Note on the Multi-Resolution Smoother

Pulong Ma
Statistical and Applied Mathematical Sciences Institute and Duke University

May 25, 2019

1 Spatio-temporal state-space model

Let $\mathbf{y}_t$ be an n_t-dimensional vector of data observed at time $t = 1, 2, \ldots, T$. Suppose that we are interested in the following linear Gaussian state-space model (SSM):

$$\mathbf{y}_t = \mathbf{H}_t\mathbf{x}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim \mathcal{N}_{n_t}(\mathbf{0}, \mathbf{R}_t), \tag{1.1}$$

$$\mathbf{x}_t = \mathbf{A}_t\mathbf{x}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim \mathcal{N}_{n_\mathcal{G}}(\mathbf{0}, \mathbf{Q}_t), \tag{1.2}$$

with the initial state $\mathbf{x}_0 \sim \mathcal{N}_{n_\mathcal{G}}(\boldsymbol{\mu}_{0|0}, \boldsymbol{\Sigma}_{0|0})$. The observation mapping matrix $\mathbf{H}_t$ relates the state vector to the observations. The observation error covariance matrix $\mathbf{R}_t$ will be assumed to be diagonal or block-diagonal. The current state evolves from previous state via the evolution matrix $\mathbf{A}_t$, which is further assumed to be a sparse matrix. Let $\mathbf{y}_{1:t} := (\mathbf{y}_1^\top, \ldots, \mathbf{y}_t^\top)^\top$ be a vector of observations from time 1 to time t. In what follows, we give the Kalman filtering and Kalman smoothing procedure, respectively.

Under Gaussianity assumption, the filtering distribution $[\mathbf{x}_t \mid \mathbf{y}_{1:t}]$ is also Gaussian. To fix the notation, we define $\boldsymbol{\mu}_{t|t} := E[\mathbf{x}_t \mid \mathbf{y}_{1:t}]$ and $\boldsymbol{\Sigma}_{t|t} := Cov(\mathbf{x}_t \mid \mathbf{y}_{1:t})$. To derive the Kalman filtering procedure, we first give the one-step ahead forecasting distribution:

$$\mathbf{x}_t \mid \mathbf{y}_{1:t-1} \sim \mathcal{N}_{n_\mathcal{G}}(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1}), \tag{1.3}$$

where $\boldsymbol{\mu}_{t|t-1} := \mathbf{A}_t\boldsymbol{\mu}_{t-1|t-1}$ and $\boldsymbol{\Sigma}_{t|t-1} := \mathbf{A}_t\boldsymbol{\Sigma}_{t-1|t-1}\mathbf{A}_t^\top + \mathbf{Q}_t$.

Based on Bayes' theorem, it follows that $[\mathbf{x}_t \mid \mathbf{y}_{1:t}] \propto [\mathbf{y}_t \mid \mathbf{x}_t][\mathbf{x}_t \mid \mathbf{y}_{1:t-1}]$. Thus, we have

$$\boldsymbol{\mu}_{t|t} := \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t(\mathbf{y}_t - \mathbf{H}_t\boldsymbol{\mu}_{t|t-1}), \tag{1.4}$$

$$\boldsymbol{\Sigma}_{t|t} := \boldsymbol{\Sigma}_{t|t-1} - \mathbf{K}_t\mathbf{H}_t\boldsymbol{\Sigma}_{t|t-1}, \tag{1.5}$$

where $\mathbf{K}_t := \boldsymbol{\Sigma}_{t|t-1}\mathbf{H}_t^\top(\mathbf{H}_t\boldsymbol{\Sigma}_{t|t-1}\mathbf{H}^\top + \mathbf{R}_t)^{-1}$ is the $n_\mathcal{G} \times n_t$ Kalman gain matrix.

As with the Kalman filter derivation, let $\boldsymbol{\mu}_{t|T} := E(\mathbf{x}_t \mid \mathbf{y}_{1:T})$ and $\boldsymbol{\Sigma}_{t|T} := Cov(\mathbf{x}_t \mid \mathbf{y}_{1:T})$. Then the linear Gaussian state-space model implies that the smoothing distribution will also be Gaussian: $\mathbf{x}_t \mid \mathbf{y}_{1:T} \sim \mathcal{N}_{n_\mathcal{G}}(\boldsymbol{\mu}_{t|T}, \boldsymbol{\Sigma}_{t|T})$. Notice that

$$[\mathbf{x}_t \mid \mathbf{y}_{1:T}] = \int [\mathbf{x}_t \mid \mathbf{x}_{t+1}, \mathbf{y}_{1:T}][\mathbf{x}_{t+1} \mid \mathbf{y}_{1:T}]\, d\mathbf{x}_{t+1},$$

where $[\mathbf{x}_t \mid \mathbf{x}_{t+1}, \mathbf{y}_{1:T}] \propto [\mathbf{x}_{t+1} \mid \mathbf{x}_t][\mathbf{x}_t \mid \mathbf{y}_{1:t}]$. It follows that the conditional mean and conditional covariance in the smoothing distribution are given by

$$\boldsymbol{\mu}_{t|T} := \boldsymbol{\mu}_{t|t} + \mathbf{J}_t(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t}), \tag{1.6}$$

$$\boldsymbol{\Sigma}_{t|T} := \boldsymbol{\Sigma}_{t|t} + \mathbf{J}_t(\boldsymbol{\Sigma}_{t+1|T} - \boldsymbol{\Sigma}_{t+1|t})\mathbf{J}_t^\top, \tag{1.7}$$

where $\mathbf{J}_t := \boldsymbol{\Sigma}_{t|t}\mathbf{A}_{t+1}^\top\boldsymbol{\Sigma}_{t+1|t}^{-1}$. Note that the smoothing distribution is well-defined for $t = 0$.

2 A multi-resolution approximation approach

In this section, we will give a multi-resolution smoother (MRS) for the spatio-temporal SSM introduced in Section 1 when the dimension of state $n_\mathcal{G}$ or the data sizes n_t are in order of 10^4 to 10^9. The key here is to efficiently compute the mean and covariance matrix of the smoothing distribution $[\mathbf{x}_t \mid \mathbf{y}_{1:T}]$ or to efficiently sample from the smoothing distribution $[\mathbf{x}_t \mid \mathbf{y}_{1:T}]$.

To derive the efficient MRS, we first review some results on multi-resolution approximation approach introduced in Katzfuss (2017) and multi-resolution filter (MRF) introduced in Jurek and Katzfuss (2018).

Result 1. *Let $\boldsymbol{\Sigma}$ be a general spatial covariance matrix, a multi-resolution decomposition (MRD) computes $\mathbf{B} = MRD(\boldsymbol{\Sigma})$ such that $\boldsymbol{\Sigma} \approx \mathbf{B}\mathbf{B}^\top$, where $\mathbf{B}$ is of the same dimension as $\boldsymbol{\Sigma}$ but possesses some sparse structure that allows efficient computations. For detailed MRD algorithm and the properties of $\mathbf{B}$, see Jurek and Katzfuss (2018).*

The multi-resolution filtering (MRF) algorithm described in Jurek and Katzfuss (2018) is given in Algorithm 1. To employ the MRD technique in the Kalman smoothing algorithm, we propose a new efficient algorithm to generate samples from the smoothing distribution $[\mathbf{x}_t \mid \mathbf{y}_{1:T}]$ for any $t = 1, \ldots, T-1$.

Algorithm 1 Multi-resolution Filter (MRF)

At the initial time $t = 0$, compute $\mathbf{B}_{0|0} := \text{MRD}(\boldsymbol{\Sigma}_{0|0})$.

1: **for** $t = 1, 2, \ldots, T$ **do**

2: **One-step-ahead forecast**: Compute $\boldsymbol{\mu}_{t|t-1} = \mathbf{A}_t\boldsymbol{\mu}_{t-1|t-1}$ and $\mathbf{B}_{t|t-1}^F := \mathbf{A}_t\mathbf{B}_{t-1|t-1}$. Perform a multi-resolution decomposition $\mathbf{B}_{t|t-1} = \text{MRD}(\boldsymbol{\Sigma}_{t|t-1}^F)$, where $\boldsymbol{\Sigma}_{t|t-1}^F = \mathbf{B}_{t|t-1}^F(\mathbf{B}_{t|t-1}^F)^\top + \mathbf{Q}_t$. Then we can obtain $\mathbf{x}_t \mid \mathbf{y}_{1:t-1} \sim \mathcal{N}_{n_\mathcal{G}}(\boldsymbol{\mu}_{t|t-1}, \boldsymbol{\Sigma}_{t|t-1})$, where $\boldsymbol{\Sigma}_{t|t-1} = \mathbf{B}_{t|t-1}\mathbf{B}_{t|t-1}^\top$.

3: **Updating filtering distribution**: Compute $\mathbf{B}_{t|t} := \mathbf{B}_{t|t-1}(\mathbf{L}_t^{-1})^\top$, where $\mathbf{L}_t$ is the lower Cholesky triangle of $\Lambda_t := \mathbf{I}_{n_\mathcal{G}} + \mathbf{B}_{t|t-1}^\top\mathbf{H}_t^\top\mathbf{R}_t^{-1}\mathbf{H}_t\mathbf{B}_{t|t-1}$. Then we can obtain $\mathbf{x}_t \mid \mathbf{y}_{1:t} \sim \mathcal{N}_{n_\mathcal{G}}(\boldsymbol{\mu}_{t|t}, \boldsymbol{\Sigma}_{t|t})$, where $\boldsymbol{\mu}_{t|t} = \boldsymbol{\mu}_{t|t-1} + \mathbf{B}_{t|t}\mathbf{B}_{t|t}^\top$ and $\boldsymbol{\Sigma}_{t|t} = \mathbf{B}_{t|t}\mathbf{B}_{t|t}^\top$

4: **end for**

2.1 Conditional simulation

To generate samples from the posterior distribution $[\mathbf{x}_{1:T} \mid \mathbf{y}_{1:T}]$, one can use the Forward Filtering Backward Sampling (FFBS) technique (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). An alternative is to use the conditional simulation technique developed in Geostatistics (Journel, 1974). This idea was developed independently in other areas such as astrophysics (Hoffman and Ribak, 1991) and time series analysis (Durbin and Koopman, 2002).

To derive this conditional simulation algorithm, we need to compute the mean of the smoothing distribution $\boldsymbol{\mu}_{t|T} = E(\mathbf{x}_t \mid \mathbf{y}_{1:T})$.

RESULT 2. *Using the MRD, we have for $t = 1, \ldots, T-1$,*

$$\boldsymbol{\mu}_{t|T} = \boldsymbol{\mu}_{t|t} + \mathbf{J}_t(\boldsymbol{\mu}_{t+1|T} - \boldsymbol{\mu}_{t+1|t}), \tag{2.1}$$

where $\mathbf{J}_t = \mathbf{B}_{t|t}\mathbf{B}_{t|t}^\top \mathbf{A}_{t+1}^\top \mathbf{B}_{t+1|t}^{-\top}\mathbf{B}_{t+1|t}^{-1}$.

As $\mathbf{B}_{t+1|t}$ is a sparse matrix, solving a sparse linear system $\mathbf{B}_{t+1|t}^{-1}\mathbf{a}$ for a vector $\mathbf{a}$ is computationally efficient, and hence $\boldsymbol{\mu}_{t|T}$ can be computed efficiently. The procedure to generate samples from $[\mathbf{x}_{1:T-1} \mid \mathbf{y}_{1:T}]$ is outlined in Algorithm 2. Let $\mathbf{X}_{1:T-1}^{\mathrm{NS}}$ be a random sample from $[\mathbf{x}_{1:T-1}]$ and $\mathbf{X}_{1:T-1}^{\mathrm{CS}}$ be a random sample from $[\mathbf{x}_{1:T-1} \mid \mathbf{y}_{1:T}]$.

Algorithm 2 Conditional simulation from the multi-resolution smoother (MRS)

At the initial time $t = 0$, compute $\mathbf{B}_{0|0} := \mathrm{MRD}(\boldsymbol{\Sigma}_{0|0})$.

for $t = 1, 2, \ldots, T-1$ **do**

  Non-conditional simulation: Generate a sample $\mathbf{X}_t^{\mathrm{NS}}$ from $[\mathbf{x}_t]$ using Equation (1.2), where only $\mathbf{B}_{0|0}$ and $\mathbf{A}_t$ are needed. Sampling from $\mathbf{w}_t$ is assumed to be computationally efficient, if $\mathbf{w}_t$ exists. Then we generate a sample $\mathbf{Y}_t^{\mathrm{NS}}$ correspondingly.

  Conditional simulation: Compute $\mathbf{X}_t^{\mathrm{CS}} := \mathbf{X}_t^{\mathrm{NS}} + E(\mathbf{x}_t \mid \mathbf{y}_{1:T}) - E(\mathbf{X}_t^{\mathrm{NS}} \mid \mathbf{Y}_{1:T}^{\mathrm{NS}})$, where $E(\mathbf{x}_t \mid \mathbf{y}_{1:T})$ and $E(\mathbf{X}_t^{\mathrm{NS}} \mid \mathbf{Y}_{1:T}^{\mathrm{NS}})$ can be computed efficiently using **Result** 2.

end for

PROPOSITION 1. *By construction, the random vectors $\mathbf{X}_{1:T-1}^{CS}$ possess the following properties.*

1. *$\mathbf{X}_{1:T-1}^{CS}$ has the same marginal distribution as $\mathbf{x}_{1:T-1}$.*
2. *The conditional distribution of $\mathbf{X}_{1:T-1}^{CS}$ given $\mathbf{y}_{1:T}$ has the same distribution as $\mathbf{x}_{1:T-1}$ given $\mathbf{y}_{1:T}$.*

These results are extensions of simple results in Ma et al. (2019); see the proof there for multivariate normal random vectors.

Even with the multi-resolution decomposition technique, it is computationally challenging to directly compute the covariance matrix in the smoothing distribution $[\mathbf{x}_t \mid \mathbf{y}_{1:T}]$ for $t = 1, \ldots, T-1$. In contrast, the capability to efficiently generate random samples from $[\mathbf{x}_t \mid \mathbf{y}_{1:T}]$ could be a fix for developing the multi-resolution smoother. This approach is different from the ensemble Kalman filter/smoother approach.

References

Carter, C. K. and Kohn, R. (1994). "On Gibbs sampling for state space models". *Biometrika*, 81(3):541–553.

Durbin, J. and Koopman, S. J. (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika*, 89(3):603–615.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15(2):183–202.

Hoffman, Y. and Ribak, E. (1991). Constrained realizations of Gaussian fields: A simple algorithm. *The Astrophysical Journal*, 380:L5–L8.

Journel, A. G. (1974). Geostatistics for conditional simulation of ore bodies. *Economic Geology*, 69(5):673–687.

Jurek, M. and Katzfuss, M. (2018). Multi-resolution filters for massive spatio-temporal data. *arXiv preprint arXiv:1810.04200*.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, 112(517):201–214.

Ma, P., Kang, E. L., Braverman, A. J., and Nguyen, H. M. (2019). Spatial statistical downscaling for constructing high-resolution nature runs in global observing system simulation experiments. *Technometrics*, 0(0):1–19.