_{1:n} \) and all other evidence factors \( Z_{1:n} \)
3. compare predictions, either directly or via a statistic, against omitted instances of the target evidence variable

The derivation for the predictive beliefs over \( O_n \) follows a nearly identical steps to the derivation for sequential filtering from Sec. 2.1, since \( p(O_n|U_{1:n}, Z_{1:n}) \) and \( p(X_n|U_{1:n}, Z_{1:n}) \) are both conditioned by the same set of variables:

\(^2\)Both \( O_n \) and \( Z_n \) represent different observed evidence variables, although the renaming of the variable \( O_n \) is to emphasize the omission of a specific target evidence variable.
By noting the similarity to the final form of Eqn. 7, a concise expression for the predictive belief of $O_n$ can be obtained by relating to $\text{belief}(X_n, X_{n-1})$ and $\text{filter}(X_{n-1})$:

$$
p(O_n|U_{1:n}, Z_{1:n}) = \left[\int \int p(O_n|X_n, X_{n-1}) \cdot \text{belief}(X_n, X_{n-1}) \, dX_n \, dX_{n-1}\right] / \left[\int \int \text{belief}(X_n, X_{n-1}) \, dX_n \, dX_{n-1}\right]
$$

(13)
References


