Algorithm 1: Measurement off policy (MOP) SMC:

1 Initialize filter particles: simulate $X_{0,j}^{F,\theta} \sim f_{X_0}(\,\cdot\,;\theta)$ for j in $1\!:\!J$
2 Initialize relative weights: $w_{0,j}^{F,\theta} = 1$ for j in $1\!:\!J$
3 **for** n *in* $1\!:\!N$ **do**
4   Simulate for prediction: $X_{n,j}^{P,\theta} \sim f_{X_n|X_{n-1}}\big(\,\cdot\,|X_{n-1,j}^{F};\theta\big)$ for j in $1\!:\!J$
5   Evaluate measurement density: $g_{n,j}^{\theta} = f_{Y_n|X_n}(y_n^*|X_{n,j}^{P,\theta};\theta)$ for j in $1\!:\!J$
6   Update relative weights to compensate for resampling: $w_{n,j}^{P,\theta} = \dfrac{w_{n-1,j}^{F,\theta}\, g_{n,j}^{\theta}}{g_{n,j}^{\phi}}$ for j in $1\!:\!J$
7   Conditional likelihood under ϕ: $L_n^{\phi} = \frac{1}{J}\sum_{m=1}^{J} g_{n,m}^{\phi}$
8   Normalize weights: $\tilde{g}_{n,j}^{\phi} = \dfrac{g_{n,j}^{\phi}}{JL_n^{\phi}}$ for j in $1\!:\!J$
9   Apply systematic resampling to select indices $k_{1:J}$ with $\text{Prob}\big(k_j = m\big) = \tilde{g}_{n,m}^{\phi}$
10   Resample: set $X_{n,j}^{F,\theta} = X_{n,k_j}^{P,\theta}$ and $w_{n,j}^{F,\theta} = w_{n,k_j}^{P,\theta}$ for j in $1\!:\!J$
11 **end**

- MOP-SMC requires that the algorithm is first run at ϕ, for which it is a vanilla particle filter. $g_{n,j}^{\phi}$ and $\tilde{g}_{n,j}^{\phi}$ are computed at this first pass. Then it is run at θ, with the seed kept fixed.

- Here, the resampling rule for particle j depends on j (and therefore ϕ and $X_{n,j}^{P,\phi}$) but not on θ or $X_{n,j}^{P,\theta}$. This does not have the standard form for a weighted particle filter, for which we can have general reweighting rules but they are usually a funciton of the location of the particle and the model parameter, θ. That may be why it is harder to see that this is a properly weighted filter. Nevertheless, inspection of the algorithm shows that each particle is properly reweighted to account for its resampling probability, in step 6, so it should be properly weighted.

- The final estimate of the likelihood is either

$$L(\theta) = \left(\frac{1}{J}\sum_{j=1}^{J} w_{N,j}^{F,\theta}\right)\prod_{n=1}^{N} L_n^{\phi} \tag{1}$$

  or

$$L'(\theta) = \left(\frac{1}{J}\sum_{j=1}^{J} w_{N,j}^{P,\theta}\, g_{N,j}^{\theta}\right)\prod_{n=1}^{N-1} L_n^{\phi} \tag{2}$$

  with $L'(\theta)$ presumably having slightly lower variance.

- Weighted samples representing the filter distribution, $f_{X_n|Y_{1:n}}(x_n\,|\,y_{1:n}^*\,;\theta)$ are either $\big\{\big(X_{n,j}^{F,\theta}, w_{n,j}^{F,\theta}\big), j \text{ in } 1\!:\!J\big\}$ or $\big\{\big(X_{n,j}^{P,\theta}, g_{n,j}^{\theta}\, w_{n,j}^{P,\theta}\big), j \text{ in } 1\!:\!J\big\}$, meaning that an expectation over $f_{X_n|Y_{1:n}}(x_n\,|\,y_{1:n}^*\,;\theta)$ is consistently estimated by a corresponding weighted average of the filter or prediction particles.

- As long as `rprocess` is a continuously differentiable function of θ for fixed seed, and `dmeasure` is a continuously differentiable function of θ, and $g_{n,j}^{\phi} \neq 0$, we see that MOP is a continuously differentiable function of θ for fixed seed. Since it also provided an unbiased estimate of the likelihood, this justifies exchanging the order of differentiation and integration to ensure that its derivative is an unbiased estimate of the deriative of the likelihood.

- Taking the derivative with respect to θ at $\theta = \phi$, step 6 looks very much like the stop gradient approach.

Here's an outline of an argument explaining why MOP is properly weighted.

Lemma 1. *Suppose that* $\{(X_j, u_j), j = 1, \ldots, J\}$ *is properly weighted for* f_X*, meaning that, if* $X \sim f_X$*,*

$$E[h(X)] \approx \sum_{j=1}^{J} \frac{w_j}{\sum_{k=1}^{J} u_k} h(X_j),$$

for some appropriate formalization of $\approx$*. Now, let* $\{(Y_j, v_j), j = 1, \ldots, J\}$ *be a sample drawn from* $\{(X_j, u_j)\}$ *where* (X_j, u_j) *is represented, on average,* $\pi_j J$ *times. This could amount to binomial resampling having* J *draws each with probability* p_j*, or systematic resampling. Suppose*

$$(Y_j, v_j) = \big(X_{a(j)}, u_{a(j)}/\pi_{a(j)}\big),$$

where a *is called the ancestor function. Then,* $\{(Y_j, v_j), j = 1, \ldots, J\}$ *is als properly weighted for* f_X*.*

Notably, Lemma 1 permits $\pi_{1:J}$ to depend on $\{(X_j, u_j)\}$ as long as the resampling is carried out independently of $\{(X_j, u_j)\}$ conditional on $\pi_{1:J}$.

Lemma 2. *Suppose that* $Z_j \sim f_{Z|X}(\cdot|X_j)$ *where* $f_{Z|X}$ *is a conditional probability density function corresponding to a joint density* $f_{X,Z}$ *with marginal densities* f_X *and* f_Z*. Then,* $\{(Z_j, u_j)\}$ *is properly weighted for* f_Z*.*

Lemma 3. *Suppose that* $(X'_j, u'_j) = \big(X_j, u_j\, f_{Z|X}(z^*|X_j)\big)$*. Then,* $\{(X'_j, u'_j)\}$ *is properly weighted for* $f_{X|Z}(\cdot|z^*)$*.*

Recursively applying Lemmas 1, 2 and 3, we obtain that the MOP filter is properly weighted. Specifically, suppose inductively that $\big\{\big(X^{F,\theta}_{n-1,j}, w^{F,\theta}_{n-1,j}\big)\big\}$ is properly weighted for $f_{X_{n-1}|Y_{1:n-1}}(x_{n-1}|y^*_{1:n-1};\theta)$. Then, Lemma 2 tells us that $\big\{\big(X^{P,\theta}_{n,j}, w^{P,\theta}_{n,j}\big)\big\}$ is properly weighted for $f_{X_n|Y_{1:n-1}}(x_n|y^*_{1:n-1};\theta)$. Lemma 3 tells us that $\big\{\big(X^{P,\theta}_{n,j}, w^{P,\theta}_{n,j} g^{\theta}_{n,j}\big)\big\}$ is therefore properly weighted for $f_{X_n|Y_{1:n}}(x_n|y^*_{1:n};\theta)$. Lemma 1 guarantees that the resampling rule, given by

$$\big(X^{F,\theta}_{n,j}, w^{F,\theta}_{n,j}\big) = \big(X^{P,\theta}_{n,a(j)}, w^{P,\theta}_{n,j} g^{\theta}_{n,j} / g^{\phi}_{n,j}\big),$$

with resampling weights proportional to $g^{\phi}_{n,j}$, is therefore also properly weighted for $f_{X_n|Y_{1:n}}(x_n|y^*_{1:n};\theta)$.

This has addressed filtering, but not quite yet the likelihood evaluation.

A properly weighted conditional likelihood estimate from MOP (give the argument above) is

$$f_{Y_n|Y_{1:n-1}}(y^*_n|y^*_{1_n-1};\theta) \approx \frac{\sum_{j=1}^{J} g^{\theta}_{n,j} w^{P,\theta}_{n,j}}{\sum_{j=1}^{J} w^{P,\theta}_{n,j}}.$$

We write the numerator as

$$L^{\phi}_n \sum_{j=1}^{J} \left[\frac{g^{\theta}_{n,j}}{g^{\phi}_{n,j}} w^{P,\theta}_{n,j}\right] \frac{g^{\phi}_{n,j}}{L^{\phi}_n}$$

Using Lemma 1, we see this is properly estimated by

$$L^{\phi}_n \sum_{j=1}^{J} w^{F,\theta}_{n,j}.$$

This corresponds to

$$L^{\theta} = \prod_{n=1}^{N} L^{\phi}_n \frac{\sum_{j=1}^{J} w^{F,\theta}_{n,j}}{\sum_{j=1}^{J} w^{P,\theta}_{n,j}}.$$

Since $w^{F,\theta}_{n,j} = w^{P,\theta}_{n+1,j}$, this is a telescoping product. The remaining terms are $\sum_{j=1}^{J} w^{P,\theta}_{0,j} = 1$ on the denominator and $\sum_{j=1}^{J} w^{F,\theta}_{N,j}$ on the numerator. This derives the MOP estimate in (1).