6 Supplementary Materials

6.1 Proof of Theorem 3.1

Proof. Let $\tilde{m}_{X_t|z_{1:T}} := \sum_{i=1}^{l} w_i^{(t+1)} m_{X_t|\tilde{X}_i, z_{1:t}}$. We then have

$$\left\| m_{X_t|z_{1:T}} - \hat{m}_{X_t|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}} \leq \left\| m_{X_t|z_{1:T}} - \tilde{m}_{X_t|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}} + \left\| \tilde{m}_{X_t|z_{1:T}} - \hat{m}_{X_t|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}}. \tag{16}$$

We consider each of the two terms in equation (16). Let $\Delta m_{X_t|z_{1:T}} := m_{X_t|z_{1:T}} - \tilde{m}_{X_t|z_{1:T}}$. For the first term,

$$\begin{aligned}
& \left\| \Delta m_{X_t|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}}^2 \\
= & \left\| m_{X_t|z_{1:T}} - \sum_{i=1}^{l} w_i^{(t+1)} m_{X_t|\tilde{X}_i, z_{1:t}} \right\|_{\mathcal{H}_\mathcal{X}}^2 \\
= & \sum_{i,j=1}^{l} w_i^{(t+1)} w_j^{(t+1)} \xi_t(\tilde{X}_i, \tilde{X}_j) \\
& -2 \sum_{i=1}^{l} w_i^{(t+1)} \int \xi_t(\tilde{X}_i, x) dP_{X_{t+1}|z_{1:T}}(x) \\
& + \int \xi_t(x, \tilde{x}) dP_{X_{t+1}|z_{1:T}}(x) dP_{X_{t+1}|z_{1:T}}(\tilde{x}) \\
= & \left\langle \Delta m_{X_{t+1}|z_{1:T}} \otimes \Delta m_{X_{t+1}|z_{1:T}}, \xi_t \right\rangle_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{X}} \\
\leq & \left\| \Delta m_{X_{t+1}|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}}^2 \left\| \xi_t \right\|_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{X}}
\end{aligned}$$

Since $\|\xi_t\|_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{X}} < \infty$, the first term decays with $O_p(l^{-2\alpha_{t+1}})$. For the second term, we have

$$\begin{aligned}
& \left\| \tilde{m}_{X_t|z_{1:T}} - \hat{m}_{X_t|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}}^2 \\
= & \left\| \sum_{i=1}^{l} w_i^{(t+1)} (m_{X_t|\tilde{X}_i, z_{1:t}} - \hat{m}_{X_t|\tilde{X}_i, z_{1:t}}) \right\|_{\mathcal{H}_\mathcal{X}}^2 \\
= & \left\| \sum_{i=1}^{l} w_i^{(t+1)} \Delta m_{X_t|\tilde{X}_i, z_{1:t}} \right\|_{\mathcal{H}_\mathcal{X}}^2 \\
= & \sum_{i,j=1}^{l} w_i^{(t+1)} w_j^{(t+1)} \Delta \xi_t(\tilde{X}_i, \tilde{X}_j) \\
= & \left\langle \hat{m}_{X_{t+1}|z_{1:T}} \otimes \hat{m}_{X_{t+1}|z_{1:T}}, \Delta \xi_t \right\rangle_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{X}} \\
\leq & \left\| \hat{m}_{X_{t+1}|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}}^2 \left\| \Delta \xi_t \right\|_{\mathcal{H}_\mathcal{X} \otimes \mathcal{H}_\mathcal{X}}.
\end{aligned}$$

Since $\|\hat{m}_{X_{t+1}|z_{1:T}}\|_{\mathcal{H}_\mathcal{X}} \to \|m_{X_{t+1}|z_{1:T}}\|_{\mathcal{H}_\mathcal{X}} < \infty$, the second term decays with $O_p(l^{-2\beta_t})$. These results lead to the statement $\left\| m_{X_t|z_{1:T}} - \hat{m}_{X_t|z_{1:T}} \right\|_{\mathcal{H}_\mathcal{X}} = O_p(l^{-\alpha_t})$, where $\alpha_t = \min\{\alpha_{t+1}, \beta_t\}$. □

6.2 Experimental Setting & Video: Tracking a Single Object (Experiment 1)

State Space Model Setting: The target's state at time t is described by $\mathbf{x}_t = (x_t, y_t, \dot{x}_t, \dot{y}_t)$ with the object's position (x_t, y_t) and the velocity $(\dot{x}_t, \dot{y}_t)$ in cartesian coordinates $\mathbb{R}^2$. The discretized dynamics is expressed with

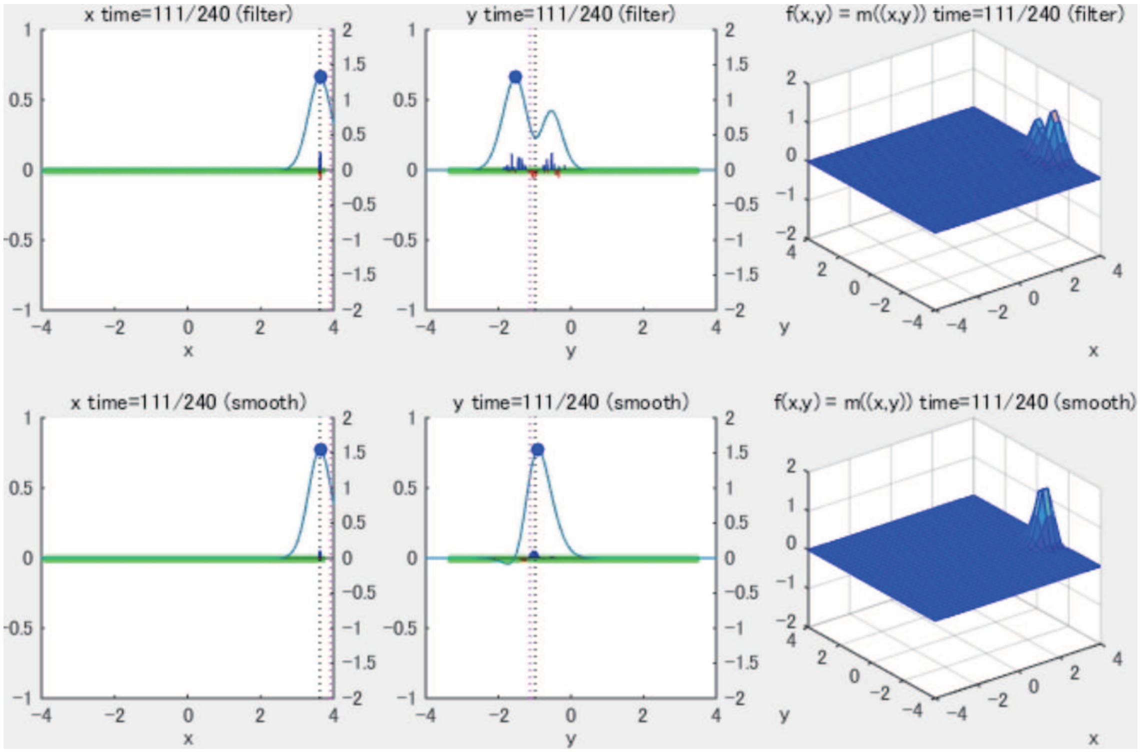

Figure 4: A supplementary video. This animation visualizes the sequential update of kernel means of the nKB-filter [10] and the nKB-smoother (Algorithm 1) for a test sequence $z_{1:240}$ in the clutter problem. The upper three figures show the sequential update of kernel means $m_{X_t|z_{1:t}}$ $(t = 1 : 240)$ of the nKB-filter. The lower three figures show the estimated smoothing kernel means $m_{X_t|z_{1:240}}$ $(t = 1 : 240)$ of the nKB-smoother. For each, the left figure shows the kernel mean projected to state x, the middle figure shows the kernel mean projected to state y, and the right figure both. Each figure visualizes the following. (Left four figures) The black dot vertical line shows the true target's state $(x,\ y)$. The magenta dot vertical line shows the (cluttered) observation $(\tilde{x},\ \tilde{y})$. The kernel mean weights are shown with left vertical axis. The positive (negative) weight values are visualized with blue (red) bars, respectively. The cyan curve shows the estimated kernel mean (estimated RKHS function) $m_P(\cdot) \in \mathcal{H}_\mathcal{X}$ as a function of $(\cdot)$ with right vertical axis. The blue dot in the top of the mountain shows the result of the mode estimation for the target's state (x, y) with the objective function value. From the two middle figures, it can be observed that the filtering estimation is bimodal for uncertainty, but smoothing estimation correctly identifies the state by using the future measurements $z_{112:240}$, so that the blue dot is on the black dot vertical line.

a time-invariant linear equation:

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + \mathbf{q}_t, \quad A := \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{17}$$

where $\mathbf{q}_t$ is discrete Gaussian white process noise having moments

$$\mathbb{E}[\mathbf{q}_t] = \mathbf{0},$$
$$\mathbb{E}[\mathbf{q}_t\mathbf{q}_t^\top] := \begin{pmatrix} \Delta t^3/3 & 0 & \Delta t^2/2 & 0 \\ 0 & \Delta t^3/3 & 0 & \Delta t^2/2 \\ \Delta t^2/2 & 0 & \Delta t & 0 \\ 0 & \Delta t^2/2 & 0 & \Delta t \end{pmatrix} q$$

with $q > 0$. The measurement process for the target is a mixture model:

$$p(\mathbf{z}_t|\mathbf{x}_t) = (1-\rho)N(\mathbf{z}_t|H\mathbf{x}_t, R) + \rho\frac{1}{|S|}, \tag{18}$$

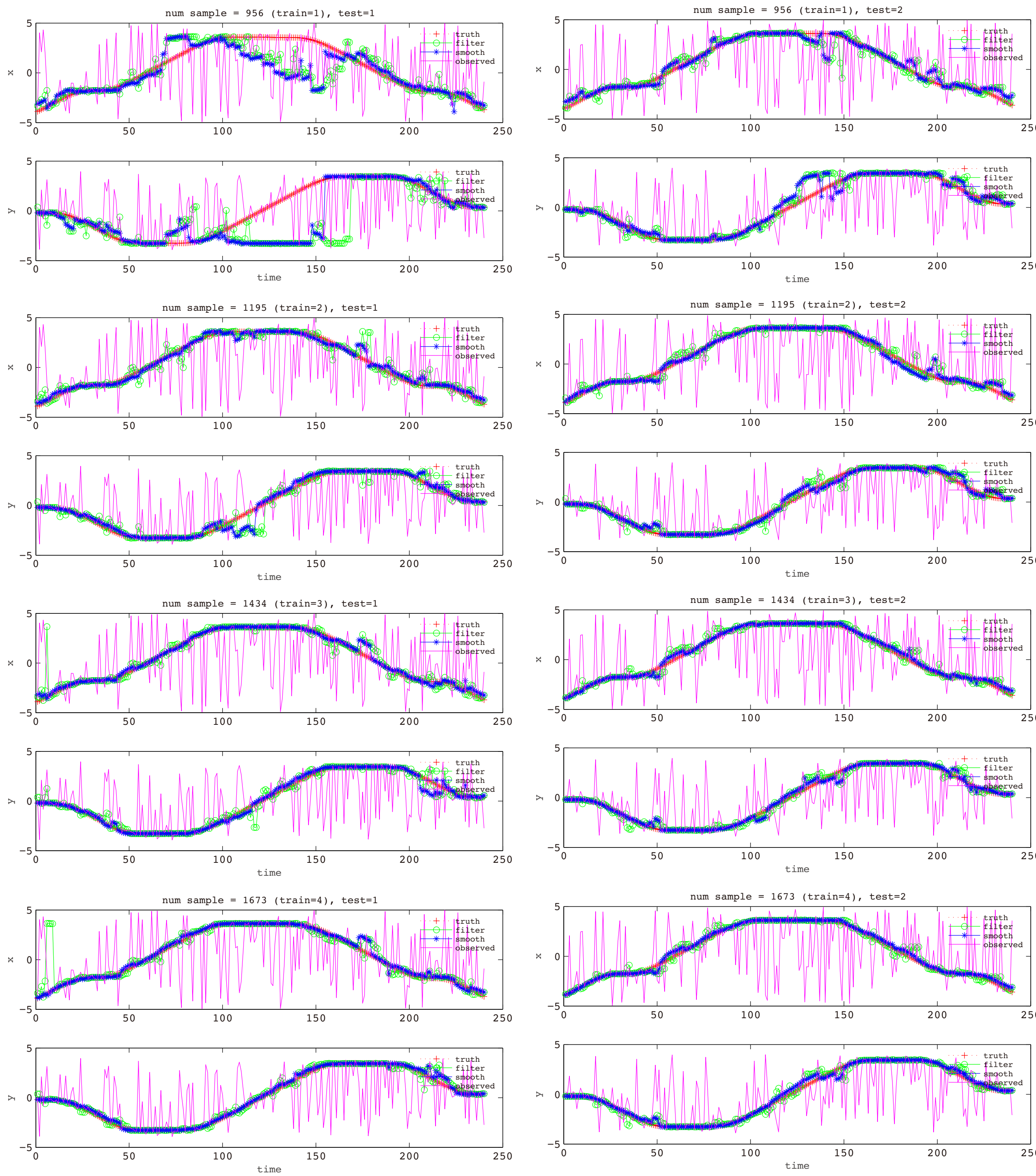

Figure 5: Performance of the nKB-filter and the nKB-smoother in different training and test data on the clutter problem. This figure shows 8 (4×2) experimental results. The upper-left two figures show the performance on the dimension x and y when the training sample size is $n = 956$, respectively. The lower figures show the results when the training sample size is increased to $n = 956, 1195, 1434, 1673$. It is observed that the performance is increased. The right eight figures show results on different test data.

where $1 - \rho$ and ρ are probabilities of measurements from the actual target and clutter, respectively. The measurement from the actual target is a Gaussian $N(\mathbf{z}_t | H\mathbf{x}_t, R)$ with the measurement model matrix H and noise covariance matrix R. The measurement from the clutter is uniform on the area S. We used the same parameter setting as the RBMCDA's demo used, i.e., the size of time step $\Delta t = 0.1$, $q = 0.1$, $\rho = 1/2$, $S = [-5, 5] \times [-4, 4]$, and

$$H = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, R = \begin{pmatrix} 0.05 & 0 \\ 0 & 0.05 \end{pmatrix}.$$

nKB-smoother setting: We used Gaussian kernels $k_{\mathcal{X}}(\mathbf{x}_1, \mathbf{x}_2) = \mathrm{e}^{-||\mathbf{x}_1 - \mathbf{x}_2||^2/2\sigma_{\mathcal{X}}^2}$ and $k_{\mathcal{Z}}(\mathbf{z}_1, \mathbf{z}_2) = \mathrm{e}^{-||\mathbf{z}_1 - \mathbf{z}_2||^2/2\sigma_{\mathcal{Z}}^2}$ for target's states and measurements, respectively, where $\sigma_{\mathcal{X}} = \sigma_{\mathcal{Z}} = 0.1$. We set regularization constants $\epsilon_n = \delta_n = \tilde{\epsilon}_n = \tilde{\delta}_n = 0.001$. Note $\tilde{\epsilon}_n$ and $\tilde{\delta}_n$ are new regularization constants introduced for KB-smoother.

A supplementary video: We present an animation which shows results of the nKB-filter [10] and the nKB-smoother (Algorithm 1) in the clutter problem. Please see a supplementary movie file (.mov). Figure 4 presents a snapshot of the animation at time step $t = 111$.

Supplementary results: Figure 5 shows other results in different training and test data on the clutter problem.

6.3 Marginal Kernel Mean Computation on Tree Graphs

In this section, we present marginal kernel mean computation on general tree graphs by using the nKB-filter and the nKB-smoother, as the extension of state space models.

6.3.1 The nKB-filter & nKB-smoother on N Branch Cases

For ease of understanding, we begin with the two branch case shown in Figure 6 (left). Let $\mathbf{x} := (x_{1:T}, \bar{x}_{t+1:\bar{T}})$ be hidden variables and $\mathbf{z} := (z_{1:T}, \bar{z}_{t+1:\bar{T}})$ be measurement variables. The joint probability density function (pdf) $p(\mathbf{x}, \mathbf{z})$ of Figure 6 (left) is given by[11]

$$p(\mathbf{x}, \mathbf{z}) = \left(\prod_{i=0}^{T-1} p(x_{i+1}|x_i)\right)\left(\prod_{i=1}^{T} p(z_i|x_i)\right) \left(\prod_{i=t}^{\bar{T}-1} p(\bar{x}_{i+1}|\bar{x}_i)\right)\left(\prod_{i=t+1}^{\bar{T}} p(\bar{z}_i|\bar{x}_i)\right),$$

where $p(x_1|x_0) := p(x_1)$ and $\bar{x}_t := x_t$. For ease of presentation, we assume that the transition process $\{p(x_{i+1}|x_i)\}_{i=1}^{T-1}$ and $\{p(\bar{x}_{i+1}|\bar{x}_i)\}_{i=t}^{\bar{T}-1}$ follow the same conditional pdf $p(x'|x)$. We also assume that the measurement process $\{p(z_i|x_i)\}_{i=1}^{T}$ and $\{p(\bar{z}_i|\bar{x}_i)\}_{i=t+1}^{\bar{T}}$ follow the same conditional pdf $p(z|x)$. It is not difficult to extend this to general inhomogenous cases, if there is a training sample for learning each of them. We assume that there are training data $\{\tilde{X}_i, \tilde{X}'_i\}_{i=1}^{l}$ and $\{X_i, Z_i\}_{i=1}^{n}$ for $p(x'|x)$ and $p(z|x)$, respectively.

The objective here is to compute the kernel means $\{m_{X_\tau|\mathbf{z}}\}_{\tau=1}^{T}$ and $\{m_{\bar{X}_\tau|\mathbf{z}}\}_{\tau=t+1}^{\bar{T}}$ of conditional distributions $\{p(x_\tau|\mathbf{z})\}_{\tau=1}^{T}$ and $\{p(\bar{x}_\tau|\mathbf{z})\}_{\tau=t+1}^{\bar{T}}$ given measurements $\mathbf{z}$, respectively. We begin with giving an order to the two branches. Wlog, we set $(x_{t+1:T}, z_{t+1:T}) > (\bar{x}_{t+1:\bar{T}}, \bar{z}_{t+1:\bar{T}})$. We have outputs of the nKB-filter and the nKB-smoother on chain $(x_{1:T}, z_{1:T})$ as

$$\begin{aligned}\hat{m}_{X_t|z_{1:t}} &= \sum_{i=1}^{n} \alpha_i^{(t)} k_{\mathcal{X}}(\cdot, X_i), \quad t = 1, \ldots, T, \\ \hat{m}_{X_t|z_{1:T}} &= \sum_{i=1}^{l} w_i^{(t)} k_{\mathcal{X}}(\cdot, \tilde{X}_i), \quad t = 1, \ldots, T-1.\end{aligned}$$

[11] For simplicity, we omitted illustrations of observable variables $\mathbf{z}$ in Figure 6.

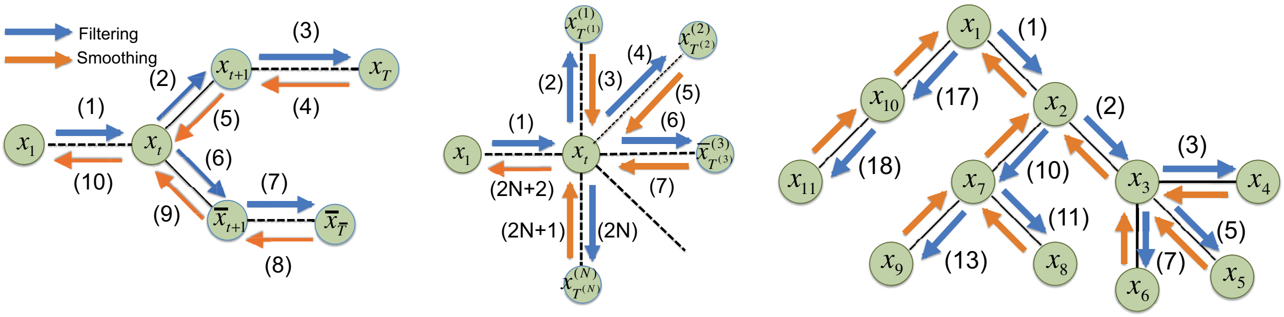

Figure 6: Marginal kernel mean computation on tree graphs using the nKB-filter and the nKB-smoother; (left) the simple two branch case, (middle) general N branch case, and (right) a tree example.

By applying the nonparametric kernel sum rule[12] (Section 2 or Song et al. [25]) to $\hat{m}_{X_t|z_{1:T}}$, we have

$$\hat{m}_{\bar{X}_{t+1}|z_{1:T}} = \hat{\mathcal{U}}_{\bar{X}_{t+1}|X_t}\hat{m}_{X_t|z_{1:T}} = \sum_{i=1}^{l}\eta_i^{(t+1)}k_{\mathcal{X}}(\cdot,\tilde{X}_i'),$$

where $\hat{\mathcal{U}}_{\bar{X}_{t+1}|X_t}$ is the nKSR operator to obtain the estimator $\hat{m}_{\bar{X}_{t+1}|z_{1:T}}$. Next, we apply the KB-filter to the other chain $(\bar{x}_{t+1:\bar{T}}, \bar{z}_{t+1:\bar{T}})$ with the initial belief $\hat{m}_{\bar{X}_{t+1}|z_{1:T}}$, so that the outputs are

$$\hat{m}_{\bar{X}_\tau|z_{1:T},\bar{z}_{t+1:\tau}} = \sum_{i=1}^{n}\bar{\alpha}_i^{(\tau)}k_{\mathcal{X}}(\cdot,X_i), \quad \tau = t+1,\ldots,\bar{T}.$$

Then, we apply the nKB-smoother to the chain $(x_{1:t}, z_{1:t})(\bar{x}_{t+1:\bar{T}}, \bar{z}_{t+1:\bar{T}})$ backward with the initial kernel mean $m_{\bar{X}_{\bar{T}}|\mathbf{z}}$, so that the outputs are

$$\begin{aligned}\hat{m}_{\bar{X}_\tau|\mathbf{z}} &= \sum_{i=1}^{l}\bar{w}_i^{(\tau)}k_{\mathcal{X}}(\cdot,\tilde{X}_i) \quad \tau = t+1,\ldots,\bar{T}-1.\\ \hat{m}_{X_\tau|\mathbf{z}} &= \sum_{i=1}^{l}\bar{w}_i^{(\tau)}k_{\mathcal{X}}(\cdot,\tilde{X}_i) \quad \tau = 1,\ldots,t.\end{aligned}$$

The numbers written in Figure 6 (left) show the order of inference of KB-filter and KB-smoother. By induction, the same applies to the N branch case in Figure 6 (middle). First, give an order to the N branches. Then, apply KB-filter and KB-smoother to one branch by one branch. As an example, Figure 6 (right) shows the order of KB-filter and KB-smoother in a tree graph. Thus, the marginal kernel mean computation on a general tree graph is obtained.

[12]By the Markov property, the conditional pdf $p(\bar{x}_{t+1}|z_{1:T})$ has the sum rule expression:

$$\begin{aligned}p(\bar{x}_{t+1}|z_{1:T}) &= \int p(\tilde{\mathbf{x}}, \tilde{z}_{t+1:\bar{T}}|z_{1:T})\delta(\tilde{x}_{t+1}-\bar{x}_{t+1})d\tilde{z}_{t+1:\bar{T}}d\tilde{\mathbf{x}}\\ &= \int p(\bar{x}_{t+1}|x_t)p(x_t|z_{1:T})dx_t,\end{aligned}$$

where $\delta(\tilde{x}_{t+1}-\bar{x}_{t+1})$ is the dirac's delta function.