DIFFERENTIABLE PARTICLE FILTERS WITH SMOOTHLY JITTERED RESAMPLING

Yichao Li, Wenshuo Wang, Ke Deng and Jun S. Liu

Supplementary Material

The supplementary materials consist of three sections: (1) The validity and computational details of the proposed algorithm; (2) A general guide for selecting the bandwidth parameter r; (3) Empirical evidence for the effectiveness of the proposed method in terms of likelihood estimation.

S1 Details of RRJ

S1.1 Derivation of Equation (5.11)

Let $\varepsilon_j \sim K_{r,j}(x)$, $j = 1, \ldots, d$, independent of each other. Then let

$$I \sim \text{Multinomial}\left(1, (1, \ldots, n), (W_1, \ldots, W_n)\right),$$

be independent of ε and

$$\tilde{X} = X_I + \varepsilon \mid X, W \sim \sum_{i=1}^{n} W_i \kappa_r \left(x - X_i\right). \tag{S1.1}$$

Now we can see that

$$\begin{aligned}
F_1^{W,X,K_r}(x_1) &= P(\tilde{X} \leq x_1 \mid W, X) \\
&= P(X_{I1} + \varepsilon_1 \leq x_1 \mid W, X) \\
&= \sum_{i=1}^{n} W_i P(X_{i1} + \varepsilon_1 \leq x_1 \mid W, X) \\
&= \sum_{i=1}^{n} W_i P(\varepsilon_1 \leq x_1 - X_{i1} \mid W, X) = \sum_{i=1}^{n} W_i K_{r,1}(x_1 - X_{i1}).
\end{aligned}$$

Then for $j > 1$,

$$P(I = i \mid W, X, \tilde{X}_{1:j-1} = x_{1:j-1}) = \frac{W_i \prod_{k=1}^{j-1} \kappa_{r,k}(x_k - X_{ik})}{\sum_{i=1}^{n} W_i \prod_{k=1}^{j-1} \kappa_{r,k}(x_k - X_{ik})},$$

so

$$
\begin{aligned}
&F_j^{W,X,K_r}(x_j; x_{1:j-1}) \\
&= P(\tilde{X}_j \le x_j \mid W, X, \tilde{X}_{1:j-1} = x_{1:j-1}) \text{ (definition)} \\
&= \sum_{i=1}^{n} P(I = i \mid W, X, \tilde{X}_{1:j-1} = x_{1:j-1}) P(\tilde{X}_j \le x_j \mid W, X, \tilde{X}_{1:j-1} = x_{1:j-1}, I = i) \\
&= \sum_{i=1}^{n} \frac{W_i \prod_{k=1}^{j-1} \kappa_{r,k}(x_k - X_{ik})}{\sum_{i=1}^{n} W_i \prod_{k=1}^{j-1} \kappa_{r,k}(x_k - X_{ik})} P(\varepsilon_j \le x_j - X_{ij} \mid W, X, \varepsilon_{1:j-1} = x_{1:j-1} - X_{i,1:j-1}, I = i) \\
&= \sum_{i=1}^{n} \frac{W_i \prod_{k=1}^{j-1} \kappa_{r,k}(x_k - X_{ik})}{\sum_{i=1}^{n} W_i \prod_{k=1}^{j-1} \kappa_{r,k}(x_k - X_{ik})} K_{r,j}(x_j - X_{ij}).
\end{aligned}
$$

S1.2 Gradient

Here, we give calculation for the Gaussian case, where

$$
K_{r,j}(x) = \Phi(x/r), \kappa_{r,j}(x) = \phi(x/r)/r.
$$

Note that the calculation can be done for general kernels as long as it has a differentiable probability density function.

We have $U_1 = F_1^{W,X,K_r}(x_1)$ and $U_j = F_j^{W,X,K_r}(x_j; x_{1:j-1})$, $j > 1$. Differentiate

$$
U_j = \frac{\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r}) \Phi(\frac{\tilde{x}_j - X_{ij}}{r})}{\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r})}
$$

with respect to

1. $W_{i'}$:

Let

$$
\begin{aligned}
A_{ij} &= \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r}), \\
b_{ij} &= \phi(\frac{\tilde{x}_j - X_{ij}}{r}), \\
B_{ij} &= \Phi(\frac{\tilde{x}_j - X_{ij}}{r}).
\end{aligned}
$$

$$
\begin{aligned}
0 = A_{i'j}B_{i'j} & \\
&+ \sum_i W_i A_{ij} \left[\frac{1}{r^2} B_{ij} \left\{ \sum_{k=1}^{j-1} (X_{ik} - \tilde{X}_k) \frac{\partial \tilde{x}_k}{\partial W_{i'}} \right\} + \frac{1}{r} b_{ij} \frac{\partial \tilde{x}_j}{\partial \tilde{W}_{i'}} \right] \\
&\qquad - A_{i'j} \frac{\sum_i W_i A_{ij} B_{ij}}{\sum_i W_i A_{ij}} \\
&\qquad - \frac{1}{r^2} \frac{\sum_i W_i A_{ij} B_{ij} \sum_i W_i A_{ij} (\sum_{k=1}^{j-1} (X_{ik} - \tilde{X}_k) \frac{\partial \tilde{x}_k}{\partial W_{i'}})}{\sum_i W_i A_{ij}}
\end{aligned}
$$

The coefficient of $\frac{\partial \tilde{x}_k}{\partial W_{i'}}$ in the jth equation, when $j = k$:

$\frac{\partial \tilde{x}_k}{\partial W_{i'}} : \frac{1}{r} \sum_i W_i A_{ik} b_{ik}$.

When $j > k$: $\frac{1}{r^2} \sum_i W_i A_{ij} B_{ij} (X_{ik} - \tilde{X}_k) - \frac{1}{r^2} \frac{\sum_i W_i A_{ij} B_{ij} \sum_i W_i A_{ij} (X_{ik} - \tilde{X}_k)}{\sum_i W_i A_{ij}}$

The constant of equation j: $A_{i'j} \frac{\sum_i W_i A_{ij} B_{ij}}{\sum_i W_i A_{ij}} - A_{i'j} B_{i'j}$

2. $U_{j'}$:

$$\delta_{j,j'} \sum_i W_i A_{ij} = \sum_i W_i A_{ij} \left\{ \frac{1}{r^2} B_{ij} \sum_{k=1}^{j-1} (X_{ik} - \tilde{x}_k) \frac{\partial \tilde{x}_k}{\partial U_{j'}} + \frac{1}{r} b_{ij} \frac{\partial \tilde{x}_j}{\partial U_{j'}} \right\}$$
$$- \frac{1}{r^2} \frac{\sum_i W_i A_{ij} B_{ij} \sum_i W_i \left\{ A_{ij} \sum_{k=1}^{j-1} (X_{ik} - \tilde{x}_k) \frac{\partial \tilde{x}_k}{\partial U_{j'}} \right\}}{\sum_i W_i A_{ij}}$$

The coefficient of $\frac{\partial \tilde{x}_k}{\partial U_{j'}}$ in the jth equation, when $j = k$:

$\frac{1}{r} \sum_i W_i A_{ij} b_{ij}$.

When $j > k$: $\frac{1}{r^2} \sum_i W_i A_{ij} B_{ij} (X_{ik} - \tilde{X}_k) - \frac{1}{r^2} \frac{\sum_i W_i A_{ij} B_{ij} \sum_i W_i A_{ij} (X_{ik} - \tilde{X}_k)}{\sum_i W_i A_{ij}}$

The constant of equation j: $\delta_{j,j'} \sum_i W_i A_{ij}$

3. $X_{i'k'}$:

$$0 = \frac{\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r}) \Phi(\frac{\tilde{x}_j - X_{ij}}{r}) \left\{ \sum_{k=1}^{j-1} \frac{\sum_{k=1}^{j-1} (X_{ik} - \tilde{x}_k) \left(\frac{\partial \tilde{x}_k}{\partial X_{i'k'}} - \delta_{(i,k),(i',k')} \right)}{r^2} \right\}}{\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r})}$$
$$+ \frac{\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r}) \phi(\frac{\tilde{x}_j - X_{ij}}{r}) \frac{\frac{\partial \tilde{x}_j}{\partial X_{i'k'}} - \delta_{(i,j),(i',k')}}{r}}{\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r})}$$
$$- \frac{\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r}) \Phi(\frac{\tilde{x}_j - X_{ij}}{r}) \left\{ \sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r}) \sum_{k=1}^{j-1} \frac{\sum_{k=1}^{j-1} (X_{ik} - \tilde{x}_k) \left(\frac{\partial \tilde{x}_k}{\partial X_{i'k'}} - \delta_{(i,k),(i',k')} \right)}{r^2} \right\}}{\left(\sum_i W_i \prod_{k=1}^{j-1} \phi(\frac{\tilde{x}_k - X_{ik}}{r}) \right)^2}$$

$$0 = \sum_i W_i A_{ij} B_{ij} \left\{ \sum_{k=1}^{j-1} \frac{(X_{ik} - \tilde{x}_k)\left(\frac{\partial \tilde{x}_k}{\partial X_{i'k'}} - \delta_{(i,k),(i',k')}\right)}{r^2} \right\}$$
$$+ \sum_i W_i A_{ij} b_{ij} \frac{\frac{\partial \tilde{x}_j}{\partial X_{i'k'}} - \delta_{(i,j),(i',k')}}{r}$$
$$- \frac{\sum_i W_i A_{ij} B_{ij} \left\{ \sum_i W_i A_{ij} \sum_{k=1}^{j-1} \frac{(X_{ik} - \tilde{x}_k)\left(\frac{\partial \tilde{x}_k}{\partial X_{i'k'}} - \delta_{(i,k),(i',k')}\right)}{r^2} \right\}}{\sum_i W_i A_{ij}}$$

The constant of equation j: $\frac{1}{r^2} W_{i'} A_{i'j} B_{i'j} \mathbb{I}(k' < j)(X_{i'k'} - \tilde{x}_{k'}) + \frac{1}{r} W_{i'} \delta_{jk'} - \frac{1}{r^2} \frac{\sum_i W_i A_{ij} B_{ij}}{\sum_i W_i A_{ij}} W_{i'} A_{i'j} \mathbb{I}(k' < j)(X_{i'k'} - \tilde{x}_{k'})$

S2 Proof of Theorem 1

Proof of Theorem 1. In state space models and Algorithm 2, we have

$$X_i^{(t)} \sim g_t\left(\cdot \mid \tilde{X}_i^{(1:t-1)}\right) \tag{S2.2}$$

$$W_i^{(t)} = f_t(y^{(t)} \mid X_i^{(t)}) \tag{S2.3}$$

$$\tilde{X}_i^{(t)} = X_{I_i^t}^{(t)} + \varepsilon_i^{(t)}, \tag{S2.4}$$

where $I_i^t \mid X^{(t)} \sim \text{Multinomial}(1, 1:n, W^{(t)})$. We let $L_0(\theta) = 1$,

$$L_t(\theta) = \int_{\mathcal{X}} \prod_{s=1}^{t} \left\{ g_t(x^{(s)} \mid x^{(s-1)}; \theta) f_t(y^{(s)} \mid x^{(s)}; \theta) \right\} dx^{(1:t)}$$

be the likelihood of the first t steps. We prove by induction for (5.12) and

$$\left| \mathrm{E}\left\{ \frac{\sum_{j=1}^{n} W_j^{(t)} \phi(X_j^{(t)})}{nL_t(\theta)/L_{t-1}(\theta)} \right\} - \int \pi_t(x^{(1:t)}) \phi(x^{(t)}) dx^{(1:t)} \right| = O(1/\sqrt{n}). \tag{S2.5}$$

For $t = 1$, (S2.5) is 0.

$$\begin{aligned}
&\left[\mathrm{E}\left\{ \frac{\sum_{i=1}^{n} W_i^{(1)} \phi(X_i^{(1)})}{\sum_{i=1}^{n} W_i^{(1)}} - \int_{\mathcal{X}} \pi_1(x^{(1)}) \phi(x^{(1)}) dx^{(1)} \right\} \right]^2 \\
&= \left[\mathrm{E}\left\{ \frac{\sum_{i=1}^{n} W_i^{(1)} \phi(X_i^{(1)})}{\sum_{i=1}^{n} W_i^{(1)}} - \frac{\sum_{i=1}^{n} W_i^{(1)} \phi(X_i^{(1)})}{nL_1(\theta)} \right\} \right]^2 \\
&= \left[\mathrm{E}\left\{ \frac{\sum_{i=1}^{n} W_i^{(1)} \phi(X_i^{(1)})}{\sum_{i=1}^{n} W_{ik}^{(1)}} \left(1 - \frac{\sum_{i=1}^{n} W_i^{(1)}}{nL_1(\theta)} \right) \right\} \right]^2 \\
&\leq \mathrm{E}\left\{ \frac{\sum_{i=1}^{n} W_i^{(1)} \phi(X_i^{(1)})}{\sum_{i=1}^{n} W_{ik}^{(1)}} \left(1 - \frac{\sum_{i=1}^{n} W_i^{(1)}}{nL_1(\theta)} \right) \right\}^2 \\
&\leq \mathrm{E}\left\{ \left(1 - \frac{\sum_{i=1}^{n} W_i^{(1)}}{nL_1(\theta)} \right)^2 \right\} \text{ (because } |\phi| \leq 1\text{)}, \\
&= \mathrm{Var}\left(\frac{\sum_{i=1}^{n} W_i^{(1)}}{nL_1(\theta)} \right) = \Omega(n^{-1})
\end{aligned} \tag{S2.6}$$

Moving on to later steps,

$$
\begin{aligned}
&\mathrm{E}\left(\frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{n}\right)\\
&= \mathrm{E}\left[\mathrm{E}\left\{\mathrm{E}\left(\frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{n} \mid \tilde{X}_i^{(1:t-1)}, X^{(1:t-1)}, W^{(t-1)}\right) \mid X^{(1:t-1)}, W^{(t-1)}\right\}\right]\\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left\{\mathrm{E}\left[\int_{\mathcal{X}} f_t(y^{(t)} \mid x) g_t(x \mid \tilde{X}_i^{(t-1)})\phi(x)dx \mid X^{(1:t-1)}, W^{(t-1)}\right]\right\}\\
&= \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left\{\sum_{j=1}^{n} \frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}} \mathrm{E}\left[\int_{\mathcal{X}} f_t(y^{(t)} \mid x) g_t(x \mid X_j^{(t-1)} + \varepsilon_j^{(t-1)})\phi(x)dx \mid X_j^{(1:t-1)}\right]\right\}
\end{aligned}
$$

Note that

$$
\tilde{\phi}_t(\cdot) = \int_{\mathcal{X}} f_t(y^{(t)} \mid x) g_t(x \mid \cdot)\phi(x)dx
$$

are bounded and ML-Lipschitz since $|\phi| \leq 1$, $|f_t| \leq M$ and $g_t(x \mid \cdot)$ is L-Lipschitz by assumption. So

$$
\begin{aligned}
&\left|\mathrm{E}\left(\frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{n}\right) - \sum_{j=1}^{n} \mathrm{E}\left\{\frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}}\tilde{\phi}_t(X_j^{(t-1)})\right\}\right|\\
&= \left|\mathrm{E}\left\{\sum_{j=1}^{n} \frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}} \mathrm{E}\left[\tilde{\phi}_t(X_j^{(t-1)} + \varepsilon_j^{(t-1)}) - \tilde{\phi}_t(X_j^{(t-1)}) \mid X_j^{(1:t-1)}\right]\right\}\right|\\
&\leq \mathrm{E}\left\{\sum_{j=1}^{n} \frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}} \mathrm{E}\left[\left|\tilde{\phi}_t(X_j^{(t-1)} + \varepsilon_j^{(t-1)}) - \tilde{\phi}_t(X_j^{(t-1)})\right| \mid X_j^{(1:t-1)}\right]\right\} \qquad \text{(S2.7)}\\
&\leq \mathrm{E}\left\{\sum_{j=1}^{n} \frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}} ML\,\mathrm{E}(\|\varepsilon_j^{(t-1)}\|)\right\}\\
&\leq \mathrm{E}\left\{\sum_{j=1}^{n} \frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}} ML\sqrt{C}dr\right\} = ML\sqrt{C}dr = O(1/\sqrt{n}).
\end{aligned}
$$

On the other hand,

$$\left|\sum_{j=1}^{n} \mathrm{E}\left\{\frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}}\tilde{\phi}_t(X_j^{(t-1)})\right\} - \frac{L_t(\theta)}{L_{t-1}(\theta)}\int \pi_t(x^{(1:t)})\phi(x^{(t)})dx^{(1:t)}\right|$$
$$= \left|\sum_{j=1}^{n} \mathrm{E}\left\{\frac{W_j^{(t-1)}}{\sum_{k=1}^{n} W_k^{(t-1)}}\tilde{\phi}_t(X_j^{(t-1)})\right\} - \int \pi_{t-1}(x^{(1:t-1)})\tilde{\phi}_t(x^{(t-1)})dx^{(1:t-1)}\right| = O(1/\sqrt{n}) \tag{S2.8}$$

by induction hypothesis. Adding equations (S2.7) and (S2.8), we have

$$\left|\mathrm{E}\left\{\frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{nL_t(\theta)/L_{t-1}(\theta)}\right\} - \int \pi_t(x^{(1:t)})\phi(x^{(t)})dx^{(1:t)}\right| = O(1/\sqrt{n}). \tag{S2.9}$$

Now we analyze the difference between $(L_t(\theta)/L_{t-1}(\theta))$-normalized estimate and weight-normalized estimate.

$$\begin{aligned}
&\left[\mathrm{E}\left\{\frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{nL_t(\theta)/L_{t-1}(\theta)} - \frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{\sum_{i=1}^{n} W_i^{(t)}}\right\}\right]^2 \\
&= \left[\mathrm{E}\left\{\frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{\sum_{i=1}^{n} W_i^{(t)}}\left(1 - \frac{\sum_{i}^{n} W_i^{(t)}}{nL_t(\theta)/L_{t-1}(\theta)}\right)\right\}\right]^2 \\
&= \mathrm{E}\left\{\frac{\sum_{i=1}^{n} W_i^{(t)}\phi(X_i^{(t)})}{\sum_{i=1}^{n} W_i^{(t)}}\left(1 - \frac{\sum_{i}^{n} W_i^{(t)}}{nL_t(\theta)/L_{t-1}(\theta)}\right)\right\}^2 \\
&\leq \mathrm{E}\left\{1 - \frac{\sum_{i=1}^{n} W_i^{(t)}}{nL_t(\theta)/L_{t-1}(\theta)}\right\}^2 \\
&= \left[\mathrm{E}\left\{1 - \frac{\sum_{i=1}^{n} W_i^{(t)}}{nL_t(\theta)/L_{t-1}(\theta)}\right\}\right]^2 + \mathrm{Var}\left\{\frac{\sum_{i=1}^{n} W_i^{(t)}}{nL_t(\theta)/L_{t-1}(\theta)}\right\}
\end{aligned}$$

The first term is $O(1/n)$ by plugging in $\phi = 1$ in (S2.9), and the second term is

$O(1/n)$ by assumption (i) since $W_i^{(t)}$ is bounded. In summary, we have

$$\left| \mathrm{E}\left\{ \frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{\sum_{j=1}^n W_j^{(t)}} \right\} - \int \pi_t(x^{(1:t)}) \phi(x^{(t)}) dx^{(1:t)} \right| = O(1/\sqrt{n}).$$

As for the variance (5.11), let $w_t = \mathrm{E}[W_i^{(t)}]$.

$$\begin{aligned}
&\mathrm{Var}\left\{ \frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{\sum_{j=1}^n W_j^{(t)}} \right\} \\
&= \mathrm{Var}\left\{ \frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{n w_t} + \frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{\sum_{j=1}^n W_j^{(t)}} \left(1 - \frac{\sum_{j=1}^n W_j^{(t)}}{n w_t} \right) \right\} \\
&\leq 2\,\mathrm{Var}\left\{ \frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{n w_t} \right\} + 2\,\mathrm{Var}\left\{ \frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{\sum_{j=1}^n W_j^{(t)}} \left(1 - \frac{\sum_{j=1}^n W_j^{(t)}}{n w_t} \right) \right\}.
\end{aligned}$$

The first term is obviously $O(1/n)$ by boundedness (note that w_t is a constant). For the second term,

$$\begin{aligned}
&\mathrm{Var}\left\{ \frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{\sum_{j=1}^n W_j^{(t)}} \left(1 - \frac{\sum_{j=1}^n W_j^{(t)}}{n w_t} \right) \right\} \\
&\leq \mathrm{E}\left\{ \left(\frac{\sum_{j=1}^n W_j^{(t)} \phi(X_j^{(t)})}{\sum_{j=1}^n W_j^{(t)}} \right)^2 \left(1 - \frac{\sum_{j=1}^n W_j^{(t)}}{n w_t} \right)^2 \right\} \\
&\leq \mathrm{E}\left\{ \left(1 - \frac{\sum_{j=1}^n W_j^{(t)}}{n w_t} \right)^2 \right\} \\
&= \mathrm{Var}\left(\frac{\sum_{j=1}^n W_j^{(t)}}{n w_t} \right) = O(1/n).
\end{aligned}$$

Now we prove (5.12). Note that

$$\ell(\theta) = \sum_{t=1}^{T} \log \frac{L_t(\theta)}{L_{t-1}(\theta)},$$
$$\hat{\ell}(\theta) = \sum_{t=1}^{T} \log \left(\frac{1}{n} \sum_{i=1}^{n} W_i^{(t)} \right).$$

Because T is fixed, it is sufficient to show that for all t,

$$\mathrm{E}\left[\left\{ \log \left(\frac{1}{n} \sum_{i=1}^{n} W_i^{(t)} \right) - \log \frac{L_t(\theta)}{L_{t-1}(\theta)} \right\}^2 \right] = O(1/n).$$

Since $\frac{1}{n}\sum_{i=1}^{n} W_i^{(t)}$ is bounded below by $\underline{e} > 0$, in the region $[\min(\underline{e}, L_t(\theta)/L_{t-1}(\theta)), \infty)$, log is Lipschtiz and

$$\left| \log \left(\frac{1}{n} \sum_{i=1}^{n} W_i^{(t)} \right) - \log \frac{L_t(\theta)}{L_{t-1}(\theta)} \right| \leq \frac{\left| \frac{1}{n}\sum_{i=1}^{n} W_i^{(t)} - \frac{L_t(\theta)}{L_{t-1}(\theta)} \right|}{\min(\underline{e}, L_t(\theta)/L_{t-1}(\theta))}.$$

Square both sides and take expectation, and we notice that the right side would be $O(1/n)$ by taking $\phi = 1$ in (S2.5). $\square$

S3 Likelihood Experiments

We conduct experiments to support our theoretical analysis of likelihood using synthetic data generated from the linear state space model from Section 6.2. In this case, the marginal likelihood $p(y^{(1:T)})$ can be computed exactly in closed form using the Kalman filter, allowing us to directly compare the output likelihood to the true likelihood. In Figure 1, we plot the log of the sample average of the likelihood estimates from 50 independent experiments, as a function of ρ in equation (6.13), while the observation $y^{(1:T)}$ is generated with $\rho = 0$ and fixed across the experiments. It can be seen that with the kernel bandwidth $r = 0.1$, RRJ has no significantly higher bias compared to multinomial resampling ($r = 0$). In Figure 2, we zoom in on cases with $\rho = 0$ and plot the same log-likelihood with respect to a continuum of r. We can see that RRJ has no significantly higher bias compared to multinomial resampling for a reasonably wide range of r.

S4 Further Details of VRNN Experiments

The dimensions of Z_t and R_t are 10 and 5, respectively. The dimension of τ_θ is 32. For the training of Ensemble Transform, the regularization parameter ϵ is chosen as 0.8. The Sinkhorn iteration number is 500 and the convergence threshold is 0.001.

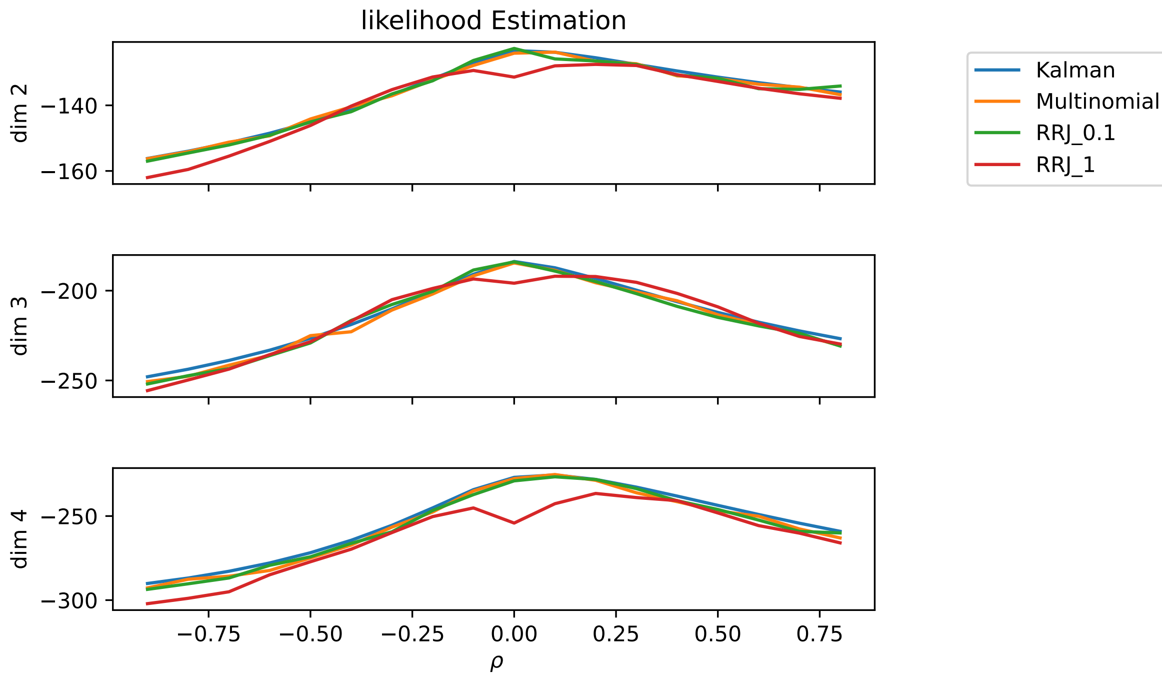

Figure 1: Likelihood estimation. The four lines are the log-likelihood computed via Kalman filter, SMC with multinomial resampling, RRJ with kernel bandwidth $r = 0.1$, and RRJ with $r = 1$, respectively.

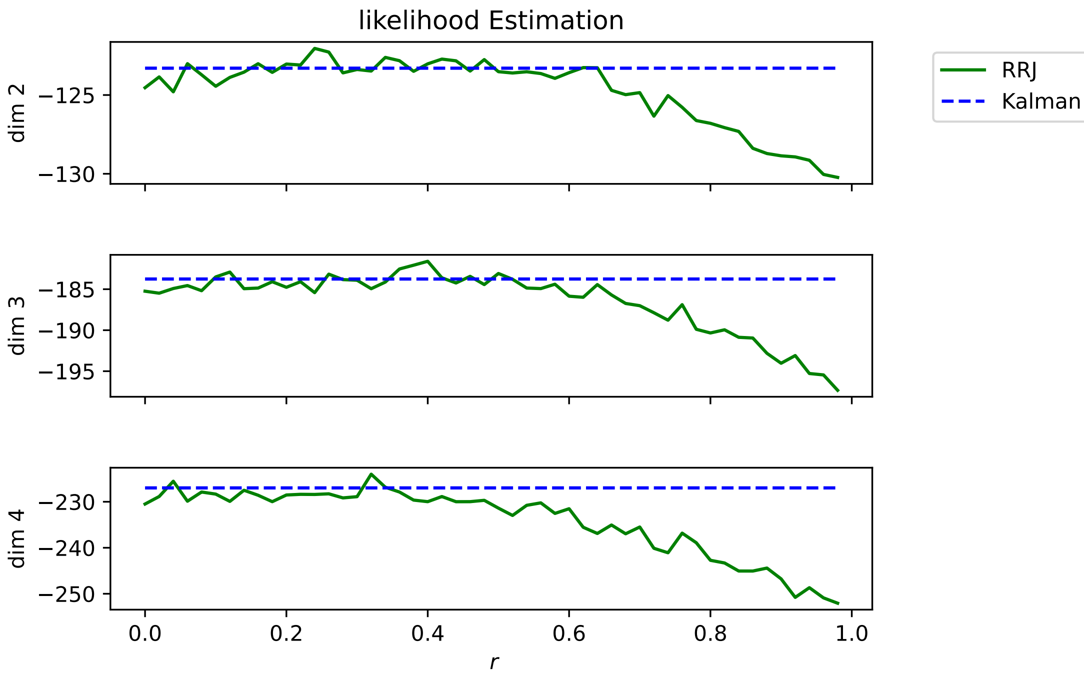

Figure 2: Likelihood estimation with $\rho = 0$. The dashed lines are the log-likelihood computed via Kalman filter. The green curves correspond to those estimated by RRJ for kernel bandwidth r ranging from 0 to 1. Note that $r = 0$ reduces to multinomial resampling.