One way of attempting to improve the performance of the RBPF is to form a better importance sampling distribution. In this section, we seek to form a better importance sampling distribution for $\pi(\lambda_t | y_{1:t}, \lambda_{0:t-1})$.

For optimality, we want $\pi(\lambda_t | y_{1:t}\lambda_{0:t-1}^{(i)}) \approx p(\lambda_t | y_{1:t}, \lambda_{0:t-1}^{(i)})$.

It is hard to formulate an analytical form for this density directly. To get around this, we can instead form the analytical joint distribution for $p(y_t, \lambda_t | y_{1:t-1}, \lambda_{0:t-1}^{(i)})$. The complete derivation is given in the appendix, with the key steps highlighted below.

We first augment the probability density given above with $x_{0:t-1}$. This allows us to use results from the previously-performed Kalman Filter step.

$$
\begin{aligned}
& p(y_t.\lambda_t, x_{0:t-1} | y_{1:t-1}, \lambda_{0:t-1}^{(i)}) \\
\propto\ & p(y_t | y_{1:t-1}, \lambda_{0:t}^{(i)}, x_{0:t-1}) p(x_{0:t-1} | y_{1:t-1}, \lambda_{0:t}^{(i)}) p(\lambda_t^{(i)} | \lambda_{0:t-1}^{(i)}) \\
\approx\ & p(y_t | y_{1:t-1}, \lambda_{0:t}^{(i)}, x_{0:t-1}) p(x_{0:t-1} | y_{1:t-1}, \lambda_{0:t-1}^{(i)}) p(\lambda_t | \lambda_{0:t-1}) \\
=\ & \mathcal{N}(y_t | \mathbf{CA}\mathbf{x}_{t-1}, \mathbf{ee}^T) \mathcal{N}(\mathbf{x}_{0:t-1} | \mu_k^{(i)}, \mathbf{\Sigma}_k^{(i)}) S_{\alpha/2}(1, 1, 0) \\
=\ & \mathcal{N}\left(\begin{bmatrix} y_t \\ \mathbf{x}_{t-1} \end{bmatrix} \Big| \begin{bmatrix} \mathbf{CA}\mu_{k-1}^{(i)} \\ \mu_{k-1}^{(i)} \end{bmatrix}, \begin{bmatrix} \mathbf{ee}^T + (\mathbf{CA})\mathbf{\Sigma}_k^{(i)}(\mathbf{CA})^T & \mathbf{CA}\mathbf{\Sigma}_k^{(i)} \\ \mathbf{\Sigma}_k^{(i)}(\mathbf{CA})^T & \mathbf{\Sigma}_k^{(i)} \end{bmatrix} \right) S_{\alpha/2}(1, 1, 0)
\end{aligned}
$$

We can then obtain the required density $p(y_t, \lambda_t^{(i)} | y_{1:t-1}, \lambda_{0:t-1})$ by marginalising out $x_{0:t-1}$:

$$
\begin{aligned}
p(y_k, \lambda_k | y_{1:k-1}, \lambda_{0:k-1}^{(i)}) &= \int p(y_k, \lambda_k, x_{0:k-1} | y_{1:k-1}, \lambda_{0:k-1}^{(i)}) dx_{0:k-1} \\
&= \mathcal{N}(y_k | \mathbf{CA}\mu_k^{(i)}, (\mathbf{Cb})(\mathbf{Cb})^T \lambda_k + d^2 + (\mathbf{CA})\mathbf{\Sigma}_k^{(i)}(\mathbf{CA})^T) S_{\alpha/2}(\lambda_t | 1, 1, 0) \\
&= \mathcal{N}(y_k | \mathbf{CA}\mu_k^{(i)}, \sigma_\lambda \lambda_k + \mu_\lambda) S_{\alpha/2}(\lambda_t | 1, 1, 0) \quad (1)
\end{aligned}
$$

where:

$$
\begin{aligned}
\sigma_\lambda &= (\mathbf{Cb})(\mathbf{Cb})^T \\
\mu_\lambda &= d^2 + (\mathbf{CA})\mathbf{\Sigma}_k^{(i)}(\mathbf{CA})^T
\end{aligned}
$$

By fixing the value of y_k in Equation 1, we can obtain the desired distribution $p(\lambda_k | y_{1:k}, \lambda_{0:k-1})$.

This formulation of the joint distribution also gives us an alternative interpretation for the sampling density $p(\lambda_t | y_{1:t}, \lambda_{0:t-1})$.

We can reinterpret this as finding the posterior density of λ_k given the observations $y_t - \mathbf{CA}\mu_k$. The prior on λ_k is the α-stable proposal distribution $S_{\alpha/2}(\lambda_k|1,1,0)$ while the likelihood is the unknown variance of a normal distribution ($\mathcal{N}(y_t|\mathbf{CA}\mu_k, \sigma_\lambda\lambda_t + \mu_\lambda)$).

Rejection Sampling

In general, there is no closed form expression for this distribution. However, we can utilise the fact that we can draw samples from the prior distribution easily using the Chamber-Mallow-Stuck method [**chambers1976method**]. This can be used as a suitable proposal function for rejection sampling.

We adapt the methods of [**godsill1999bayesian**] to draw samples from this distribution.

The target distribution for the rejection sampling scheme is given by:

$$
\begin{aligned}
f(y_t, \lambda_t|y_{1:t-1}, \lambda_{0:t-1}^{(i)}) &= \mathcal{N}(y_t|\mathbf{CA}\mu_k^{(i)}, \sigma_\lambda\lambda_t + \mu_\lambda)S_{\alpha/2}(\lambda_t|1,1,0) \\
&= \mathcal{N}(y_t - \mathbf{CA}\mu_k^{(i)}|, \sigma_\lambda\lambda_t + \mu_\lambda)S_{\alpha/2}(\lambda_t|1,1,0)
\end{aligned}
$$

Using a proposal distribution $g(\lambda_t|y_{1:t}, \lambda_{0:t-1}) = S_{\alpha/2}(1,1,0)$, we can use the likelihood as a valid rejection function as it is bounded from the above:

$$
\begin{aligned}
\mathcal{N}(y_t - \mathbf{CA}\mu_k^{(i)}|0, \sigma_\lambda\lambda_t + \mu_\lambda) &\leq \frac{f(y_t, \lambda_t|y_{1:t-1}, \lambda_{0:t-1})}{g(y_t, \lambda_t|y_{1:t-1}, \lambda_{0:t-1})} \\
&= M
\end{aligned}
$$

$$
M = \begin{cases} \frac{1}{\sqrt{2\pi(y_t - \mathbf{CA}\mu_k^{(i)})^2}}\exp(-0.5) & (y_t - \mathbf{CA}\mu_k^{(i)})^2 \geq \mu_\lambda \\ \frac{1}{\sqrt{2\pi d^2}}\exp\left(-\frac{(y_t - \mathbf{CA}\mu_k^{(i)})^2}{2d^2}\right) & (y_t - \mathbf{CA}\mu_k^{(i)})^2 \leq \mu_\lambda \end{cases} \tag{2}
$$

This gives a suitable rejection sampler as:

1. Draw $\lambda_k \sim S_{\alpha_2}(1,1,0)$

2. Draw $u \sim U(0, M)$

3. If $u \geq \mathcal{N}\left(y_t - \mathbf{CA}\mu_k^{(i)}|0, \sigma_\lambda\lambda_k + \mu_\lambda\right)$, reject λ_t and go to Step (1)

Improvements on the rejection sampler:

The rejection sampler suffers from poor acceptance rates when $y_k - \mathbf{CA}\mu_k^{(i)}$ is very large. Most of the samples generated by the prior fall into the left tail of the likelihood, and have extremely low probabilities of being accepted (Figure 1).
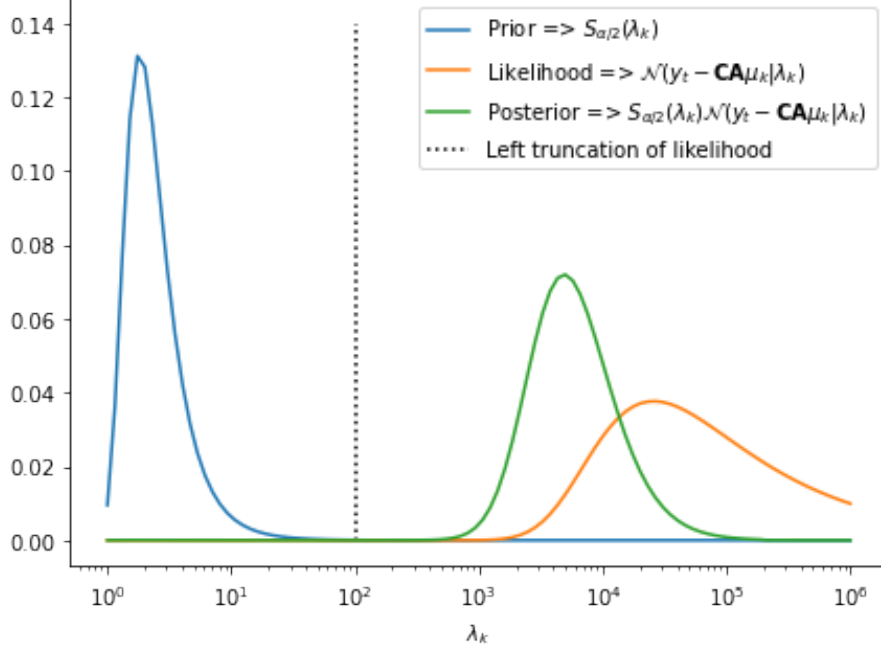

Figure 1: Illustration of the poor acceptance rate regime of prior/likelihood interaction

Due to the additional observation noise term (d), we are unable to use the Inverse Gamma approximation to the full posterior as detailed in [**godsill1999bayesian**]. Instead, we attempt to improve the sampling efficiency by truncating the left tail of the likelihood distribution.

Due to the intractability of calculating the Cumulative Distribution Function (CDF) of the likelihood analytically, we left-truncate the likelihood where it is smaller than a chosen epsilon. In this project, this small epsilon was chosen to be $\epsilon_{trunc} = 1 \times 10^{-9}$. Further checks are also performed to ensure that this truncation only occurs well into the right tail of the prior.

This allows us to generate samples from a left-truncated form of the α-stable prior instead of the full prior. These generated samples are much more likely to be in the high probability regions of the likelihood, thus improving our sampling efficiency greatly.

We sample from the left truncated α-stable distribution by applying a paretian tail approximation to the α-stable distribution , then sampling from the paretian tail approximation

using **??**.

This improved rejection sampler can be described as:

1. Define hyperparameter ϵ_{trunc}

2. Calculate 95th percentile of the prior $(\lambda_{0.95})$ where: $\int_0^{\lambda_{0.95}} S_{\alpha/2}(\lambda_k|1,1,0) = 0.95$

3. Draw proposals $\lambda_k^{(i)}$

   - Calculate a possible truncation value $\lambda_{k,trunc}$ where $\mathcal{N}(y_k - \mathbf{CA}\mu_k^{(i)}|\lambda_{k,trunc}) = \epsilon_{trunc}$.

   - If this truncation value also lies in the right tail of the prior (i.e. $\lambda_{k,trunc} > \lambda_{0.95}$) then sample particle from the paretian tail approximation

   $$\lambda_k^{(i)} \sim f_{bounded\ pareto}(x|\alpha/2, \beta = 1, \lambda_{k,trunc})$$

   - Otherwise, sample particle from the original prior:

   $$\lambda_k^{(i)} \sim S_{\alpha/2}(1,1,0)$$

4. Draw $u \sim U(0, M)$

5. If $u \geq \mathcal{N}\left(y_k - \mathbf{CA}\mu_k^{(i)}|0, \sigma_\lambda \lambda_k + \mu_\lambda\right)$, reject $\lambda_k^{(i)}$ and go to Step (3)

List of Figures

List of Tables