A Proof of validity for Recycled ESS

We give the complete proof of validity of Recycled ESS here. The Markov chain at step i has state space $\{\mathbf{x}_1^{(i)}, ..., \mathbf{x}_M^{(i)}\}$. We prove that it converges to the stationary distribution with each element identically distributed according to the target distribution p^*.

Lemma A.1. *Let T_j correspond to the transition operator from $\mathbf{x}_1^{(i-1)} \rightarrow \hat{\mathbf{x}}_j^{(i)}$ (the vector returned after the first j applications of line 9 of Algorithm 1). Then T_j is invariant to p^*.*

Proof. Let $\{\theta_{j,k}\},\ k = 1, 2, \ldots K_j$, be the sequence of angles sampled during T_j. The joint distribution over the current state $\mathbf{x}_1^{(i-1)}$ and random variables $y, \boldsymbol{\nu}, \{\theta_{j,k}\}$ generated to transition to $\hat{\mathbf{x}}_j^{(i)}$ is

$$\begin{aligned}
&p(\mathbf{x}_1^{(i-1)}, y, \boldsymbol{\nu}, \{\theta_{j,k}\}) \\
&= p^*(\mathbf{x}_1^{(i-1)}) \cdot p(y|\mathbf{x}_1^{(i-1)}) \cdot p(\boldsymbol{\nu}) \cdot p(\{\theta_{j,k}\}|\mathbf{x}_1^{(i-1)}), y, \boldsymbol{\nu}) \\
&= \frac{1}{Z}\mathcal{N}(\mathbf{x}_1^{(i-1)}; \mathbf{0}, \boldsymbol{\Sigma}) \cdot \mathcal{N}(\boldsymbol{\nu}; \mathbf{0}, \boldsymbol{\Sigma}) \cdot p(\{\theta_{j,k}\}|\mathbf{x}_1^{(i-1)}), y, \boldsymbol{\nu})
\end{aligned}$$

where $p(y|\mathbf{x}_1^{(i-1)}) = 1/\mathcal{L}(\mathbf{x}_1^{(i-1)})$. We show that the algorithm is reversible i.e.

$$p(\mathbf{x}_1^{(i-1)}, y, \boldsymbol{\nu}, \{\theta_{j,k}\}) = p(\hat{\mathbf{x}}_j^{(i)}, y, \hat{\boldsymbol{\nu}}, \{\hat{\theta}_{j,k}\}) \tag{5}$$

where

$$\begin{aligned}
\hat{\boldsymbol{\nu}} &= \boldsymbol{\nu}\cos(\theta_{j,K}) - \mathbf{x}_1^{(i-1)}\sin(\theta_{j,K}) \\
\hat{\theta}_{j,k} &= \begin{cases} \theta_{j,k} - \theta_{j,K} & \text{if } k < K \\ -\theta_{j,K} & \text{if } k = K. \end{cases}
\end{aligned}$$

To prove Equation (5), we first show that:

$$p(\{\theta_{j,k}\}|\mathbf{x}_1^{(i-1)}, y, \boldsymbol{\nu}) = p(\{\hat{\theta}_{j,k}\}|\hat{\mathbf{x}}_j^{(i)}, y, \hat{\boldsymbol{\nu}}) \tag{6}$$

The argument is as follows: probability of the first angle $\theta_{j,1}$ is always $1/2\pi$. The intermediate angles were drawn with probabilities $1/(\theta_{j,k}^{max} - \theta_{j,k}^{min})$ where $(\theta_{j,k}^{min}, \theta_{j,k}^{max})$ denotes the angle bracket for $\theta_{j,k}$. Whenever the bracket was shrunk, it was done so that $\hat{\mathbf{x}}_j^{(i)}$ remained selectable. Now lets consider the reverse transitions starting from $\hat{\mathbf{x}}_j^{(i)}$. The reverse transitions begin by drawing the same threshold y and then make the same intermediate proposals. This involves making the same shrinking decisions as forward transitions. Since same angle brackets $(\theta_{j,k}^{\hat{min}}, \theta_{j,k}^{\hat{max}})$ are sampled, the probabilities for drawing angles in forward and reverse transitions remains the same.

Additionally, we use the fact that

$$\mathcal{N}(\mathbf{x}_1^{(i-1)}; \mathbf{0}, \boldsymbol{\Sigma}) \cdot \mathcal{N}(\boldsymbol{\nu}; \mathbf{0}, \boldsymbol{\Sigma}) = \mathcal{N}(\hat{\mathbf{x}}_j^{(i)}; \mathbf{0}, \boldsymbol{\Sigma}) \cdot \mathcal{N}(\hat{\boldsymbol{\nu}}; \mathbf{0}, \boldsymbol{\Sigma}) \tag{7}$$

as $\hat{\mathbf{x}}_j^{(i)}$ and $\hat{\boldsymbol{\nu}}$ are simply obtained by rotations of $\mathbf{x}_1^{(i-1)}$ and $\boldsymbol{\nu}$ by $\hat{\theta}_{j,K}$. This combined with the result in equation(6) proves equation(5) and shows that T_j is invariant to p^*. $\square$