Temporal Parallelization of Bayesian Smoothers

Date: August 4, 2023
Presenter: Kei Ishikawa

TL;DR

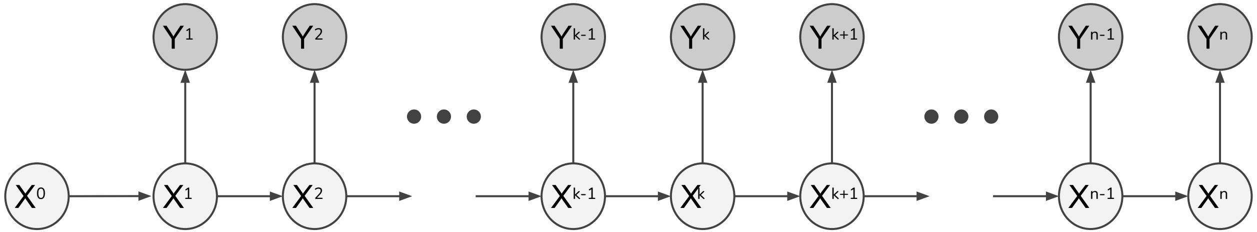
- Bayesian filtering & smoothing algorithms (e.g. Kalman filtering, Viterbi for HMMs) for a sequence of length n can be O(log n) time!
    - Parallel-scan for prefix sum is used.
    - Note that the total complexity is O(n).

Model

A. Bayesian filtering and smoothing

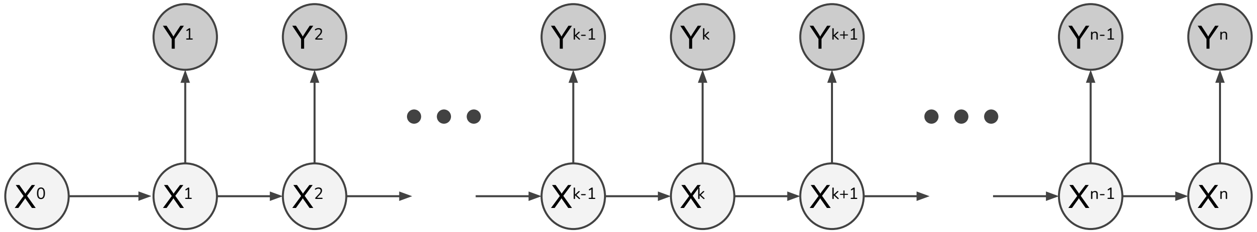
Bayesian filtering and smoothing methods [6] are algorithms for statistical inference in probabilistic state-space models of the form

$$\begin{aligned} x_k &\sim p(x_k \mid x_{k-1}), \\ y_k &\sim p(y_k \mid x_k). \end{aligned} \tag{1}$$



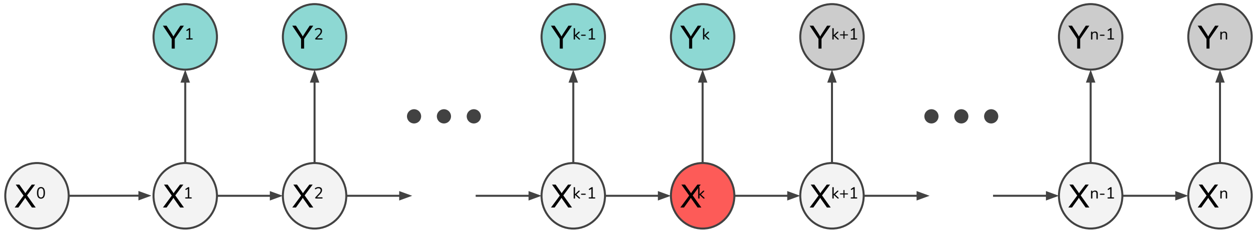
Bayesian Filtering & Smoothing

- Filtering:
  - Given observations up to now (y_1, ..., y_k), calculate the posterior of current state x_k
  - i.e. calculate p(x_k| y_1, ..., y_k)
- Smoothing:
  - Given observations of whole sequence (y_1, ..., y_n), calculate the posterior of any state x_k
  - i.e. calculate p(x_k| y_1, ..., y_n)



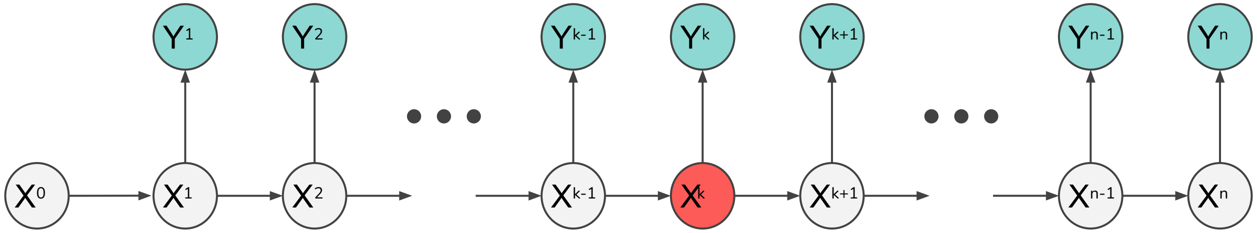
Bayesian Filtering & Smoothing

- Filtering:
    - Given observations up to now ($y_1, \dots, y_k$), calculate the posterior of current state x_k
    - i.e. calculate $p(x_k | y_1, \dots, y_k)$
- Smoothing:
    - Given observations of whole sequence ($y_1, \dots, y_n$), calculate the posterior of any state x_k
    - i.e. calculate $p(x_k | y_1, \dots, y_n)$



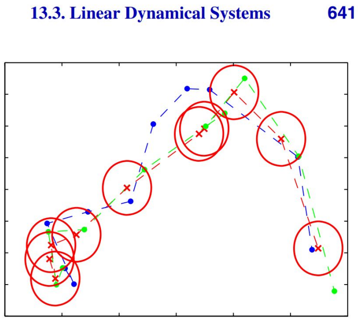
Bayesian Filtering & Smoothing

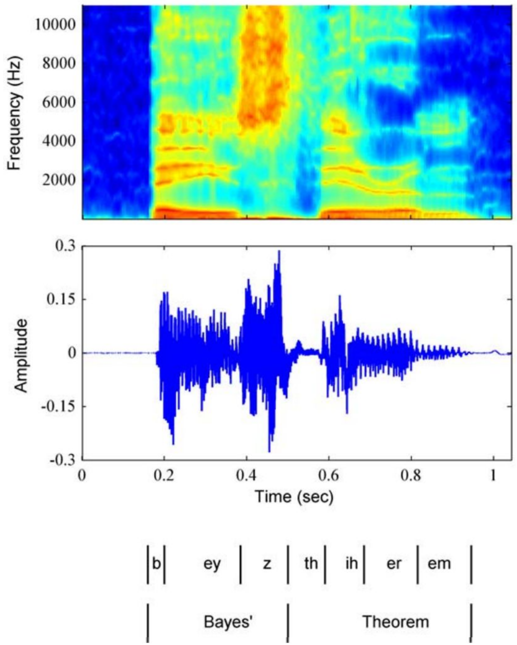
- Filtering:
  - Given observations up to now (y_1, …, y_k), calculate the posterior of current state x_k
  - i.e. calculate p(x_k| y_1, …, y_k)
- Smoothing:
  - Given observations of whole sequence (y_1, …, y_n), calculate the posterior of any state x_k
  - i.e. calculate p(x_k| y_1, …, y_n)



Examples of (Probabilistic) State Space Models

- Hidden Markov models (HMMs)
  - Discrete state x_k
  - Arbitrary emission model p(y_k| x_k)
  - Applications include: speech recognition, handwriting recognition
- Linear Gaussian state space models
  - $x_k \sim \mathcal{N}(x_k | A x_{k-1}, \Gamma)$
  - $y_k \sim \mathcal{N}(y_k | C x_k, \Sigma)$
  - Applied to Kalman filtering & smoothing for object tracking





The Classic Bayesian Filtering & Smoothing: the Forward-Backward Algorithm

WIKIPEDIA
The Free Encyclopedia

Create account Log in

Forward–backward algorithm

6 languages

Article Talk Read Edit View history Tools

From Wikipedia, the free encyclopedia

This article includes a list of references, related reading, or external links, **but its sources remain unclear because it lacks inline citations**. Please help to improve this article by introducing more precise citations. *(April 2018) (Learn how and when to remove this template message)*

The **forward–backward algorithm** is an inference algorithm for hidden Markov models which computes the posterior marginals of all hidden state variables given a sequence of observations/emissions $o_{1:T} := o_1, \dots, o_T$, i.e. it computes, for all hidden state variables $X_t \in \{X_1, \dots, X_T\}$, the distribution $P(X_t \mid o_{1:T})$. This inference task is usually called **smoothing**. The algorithm makes use of the principle of dynamic programming to efficiently compute the values that are required to obtain the posterior marginal distributions in two passes. The first pass goes forward in time while the second goes backward in time; hence the name *forward–backward algorithm*.

The Classic Bayesian Filtering & Smoothing: the Forward-Backward Algorithm

The Bayesian filter is a sequential algorithm, which iterates the following prediction and update steps:

$$p(x_k \mid y_{1:k-1}) = \int p(x_k \mid x_{k-1})\, p(x_{k-1} \mid y_{1:k-1})\, \mathrm{d}x_{k-1}, \quad (2)$$

$$p(x_k \mid y_{1:k}) = \frac{p(y_k \mid x_k)\, p(x_k \mid y_{1:k-1})}{\int p(y_k \mid x_k)\, p(x_k \mid y_{1:k-1})\, \mathrm{d}x_k}. \quad (3)$$

Given the filtering outputs for $k = 1, \ldots, n$, the Bayesian forward-backward smoother consists of the following backward iteration for $k = n - 1, \ldots, 1$:

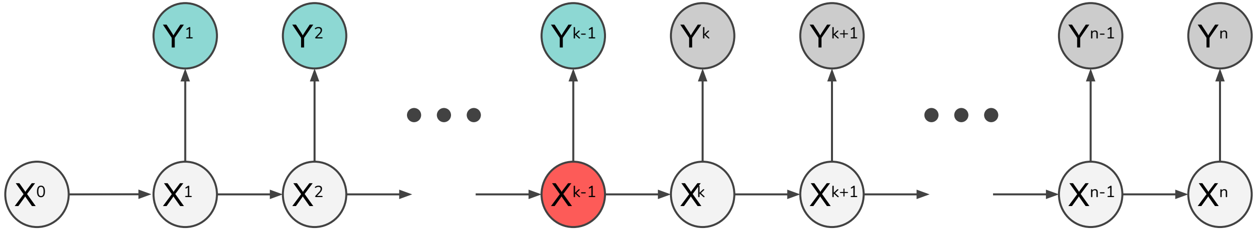
$$p(x_k \mid y_{1:n}) = p(x_k \mid y_{1:k}) \int \frac{p(x_{k+1} \mid x_k)\, p(x_{k+1} \mid y_{1:n})}{p(x_{k+1} \mid y_{1:k})} \mathrm{d}x_{k+1}. \quad (4)$$

The Classic Bayesian Filtering & Smoothing: the Forward-Backward Algorithm

The Bayesian filter is a sequential algorithm, which iterates the following prediction and update steps:

$$p(x_k \mid y_{1:k-1}) = \int p(x_k \mid x_{k-1})\, p(x_{k-1} \mid y_{1:k-1})\, \mathrm{d}x_{k-1}, \quad (2)$$

$$p(x_k \mid y_{1:k}) = \frac{p(y_k \mid x_k)\, p(x_k \mid y_{1:k-1})}{\int p(y_k \mid x_k)\, p(x_k \mid y_{1:k-1})\, \mathrm{d}x_k}. \quad (3)$$

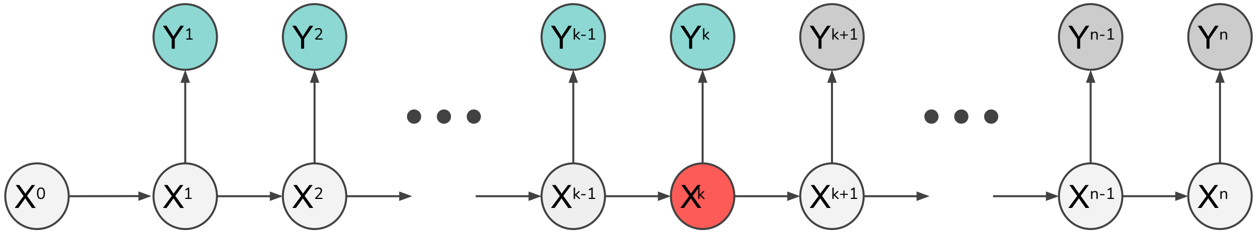


The Classic Bayesian Filtering & Smoothing: the Forward-Backward Algorithm

The Bayesian filter is a sequential algorithm, which iterates the following prediction and update steps:

$$p(x_k \mid y_{1:k-1}) = \int p(x_k \mid x_{k-1})\, p(x_{k-1} \mid y_{1:k-1})\, \mathrm{d}x_{k-1}, \quad (2)$$

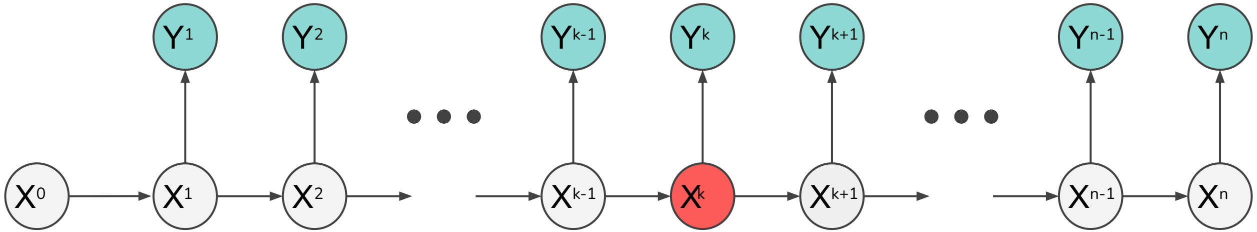
$$p(x_k \mid y_{1:k}) = \frac{p(y_k \mid x_k)\, p(x_k \mid y_{1:k-1})}{\int p(y_k \mid x_k)\, p(x_k \mid y_{1:k-1})\, \mathrm{d}x_k}. \quad (3)$$



The Classic Bayesian Filtering & Smoothing: the Forward-Backward Algorithm

Given the filtering outputs for $k = 1, \ldots, n$, the Bayesian forward-backward smoother consists of the following backward iteration for $k = n - 1, \ldots, 1$:

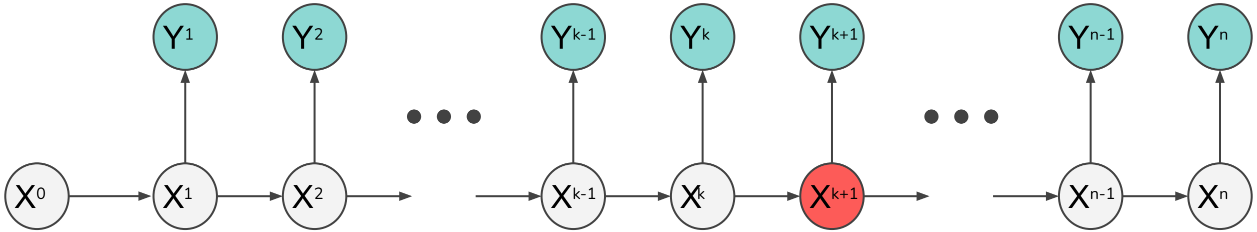
$$p(x_k \mid y_{1:n}) = p(x_k \mid y_{1:k}) \int \frac{p(x_{k+1} \mid x_k)\, p(x_{k+1} \mid y_{1:n})}{p(x_{k+1} \mid y_{1:k})} \mathrm{d}x_{k+1}. \tag{4}$$



The Classic Bayesian Filtering & Smoothing: the Forward-Backward Algorithm

Given the filtering outputs for $k = 1, \ldots, n$, the Bayesian forward-backward smoother consists of the following backward iteration for $k = n-1, \ldots, 1$:

$$p(x_k \mid y_{1:n}) = p(x_k \mid y_{1:k}) \int \frac{p(x_{k+1} \mid x_k)\, p(x_{k+1} \mid y_{1:n})}{p(x_{k+1} \mid y_{1:k})} \mathrm{d}x_{k+1}. \quad (4)$$



Application of Smoothing: the EM algorithm

The General EM Algorithm

Given a joint distribution $p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ over observed variables $\mathbf{X}$ and latent variables $\mathbf{Z}$, governed by parameters $\boldsymbol{\theta}$, the goal is to maximize the likelihood function $p(\mathbf{X}|\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$.

1. Choose an initial setting for the parameters $\boldsymbol{\theta}^{\text{old}}$.

analytical posterior required here

2. **E step** Evaluate $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$.
3. **M step** Evaluate $\boldsymbol{\theta}^{\text{new}}$ given by

$$\boldsymbol{\theta}^{\text{new}} = \arg\max_{\boldsymbol{\theta}} \mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) \tag{9.32}$$

where

$$\mathcal{Q}(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}). \tag{9.33}$$

4. Check for convergence of either the log likelihood or the parameter values. If the convergence criterion is not satisfied, then let

$$\boldsymbol{\theta}^{\text{old}} \leftarrow \boldsymbol{\theta}^{\text{new}} \tag{9.34}$$

and return to step 2.

Application of Smoothing: the EM algorithm

- Maximum likelihood estimator θ of the following state space model can be obtained by the EM algorithm:

$$x_k \sim p_\theta(x_k|x_{k-1})$$
$$y_k \sim p_\theta(y_k|x_k)$$

Algorithm 1 The EM algorithm as the coordinate descent of $\text{ELBO}[\theta, q]$

Let $\text{ELBO}[\theta, q] := \log p_\theta(y_{1:n}) - \text{KL}[q(X_{1:n}|y_{1:n})||p_\theta(X_{1:n}|y_{1:n})] \leq \log p_\theta(y_{1:n})$.

while (θ, q) has not converged **do**

$q \leftarrow p_\theta(X_{1:n}|y_{1:n}) = \arg\max_q \text{ELBO}[\theta, q]$

$\theta \leftarrow \arg\max_{\theta \in \Theta} \text{ELBO}[\theta, q]$

end while

Prefix Sum

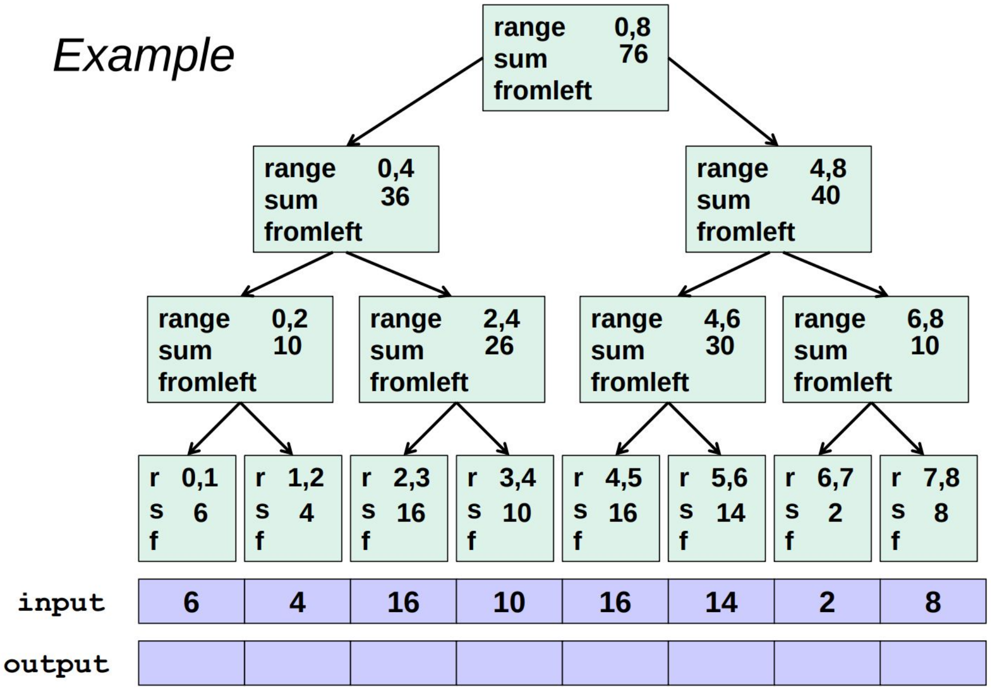
Definition 1: Given a sequence of elements $(a_1, a_2, \ldots, a_n)$, where a_i belongs to a certain set, along with an associative binary operator $\otimes$ on this set, the all-prefix-sums operation computes the sequence

$$(a_1, a_1 \otimes a_2, \ldots, a_1 \otimes \cdots \otimes a_n). \tag{5}$$

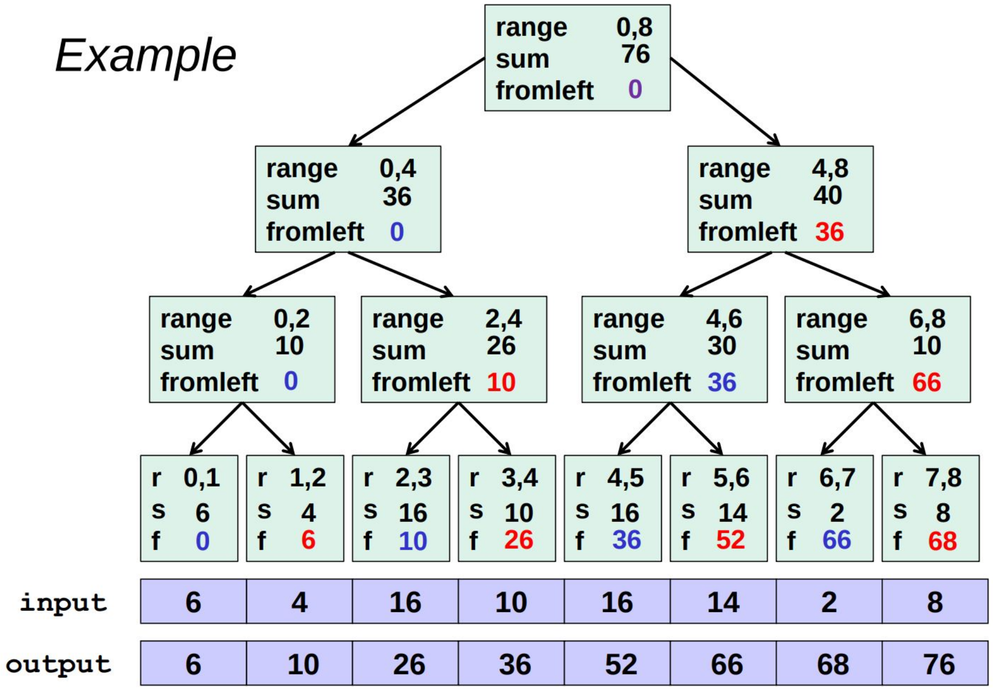
An Example of Cumulative Sum

- Consider cumulative sum of [6, 4, 16, 10, 16, 14, 2, 8], which is [6, 6 + 4, 6 + 4 + 16, ..., 6 + 4 + 16 + 10 + 16 + 14 + 2 + 8]
- For sequence of length n, this can take O(log n) time with **parallel scan**!

Parallel Scan: An Example of Cumulative Sum



Parallel Scan: An Example of Cumulative Sum



Parallel Scan: An Example of Cumulative Sum

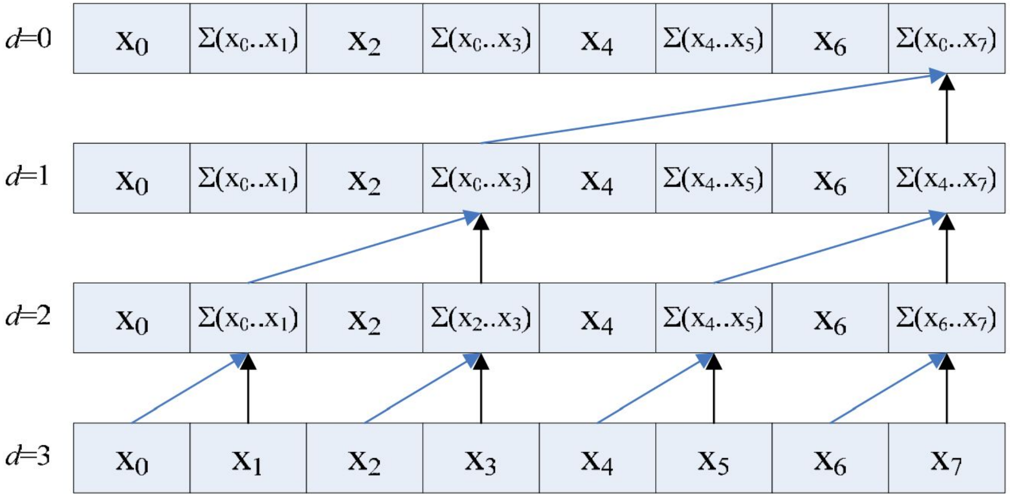


Figure 2: An illustration of the up-sweep, or reduce, phase of a work-efficient sum scan algorithm.

Parallel Scan: An Example of Cumulative Sum

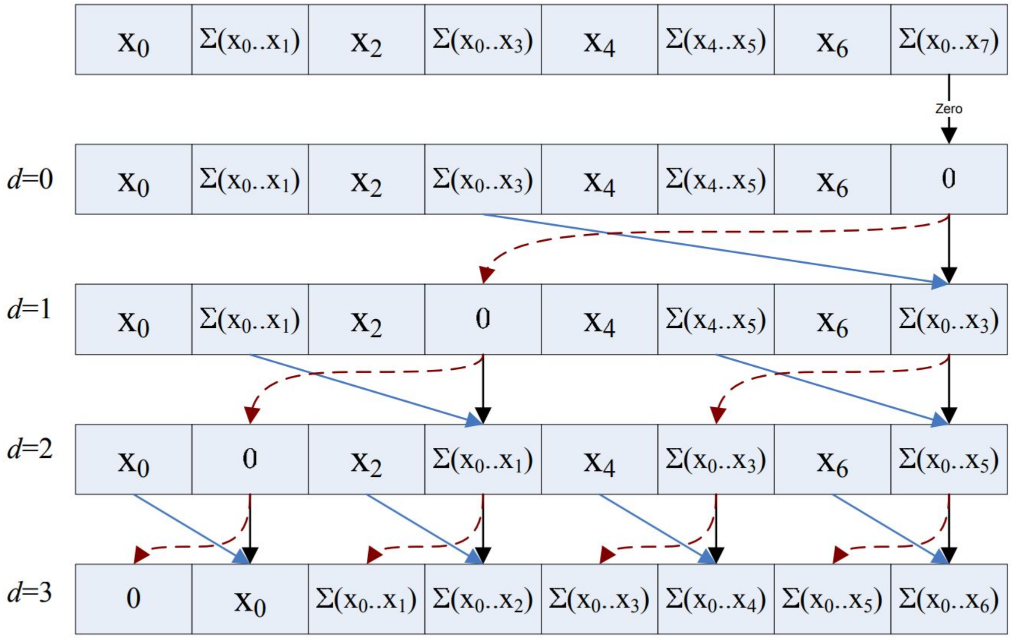


Figure 3: An illustration of the down-sweep phase of the work-efficient parallel sum scan algorithm. Notice that the first step zeros the last element of the array.

Bayesian Filtering as Prefix Sum

A. Bayesian filtering

In order to perform parallel Bayesian filtering, we need to find the suitable element a_k and the binary associative operator $\otimes$. As we will see in this section, an element a consists of a pair $(f, g) \in \mathcal{F}$ where $\mathcal{F}$ is

$$\mathcal{F} = \left\{ (f, g) : \int f(y \mid z) \, \mathrm{d}y = 1 \right\}, \tag{6}$$

and $f : \mathbb{R}^{nx} \times \mathbb{R}^{nx} \to [0, \infty)$ represents a conditional density, and $g : \mathbb{R}^{nx} \to [0, \infty)$ represents a likelihood.

Definition 2: Given two elements $(f_i, g_i) \in \mathcal{F}$ and $(f_j, g_j) \in \mathcal{F}$, the binary associative operator $\otimes$ for Bayesian filtering is

$$(f_i, g_i) \otimes (f_j, g_j) = (f_{ij}, g_{ij}),$$

where

$$f_{ij}(x \mid z) = \frac{\int g_j(y) \, f_j(x \mid y) \, f_i(y \mid z) \, \mathrm{d}y}{\int g_j(y) \, f_i(y \mid z) \, \mathrm{d}y},$$

$$g_{ij}(z) = g_i(z) \int g_j(y) \, f_i(y \mid z) \, \mathrm{d}y.$$

Theorem 3: Given the element $a_k = (f_k, g_k) \in \mathcal{F}$ where

$$f_k(x_k \mid x_{k-1}) = p(x_k \mid y_k, x_{k-1}),$$
$$g_k(x_{k-1}) = p(y_k \mid x_{k-1}),$$

$p(x_1 \mid y_1, x_0) = p(x_1 \mid y_1)$, and $p(y_1 \mid x_0) = p(y_1)$, the k-th prefix sum is

$$a_1 \otimes a_2 \otimes \cdots \otimes a_k = \begin{pmatrix} p(x_k \mid y_{1:k}) \\ p(y_{1:k}) \end{pmatrix}.$$

Bayesian Smoothing as Prefix Sum

B. Bayesian smoothing

The Bayesian smoothing pass requires that the filtering densities have been obtained beforehand. In smoothing, we consider a different type of element a and binary operator $\otimes$ than those used in filtering. As we will see in this section, an element a is a function $a : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \rightarrow [0, \infty)$ that belongs to the set

$$\mathcal{S} = \left\{ a : \int a\left(x \mid z\right) \mathrm{d}x = 1 \right\}.$$

Definition 5: Given two elements $a_i \in \mathcal{S}$ and $a_j \in \mathcal{S}$, the binary associative operator $\otimes$ for Bayesian smoothing is

$$a_i \otimes a_j = a_{ij},$$

where

$$a_{ij}\left(x \mid z\right) = \int a_i\left(x \mid y\right) a_j\left(y \mid z\right) \mathrm{d}y.$$

Theorem 6: Given the element $a_k = p\left(x_k \mid y_{1:k}, x_{k+1}\right) \in \mathcal{S}$, with $a_n = p\left(x_n \mid y_{1:n}\right)$, we have that

$$a_k \otimes a_{k+1} \otimes \cdots \otimes a_n = p\left(x_k \mid y_{1:n}\right).$$

Benchmarks

- Linear Gaussian state space model
  - Kalman filter (KF)
    - filtering
    - forward algorithm
  - Rauch-Tung-Striebel (RTS)
    - smoothing
    - backward algorithm

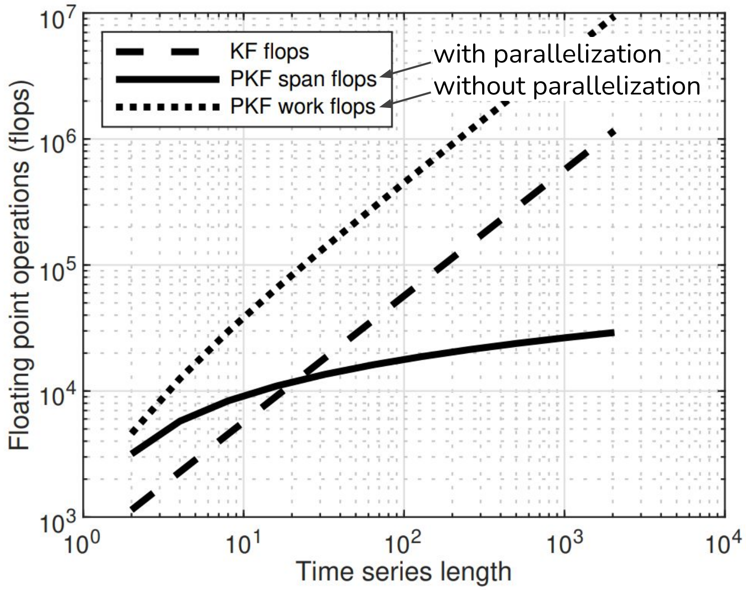


Fig. 3. The flops and the span and work flops for the sequential Kalman filter (KF) and the parallel Kalman filter (PKF).

Benchmarks

- Linear Gaussian state space model
  - Kalman filter (KF)
    - filtering
    - forward algorithm
  - Rauch-Tung-Striebel (RTS)
    - smoothing
    - backward algorithm

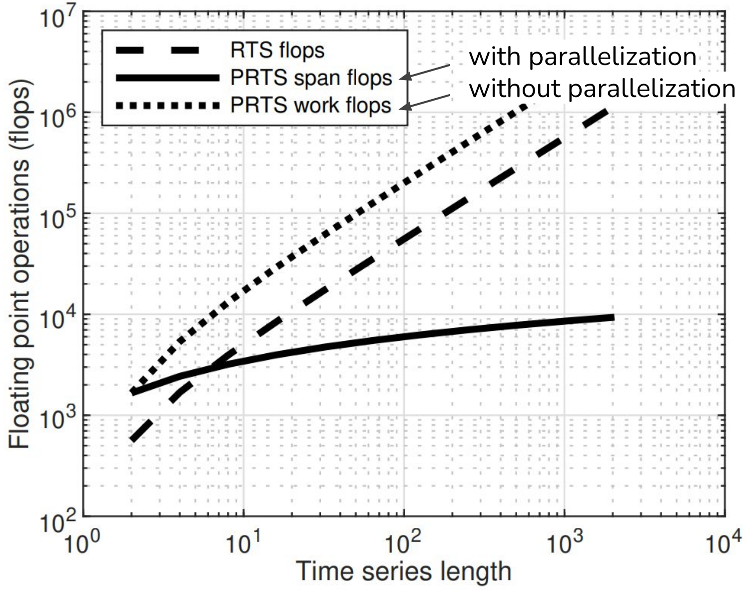


Fig. 4. The flops and the span and work flops for the (sequential) RTS smoother and the parallel RTS (PRTS) smoother.

References

- Särkkä, S., & García-Fernández, Á. F. (2020). Temporal parallelization of Bayesian smoothers. IEEE Transactions on Automatic Control, 66(1), 299-306.
- Bishop, C. M., & Nasrabadi, N. M. (2006). Pattern recognition and machine learning (Vol. 4, No. 4, p. 738). New York: springer.
- Harris, M., Sengupta, S., & Owens, J. D. (2007). Parallel prefix sum (scan) with CUDA. GPU gems, 3(39), 851-876.
- Dan Grossman. (2007). A Sophomoric Introduction to Shared-Memory Parallelism and Concurrency, Lecture 3: Parallel Prefix, Pack, and Sorting. https://homes.cs.washington.edu/~djg/teachingMaterials/spac/grossmanSPAC_lec3.pdf

Sketch of Proof

Bayesian Filtering as Prefix Sum

A. Bayesian filtering

In order to perform parallel Bayesian filtering, we need to find the suitable element a_k and the binary associative operator $\otimes$. As we will see in this section, an element a consists of a pair $(f, g) \in \mathcal{F}$ where $\mathcal{F}$ is

$$\mathcal{F} = \left\{ (f, g) : \int f(y \mid z) \, \mathrm{d}y = 1 \right\}, \tag{6}$$

and $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to [0, \infty)$ represents a conditional density, and $g : \mathbb{R}^{n_x} \to [0, \infty)$ represents a likelihood.

Definition 2: Given two elements $(f_i, g_i) \in \mathcal{F}$ and $(f_j, g_j) \in \mathcal{F}$, the binary associative operator $\otimes$ for Bayesian filtering is

$$(f_i, g_i) \otimes (f_j, g_j) = (f_{ij}, g_{ij}),$$

where

$$f_{ij}(x \mid z) = \frac{\int g_j(y) \, f_j(x \mid y) \, f_i(y \mid z) \, \mathrm{d}y}{\int g_j(y) \, f_i(y \mid z) \, \mathrm{d}y},$$

$$g_{ij}(z) = g_i(z) \int g_j(y) \, f_i(y \mid z) \, \mathrm{d}y.$$

Theorem 3: Given the element $a_k = (f_k, g_k) \in \mathcal{F}$ where

$$f_k(x_k \mid x_{k-1}) = p(x_k \mid y_k, x_{k-1}),$$
$$g_k(x_{k-1}) = p(y_k \mid x_{k-1}),$$

$p(x_1 \mid y_1, x_0) = p(x_1 \mid y_1)$, and $p(y_1 \mid x_0) = p(y_1)$, the k-th prefix sum is

$$a_1 \otimes a_2 \otimes \cdots \otimes a_k = \begin{pmatrix} p(x_k \mid y_{1:k}) \\ p(y_{1:k}) \end{pmatrix}.$$

Bayesian Filtering as Prefix Sum

Let us define

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

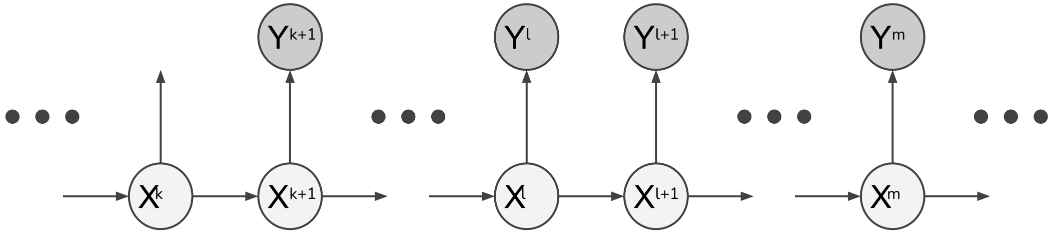
so that

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_\ell$$



Bayesian Filtering as Prefix Sum

Let us define

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

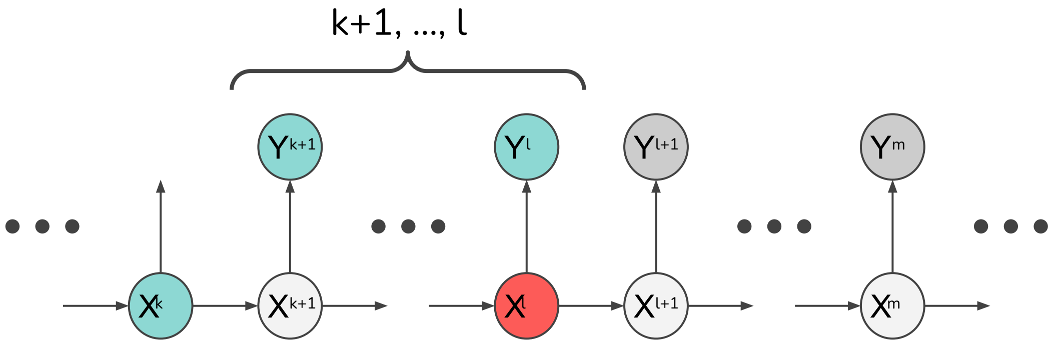
so that

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_\ell$$



Bayesian Filtering as Prefix Sum

Let us define

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

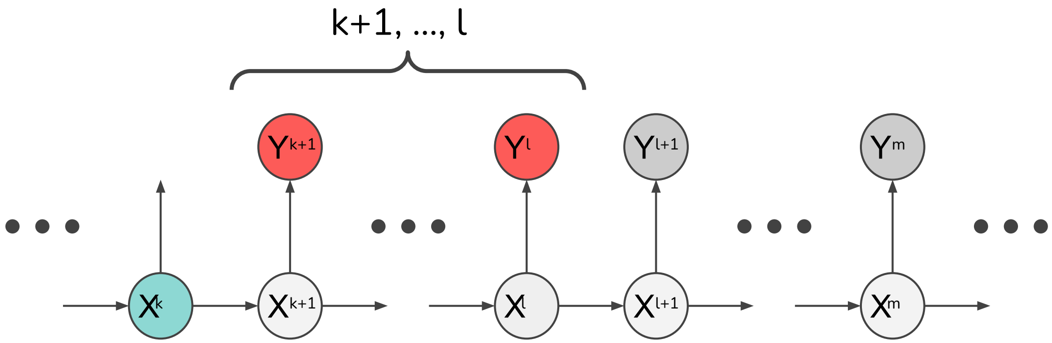
so that

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_\ell$$



Bayesian Filtering as Prefix Sum

Let us define

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

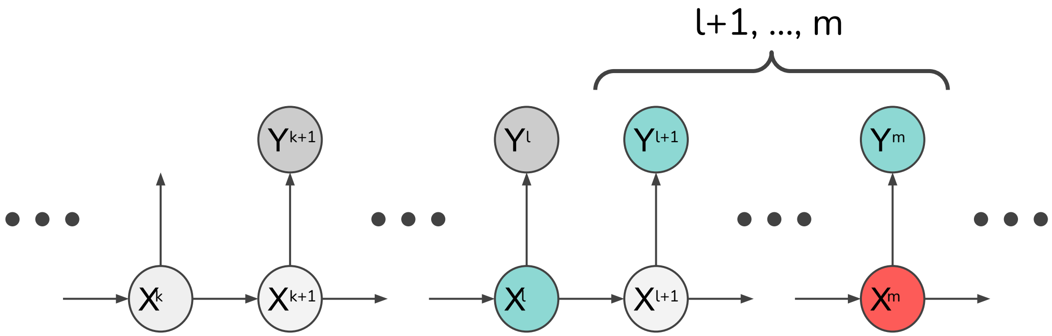
so that

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_\ell$$



Bayesian Filtering as Prefix Sum

Let us define

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

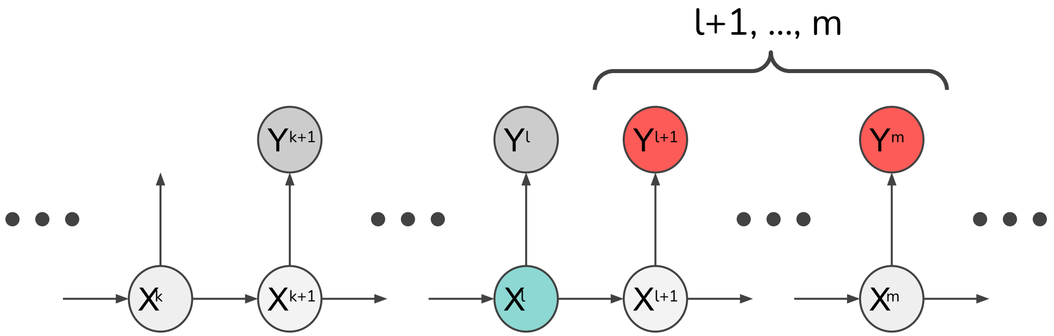
so that

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_\ell$$



Bayesian Filtering as Prefix Sum

Let us define

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

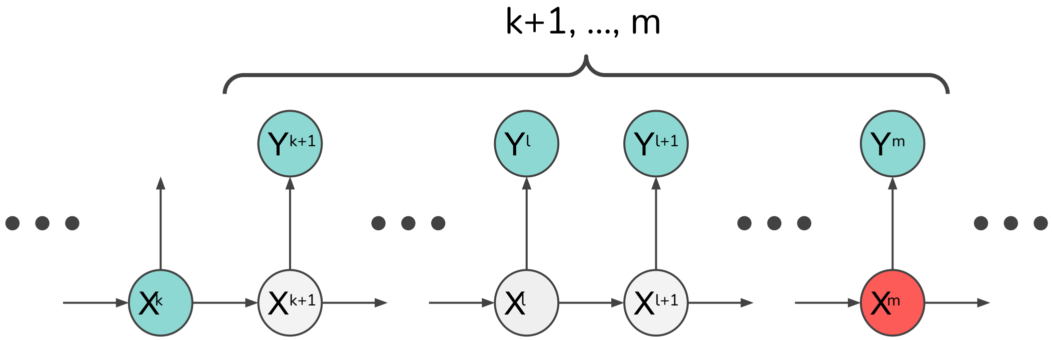
so that

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_\ell$$



Bayesian Filtering as Prefix Sum

Let us define

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

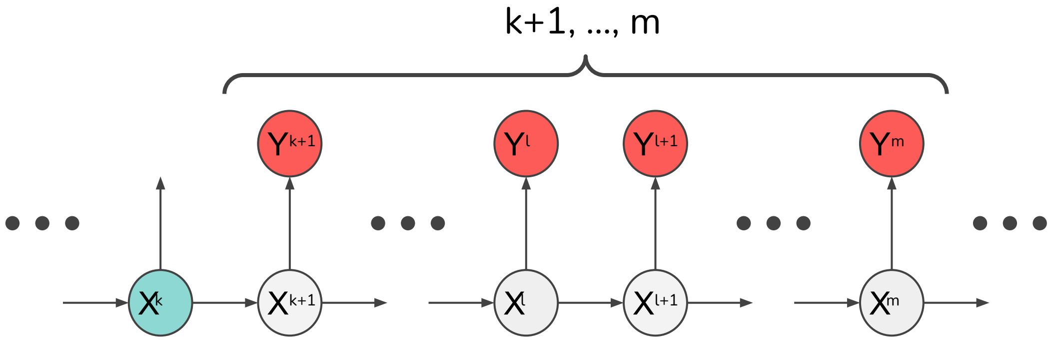
so that

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_\ell$$



Bayesian Filtering as Prefix Sum

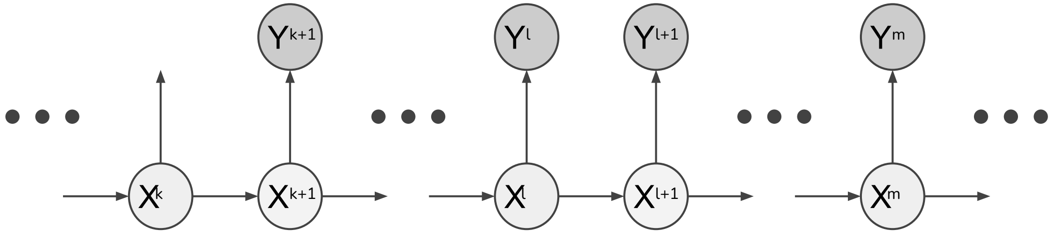
We defined

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

By definition, we can see

$$a_{\ell:\ell} = a_\ell = \begin{pmatrix} p(x_\ell | y_\ell, x_{\ell-1}) \\ p(y_\ell | x_{\ell-1}) \end{pmatrix}$$

for l > 1.



Bayesian Filtering as Prefix Sum

We defined

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

By definition, we can see

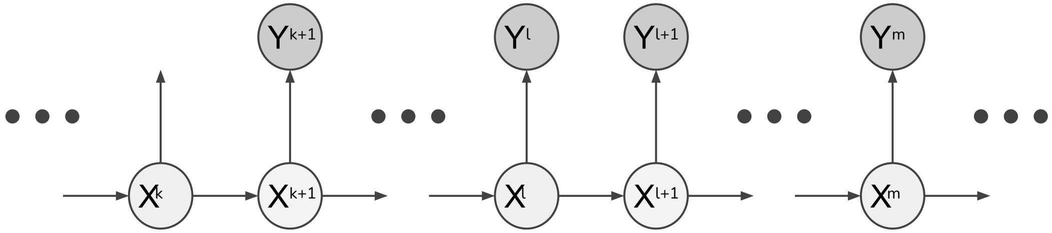
$$a_{\ell:\ell} = a_\ell = \begin{pmatrix} p(x_\ell | y_\ell, x_{\ell-1}) \\ p(y_\ell | x_{\ell-1}) \end{pmatrix}$$

for l > 1.

For l = 1, we set

$$a_{0:0} = \begin{pmatrix} p(x_0 | y_0, x_{-1}) \\ p(y_0 | x_{-1}) \end{pmatrix} = \begin{pmatrix} p(x_0) \\ 1 \end{pmatrix}$$

$$a_1 = a_{0:1}$$



Bayesian Filtering as Prefix Sum

We defined

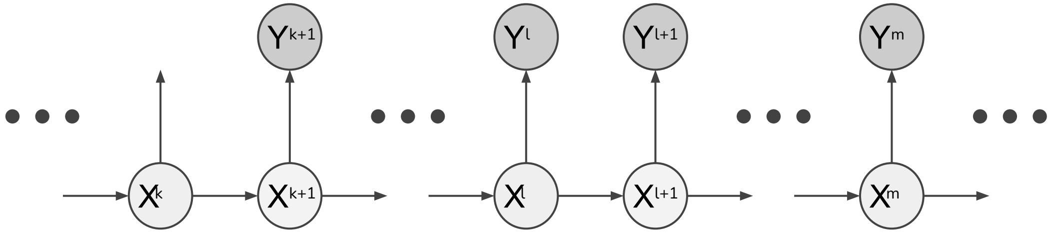
$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

Now, it remains to check

$$a_{k+1:\ell} \otimes a_{\ell+1:m} = a_{k+1:m}$$

so that

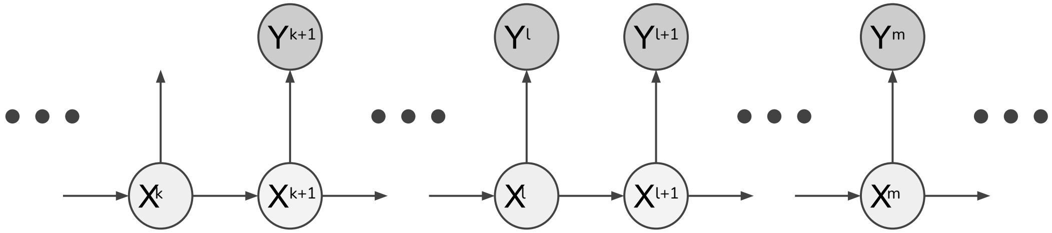
$$a_{0:k} = a_1 \otimes \cdots \otimes a_k = \begin{pmatrix} p(x_k | y_{1:k}) \\ p(y_{1:k}) \end{pmatrix}$$



Bayesian Filtering as Prefix Sum

Indeed,

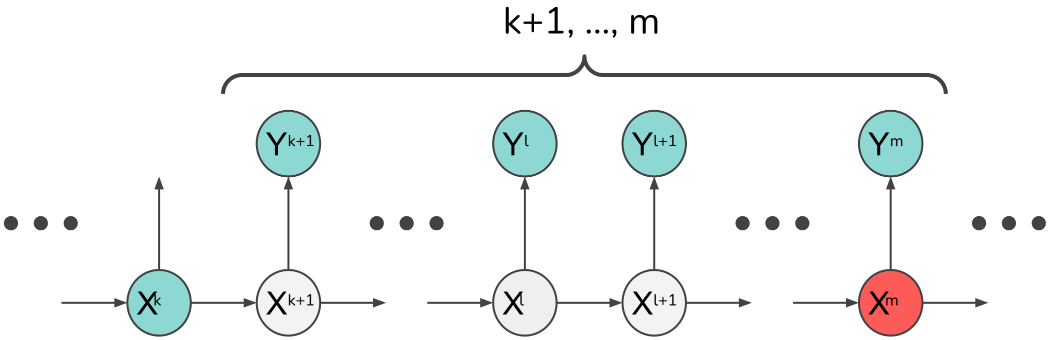
$$\begin{aligned}
&p(x_m|y_{k+1:m}, x_k)\\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k)\\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)}\\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}\\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}\\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}$$



Bayesian Filtering as Prefix Sum

Indeed,

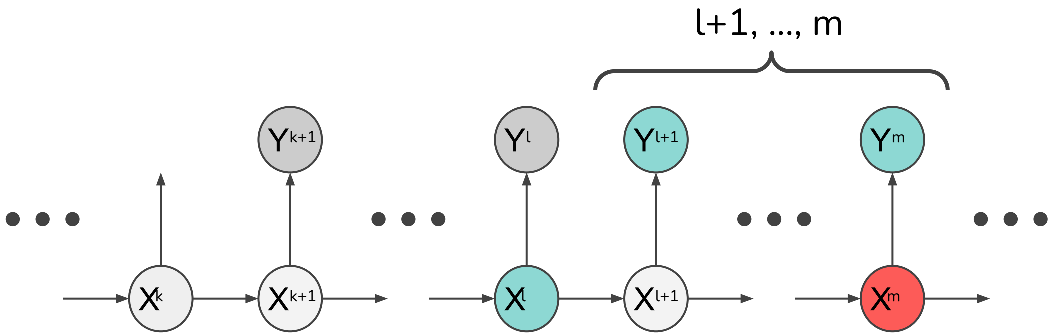
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

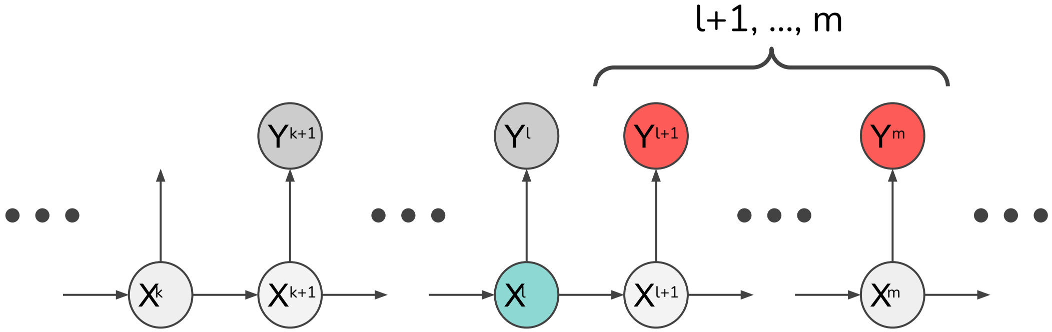
$$
\begin{aligned}
&p(x_m | y_{k+1:m}, x_k) \\
&= p(x_m | y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m} | y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m} | y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k) p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell) p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

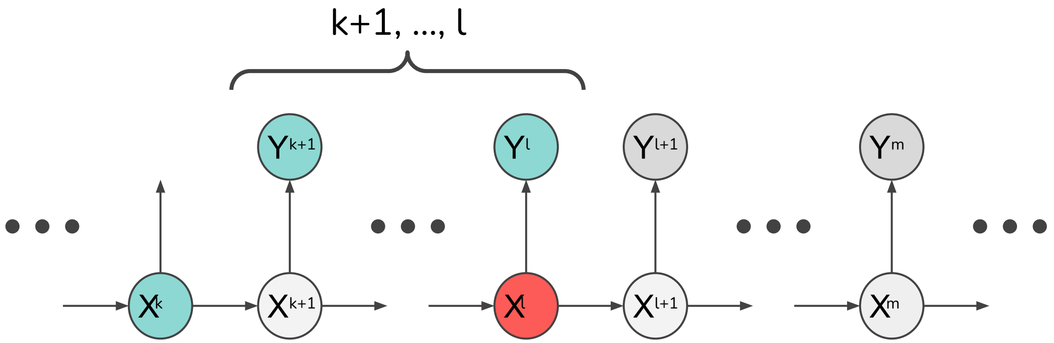
$$
\begin{aligned}
&p(x_m | y_{k+1:m}, x_k) \\
&= p(x_m | y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m} | y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m} | y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k) p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell) p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

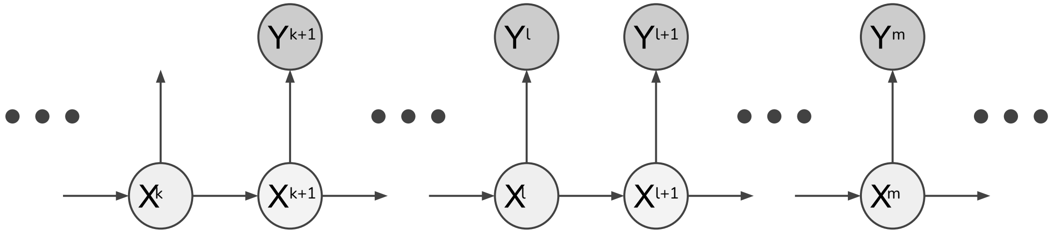
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

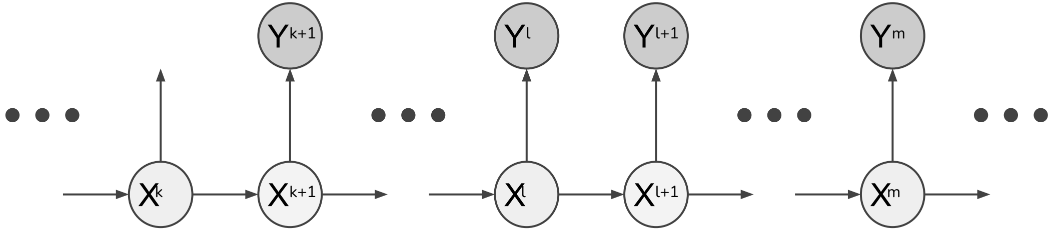
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

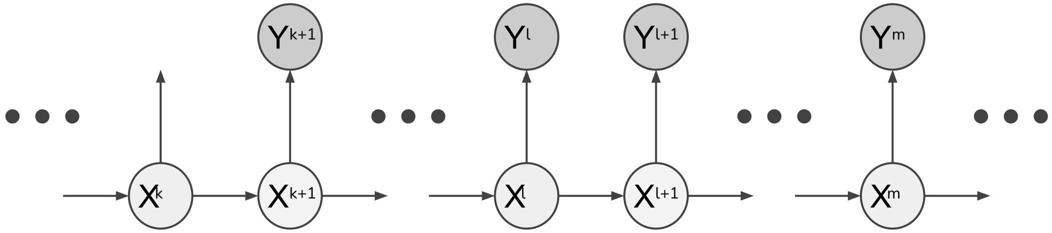
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

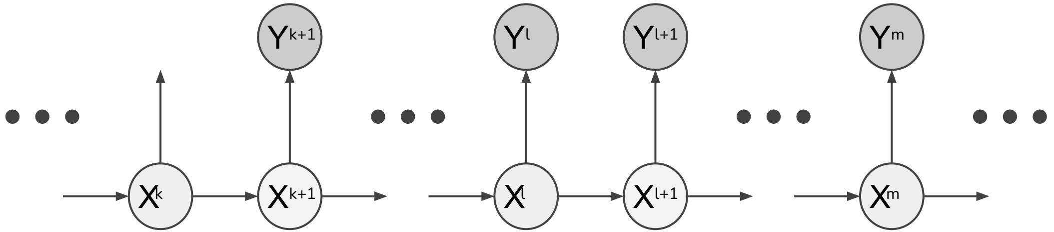
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k)\\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k)\\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)}\\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}\\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}\\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

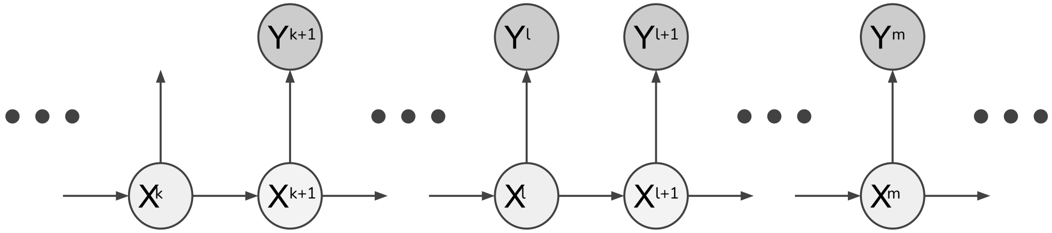
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

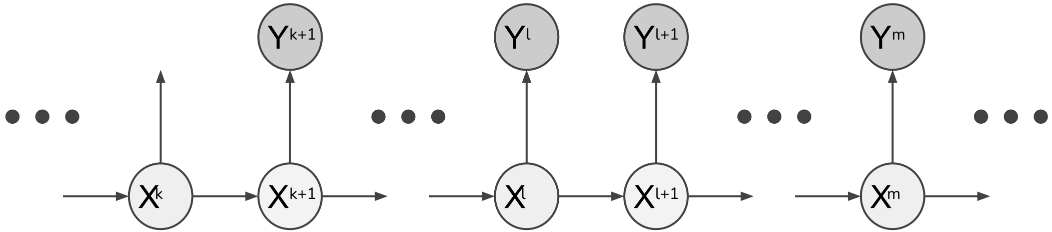
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

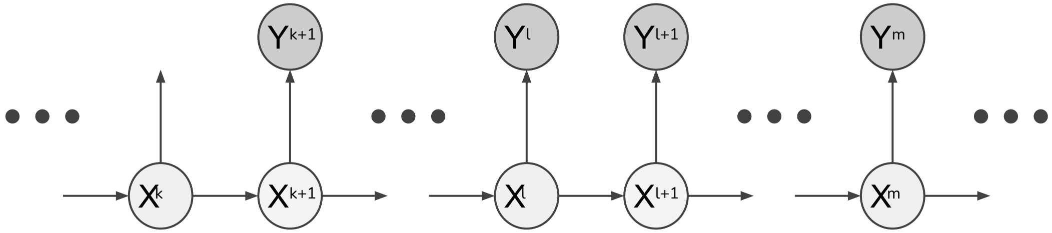
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

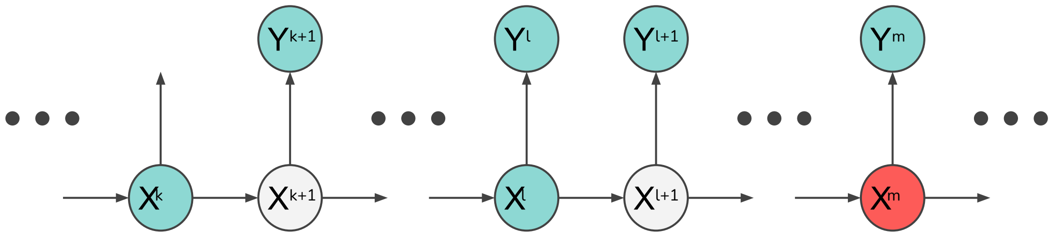
$$
\begin{aligned}
&p(x_m | y_{k+1:m}, x_k) \\
&= p(x_m | y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m} | y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m} | y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k) p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell) p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

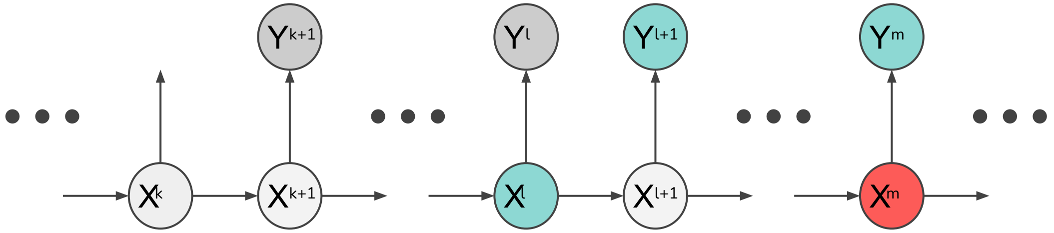
$$
\begin{aligned}
&p(x_m | y_{k+1:m}, x_k) \\
&= p(x_m | y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m} | y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m} | y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k) p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell} \\
&= \frac{\int p(x_m | y_{\ell+1:m}, x_\ell) p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}{\int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) \mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

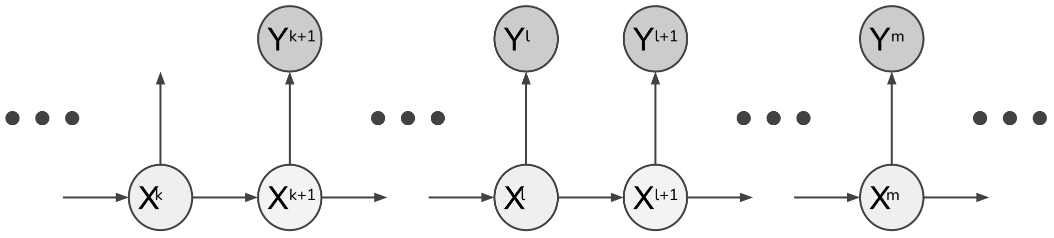
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Indeed,

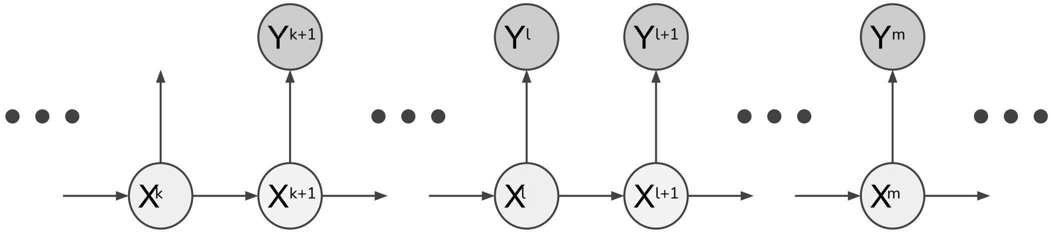
$$
\begin{aligned}
&p(x_m|y_{k+1:m}, x_k) \\
&= p(x_m|y_{k+1:\ell}, y_{\ell+1:m}, x_k) \\
&= \frac{p(x_m, y_{\ell+1:m}|y_{k+1:\ell}, x_k)}{p(y_{\ell+1:m}|y_{k+1:\ell}, x_k)} \\
&= \frac{\int p(x_m, y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}, x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell, y_{k+1:\ell}, x_k)p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell} \\
&= \frac{\int p(x_m|y_{\ell+1:m}, x_\ell)p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}{\int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)\mathrm{d}x_\ell}
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

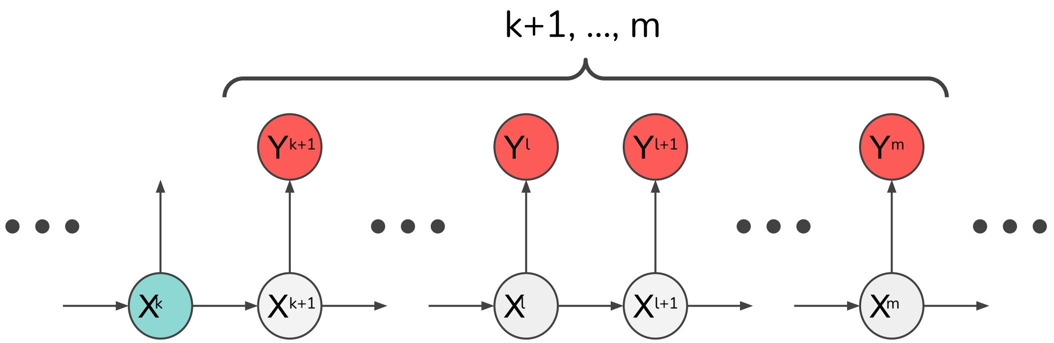
$$
\begin{aligned}
& p(y_{k+1:m}|x_k) \\
&= \int p(y_{k+1:m}, x_\ell | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m} | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

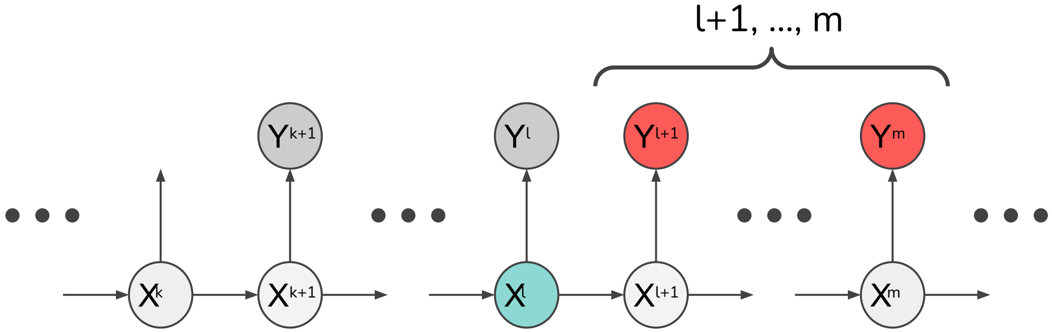
$$
\begin{aligned}
p(y_{k+1:m}|x_k) &= \int p(y_{k+1:m}, x_\ell | x_k)\mathrm{d}x_\ell \\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m} | x_k)\mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

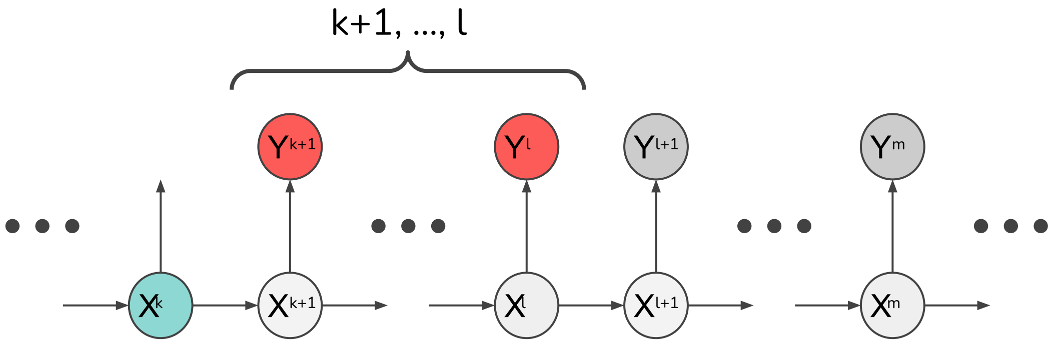
$$
\begin{aligned}
&p(y_{k+1:m}|x_k) \\
&= \int p(y_{k+1:m}, x_\ell|x_k)\mathrm{d}x_\ell \\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m}|x_k)\mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

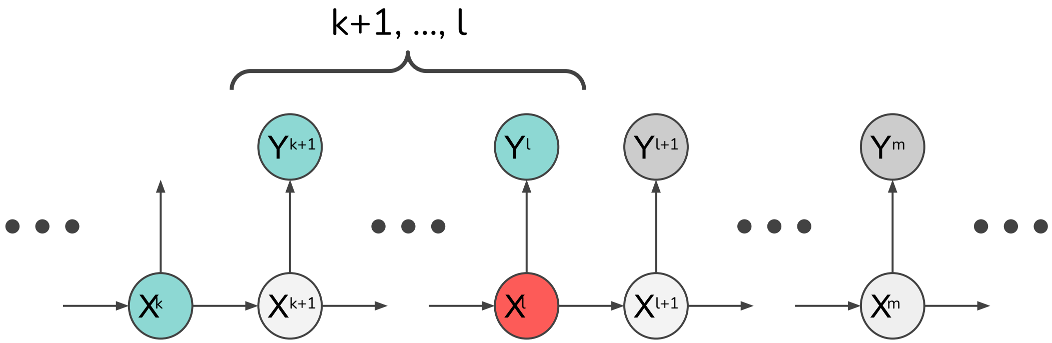
$$
\begin{aligned}
&p(y_{k+1:m}|x_k) \\
&= \int p(y_{k+1:m}, x_\ell | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m} | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

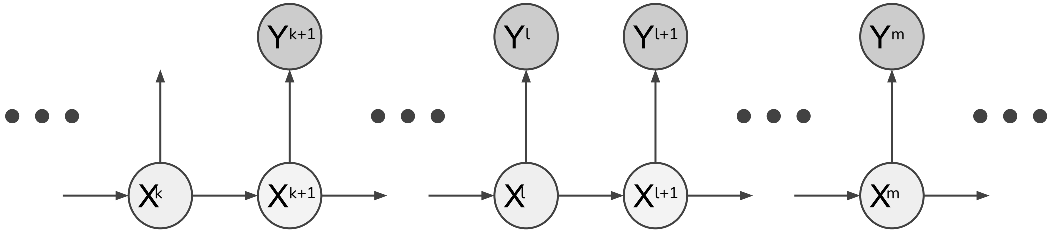
$$\begin{aligned} p(y_{k+1:m}|x_k) &= \int p(y_{k+1:m}, x_\ell|x_k)\mathrm{d}x_\ell \\ &= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m}|x_k)\mathrm{d}x_\ell \\ &= \int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell \\ &= \int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell \end{aligned}$$



Bayesian Filtering as Prefix Sum

Also, we have

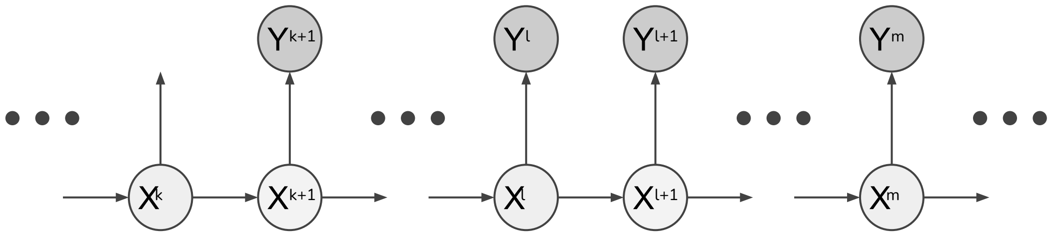
$$
\begin{aligned}
&p(y_{k+1:m}|x_k)\\
&= \int p(y_{k+1:m}, x_\ell|x_k)\mathrm{d}x_\ell\\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m}|x_k)\mathrm{d}x_\ell\\
&= \int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell\\
&= \int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

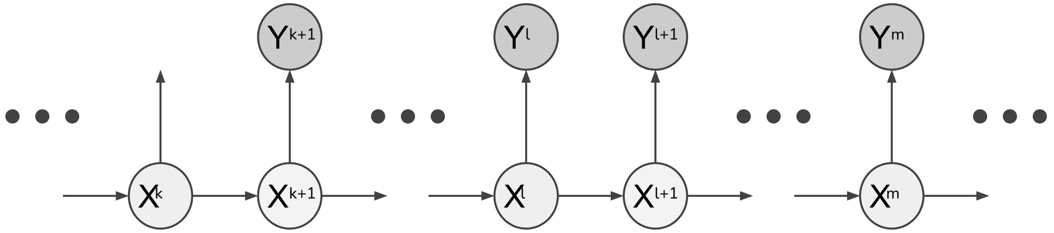
$$
\begin{aligned}
&p(y_{k+1:m}|x_k) \\
&= \int p(y_{k+1:m}, x_\ell|x_k)\mathrm{d}x_\ell \\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m}|x_k)\mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m}|x_\ell, y_{k+1:\ell}, x_k)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m}|x_\ell)p(x_\ell|y_{k+1:\ell}, x_k)p(y_{k+1:\ell}|x_k)\mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

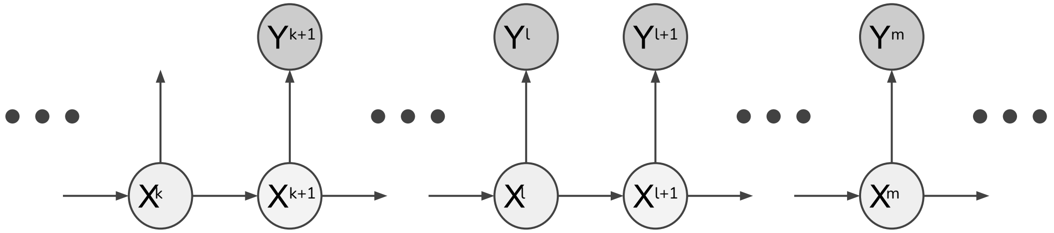
$$
\begin{aligned}
&p(y_{k+1:m}|x_k)\\
&= \int p(y_{k+1:m}, x_\ell | x_k) \mathrm{d}x_\ell\\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m} | x_k) \mathrm{d}x_\ell\\
&= \int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell\\
&= \int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Also, we have

$$
\begin{aligned}
&p(y_{k+1:m}|x_k) \\
&= \int p(y_{k+1:m}, x_\ell | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{k+1:\ell}, x_\ell, y_{\ell+1:m} | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m} | x_\ell, y_{k+1:\ell}, x_k) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell \\
&= \int p(y_{\ell+1:m} | x_\ell) p(x_\ell | y_{k+1:\ell}, x_k) p(y_{k+1:\ell} | x_k) \mathrm{d}x_\ell
\end{aligned}
$$



Bayesian Filtering as Prefix Sum

Thus, for

$$a_{k+1:\ell} := \begin{pmatrix} p(x_\ell | y_{k+1:\ell}, x_k) \\ p(y_{k+1:\ell} | x_k) \end{pmatrix}$$

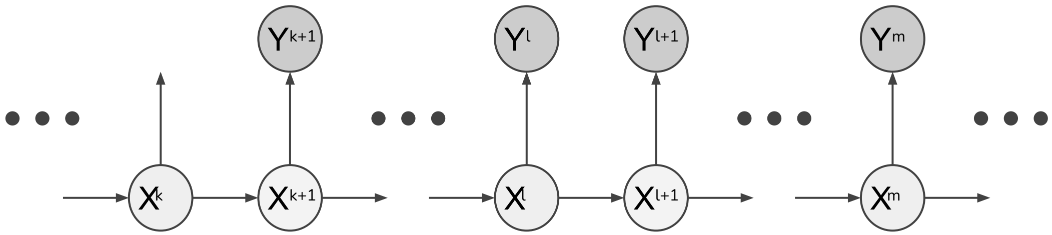
and

$$a_{\ell+1:m} = \begin{pmatrix} p(x_m | y_{\ell+1:m}, x_\ell) \\ p(y_{\ell+1:m} | x_\ell) \end{pmatrix}$$

$$a_{k+1:m} = \begin{pmatrix} p(x_m | y_{k+1:m}, x_k) \\ p(y_{k+1:m} | x_k) \end{pmatrix}$$

we proved that

$$a_{k+1:\ell} \otimes a_{\ell+1:m} = a_{k+1:m}$$



Bayesian Smoothing as Prefix Sum

B. Bayesian smoothing

The Bayesian smoothing pass requires that the filtering densities have been obtained beforehand. In smoothing, we consider a different type of element a and binary operator $\otimes$ than those used in filtering. As we will see in this section, an element a is a function $a : \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to [0, \infty)$ that belongs to the set

$$\mathcal{S} = \left\{ a : \int a\left(x \mid z\right) \mathrm{d}x = 1 \right\}.$$

Definition 5: Given two elements $a_i \in \mathcal{S}$ and $a_j \in \mathcal{S}$, the binary associative operator $\otimes$ for Bayesian smoothing is

$$a_i \otimes a_j = a_{ij},$$

where

$$a_{ij}\left(x \mid z\right) = \int a_i\left(x \mid y\right) a_j\left(y \mid z\right) \mathrm{d}y.$$

Theorem 6: Given the element $a_k = p\left(x_k \mid y_{1:k}, x_{k+1}\right) \in \mathcal{S}$, with $a_n = p\left(x_n \mid y_{1:n}\right)$, we have that

$$a_k \otimes a_{k+1} \otimes \cdots \otimes a_n = p\left(x_k \mid y_{1:n}\right).$$

Bayesian Smoothing as Prefix Sum

Let us define

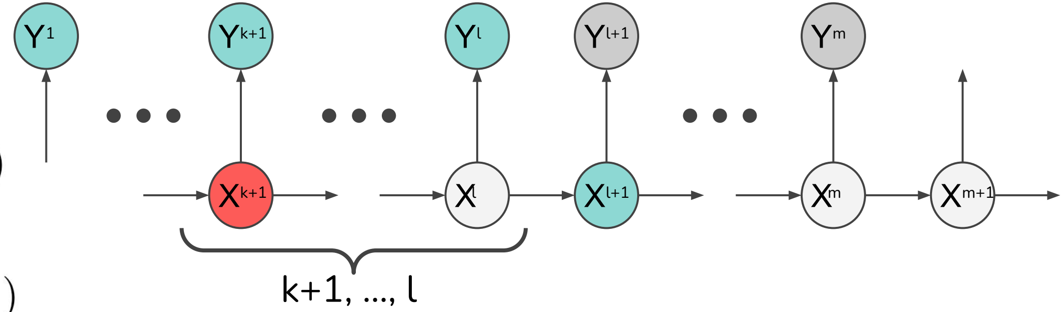
$$a_{k+1:\ell} := p(x_{k+1} | y_{1:\ell}, x_{\ell+1})$$

so that

$$a_{\ell+1:m} = p(x_{\ell+1} | y_{1:m}, x_{m+1})$$
$$a_{k+1:\ell} = p(x_{k+1} | y_{1:m}, x_{m+1})$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_{\ell}$$



Bayesian Smoothing as Prefix Sum

Let us define

$$a_{k+1:\ell} := p(x_{k+1} | y_{1:\ell}, x_{\ell+1})$$

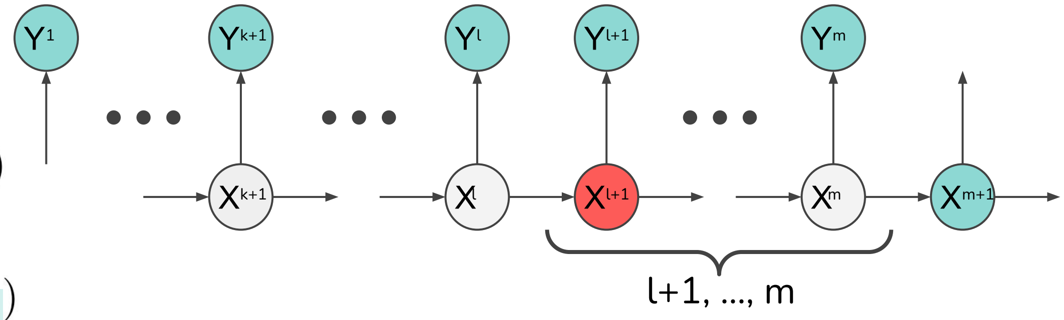
so that

$$a_{\ell+1:m} = p(x_{\ell+1} | y_{1:m}, x_{m+1})$$

$$a_{k+1:\ell} = p(x_{k+1} | y_{1:m}, x_{m+1})$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_{\ell}$$



Bayesian Smoothing as Prefix Sum

Let us define

$$a_{k+1:\ell} := p(x_{k+1} | y_{1:\ell}, x_{\ell+1})$$

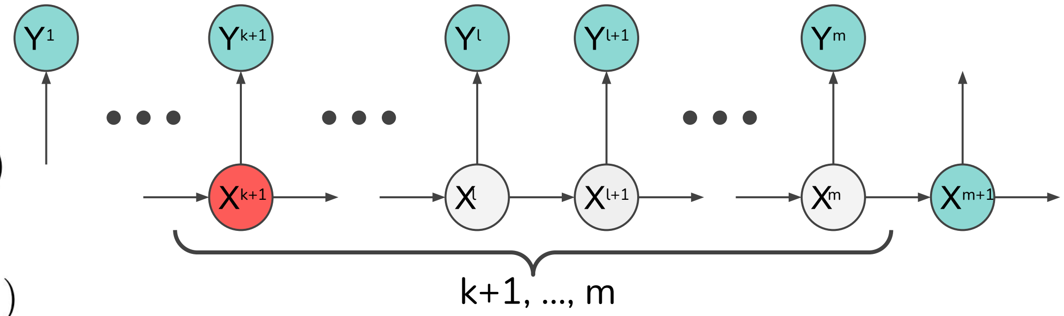
so that

$$a_{\ell+1:m} = p(x_{\ell+1} | y_{1:m}, x_{m+1})$$

$$a_{k+1:\ell} = p(x_{k+1} | y_{1:m}, x_{m+1})$$

for k < l < m. Then, we can show that

$$a_{k+1:\ell} = a_{k+1} \otimes \cdots \otimes a_{\ell}$$



Bayesian Smoothing as Prefix Sum

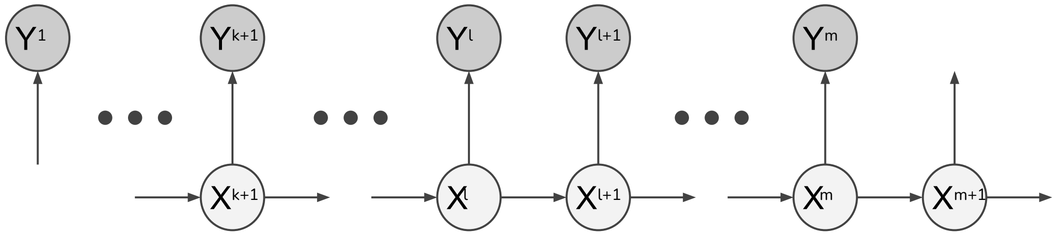
As we defined

$$a_{k+1:\ell} := p(x_{k+1}|y_{1:\ell}, x_{\ell+1})$$

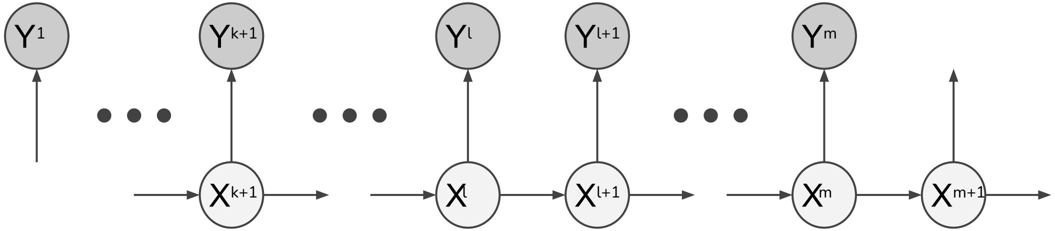
it follows from the definition that

$$a_{\ell:\ell} = a_\ell = p(x_\ell|y_{1:\ell}, x_{\ell+1})$$

for l < n.



Bayesian Smoothing as Prefix Sum



As we defined

$$a_{k+1:\ell} := p(x_{k+1}|y_{1:\ell}, x_{\ell+1})$$

it follows from the definition that

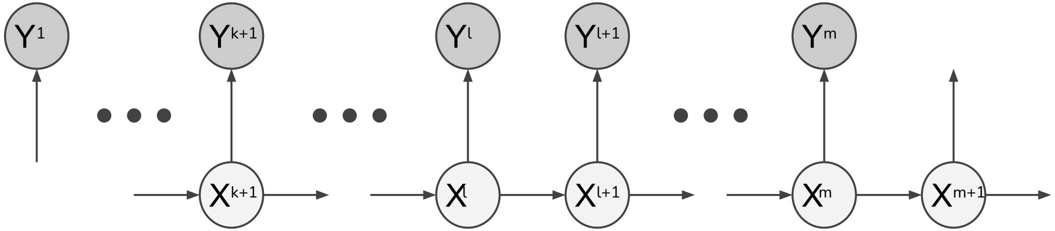
$$a_{\ell:\ell} = a_\ell = p(x_\ell|y_{1:\ell}, x_{\ell+1})$$

for l < n.

We handle the boundary condition by setting

$$a_{n:n} = p(x_n|y_{1:n}, x_{n+1}) = p(x_n|y_{1:n}) = a_n$$

Bayesian Smoothing as Prefix Sum



As we defined

$$a_{k+1:\ell} := p(x_{k+1}|y_{1:\ell}, x_{\ell+1})$$

it follows from the definition that

$$a_{\ell:\ell} = a_\ell = p(x_\ell|y_{1:\ell}, x_{\ell+1})$$

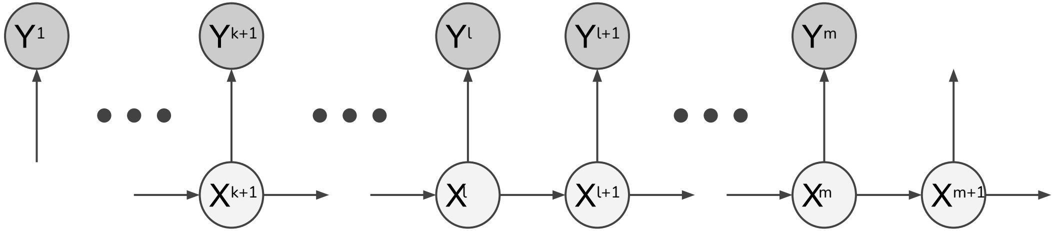
for l < n.

We handle the boundary condition by setting

$$a_{n:n} = p(x_n|y_{1:n}, x_{n+1}) = p(x_n|y_{1:n}) = a_n$$

Bayesian Smoothing as Prefix Sum

Finally, we prove the recursion by

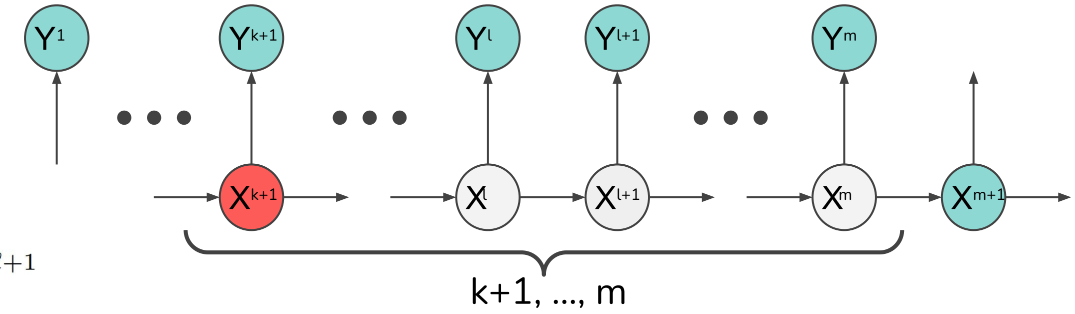


$$
\begin{aligned}
&p(x_{k+1}|y_{1:m}, x_{m+1}) \\
&= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1}
\end{aligned}
$$

Bayesian Smoothing as Prefix Sum

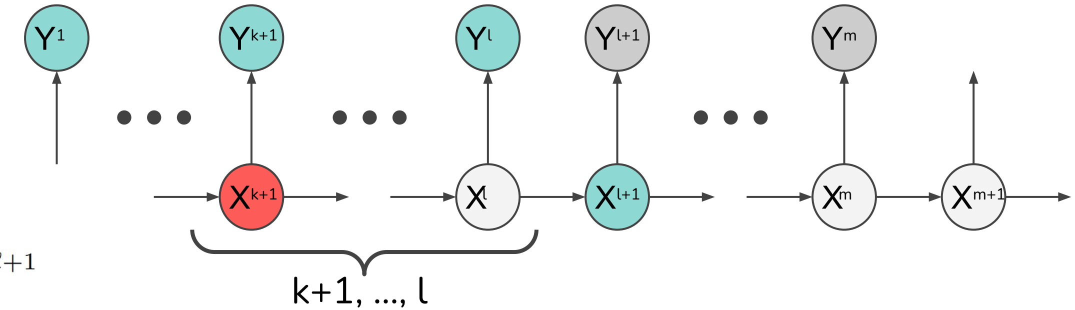
Finally, we prove the recursion by

$$p(x_{k+1}|y_{1:m}, x_{m+1})$$
$$= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1}$$
$$= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1}$$
$$= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1}$$



Bayesian Smoothing as Prefix Sum

Finally, we prove the recursion by

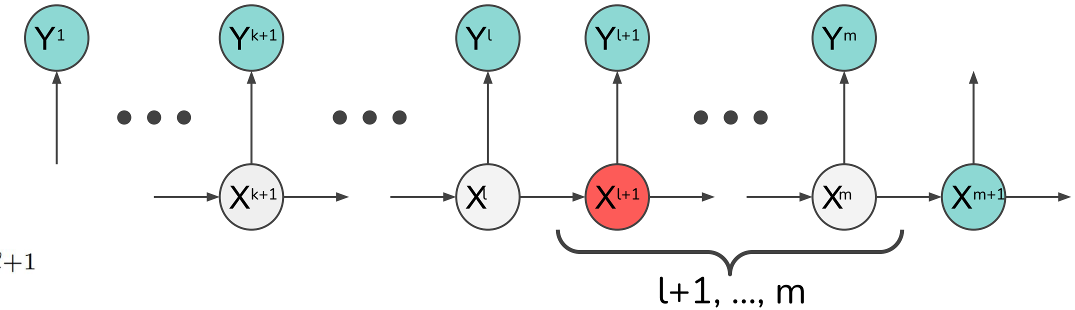


$$
\begin{aligned}
&p(x_{k+1}|y_{1:m}, x_{m+1}) \\
&= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1}
\end{aligned}
$$

Bayesian Smoothing as Prefix Sum

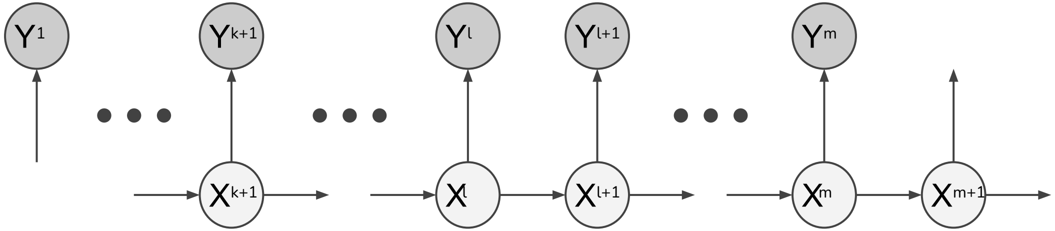
Finally, we prove the recursion by

$$p(x_{k+1}|y_{1:m}, x_{m+1})$$
$$= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1}$$
$$= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1}$$
$$= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1}$$



Bayesian Smoothing as Prefix Sum

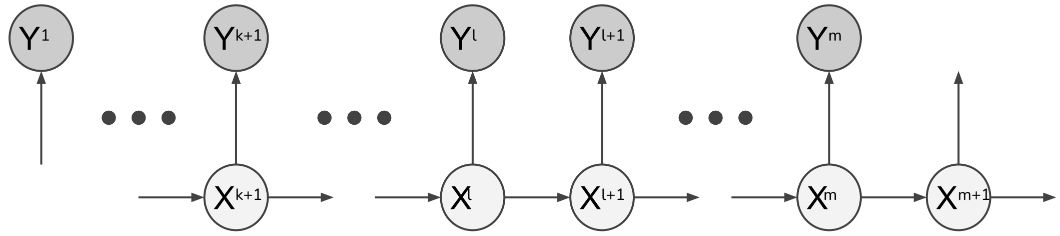
Finally, we prove the recursion by



$$
\begin{aligned}
&p(x_{k+1}|y_{1:m}, x_{m+1}) \\
&= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1}
\end{aligned}
$$

Bayesian Smoothing as Prefix Sum

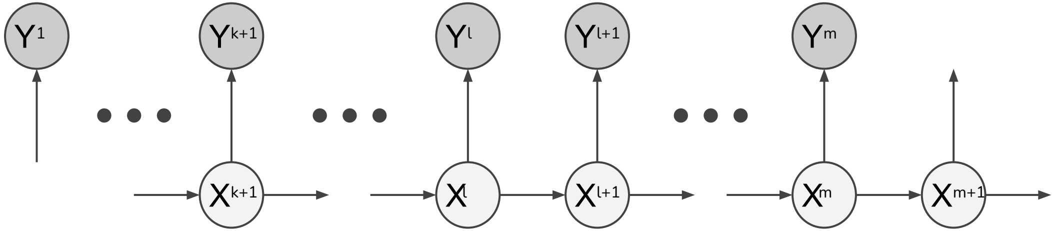
Finally, we prove the recursion by



$$
\begin{aligned}
&p(x_{k+1}|y_{1:m}, x_{m+1}) \\
&= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1}
\end{aligned}
$$

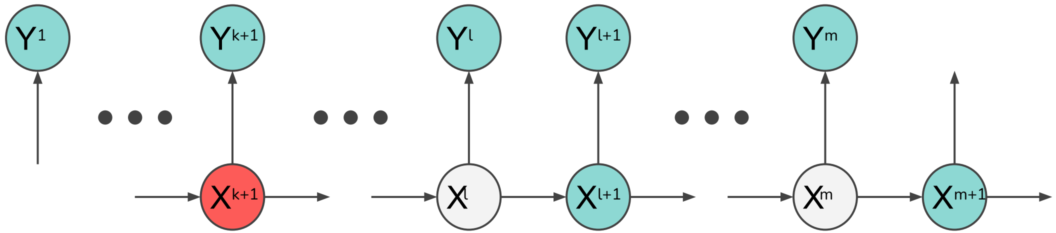
Bayesian Smoothing as Prefix Sum

Finally, we prove the recursion by



$$
\begin{aligned}
&p(x_{k+1}|y_{1:m}, x_{m+1}) \\
&= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell}) p(x_{\ell+1}|y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1}
\end{aligned}
$$

Bayesian Smoothing as Prefix Sum

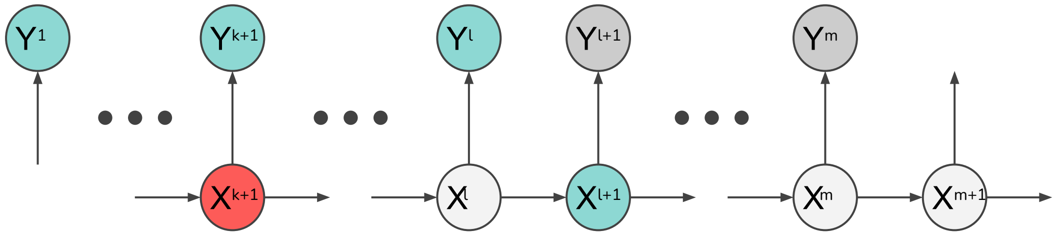


Finally, we prove the recursion by

$$
\begin{aligned}
&p(x_{k+1} | y_{1:m}, x_{m+1}) \\
&= \int p(x_{k+1}, x_{\ell+1} | y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1} | x_{\ell+1}, y_{1:m}, x_{m+1}) p(x_{\ell+1} | y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1} | x_{\ell+1}, y_{1:\ell}) p(x_{\ell+1} | y_{1:m}, x_{m+1}) \mathrm{d}x_{\ell+1}
\end{aligned}
$$

Bayesian Smoothing as Prefix Sum

Finally, we prove the recursion by



$$
\begin{aligned}
&p(x_{k+1}|y_{1:m}, x_{m+1}) \\
&= \int p(x_{k+1}, x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:m}, x_{m+1})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1} \\
&= \int p(x_{k+1}|x_{\ell+1}, y_{1:\ell})p(x_{\ell+1}|y_{1:m}, x_{m+1})\mathrm{d}x_{\ell+1}
\end{aligned}
$$