Software practical — Uncertainty Quantification

Implementation of a C++ Particle Filter Library

Winter semester 2019

Submitted by: **Nils Friess**

February 26, 2020

Supervisors: Prof. Robert Scheichl, Gianluca Detommaso

Heidelberg University

Introduction

{sec:intro}

In this report we present an implementation of a Particle Filter in C++. Before discussing the actual implementation we give the theoretical details of the method in this section. We start by giving a brief overview of the method before discussing the individual steps in more detail.

We remark here, that the naming in these methods is highly ambiguous and varies greatly from author to author and even in different publications of the same authors. We use – with some exceptions – the naming and notation used in [3] and highlight parts where the naming differs from other publications.

Overview of a Particle Filter

A Particle Filter is a Sequential Monte Carlo (SMC) method[1] that is used to estimate the state of a system that changes over time using only noisy and/ or partial observations of the system's state. This will be done in a Bayesian framework where one attempts to construct the posterior probability density function (pdf) of the state based on the observations. We make the following assumptions:

(A1) The model describing the initial state and the evolution of the internal state in time is available in a probablistic form.

(A2) The model that relates the observations to the internal state is available in a probabilistic form.

(A3) The observations are only available sequentially, not as a batch (ie. we assume that we receive new measurments sequentially in time).

Due to (A3) we aim at a recursive method that does neither require to store nor to reprocess all the previous information when a new observation becomes available. To formalise the first two assumptions we will use the notion of *hidden markov models*.

[1] Some authors use the terms *Particle Filter* and *SMC method* synonymously. Doucet and Johansen develop in [3] a framework in which Particle Filters are only one specific method in the much broader class of SMC methods. They argue that this distinction allows for a better understanding of these methods. In this report, we are only interested in the *filtering problem* and we will introduce it without discussing the more general notion of SMC methods as given by Doucet and Johansen.

Such models consist of the triplet

$$\begin{aligned} X_0 &\sim \mu(x_0), \\ X_n \mid (X_{n-1} = x_{n-1}) &\sim p(x_n \mid x_{n-1}), \\ Y_n \mid (X_n = x_n)c &\sim p(y_n \mid x_n), \end{aligned}$$

where

- $n \in \mathbb{N}$ denotes discrete time;
- X_n is the d_x-dimensional state of the system taking values in $\mathbb{R}^{d_x}$;
- $p(x_0)$ is the prior probability density fumction (pdf) of the system's state;[2]
- Y_n is the d_y-dimensional vector of observations which is assumed to be conditionally iindependent of all other observations given the state X_n,
- $p(y_n \mid x_n)$ is the conditional pdf of Y_n given $X_n = x_n$.

Assumptions (A1) and (A2) then state that all these pdfs are known. Our goal is now to estimate the distribution $p(x_n \mid y_{1:n})$, where $y_{1:n} \coloneqq (y_1, y_2, \ldots, y_n)$. This is often referred to as the *filtering problem* or *tracking*.[3]

1-step-predictor (Chapman-Kolmogorov)

In a restrictive set of cases this distribution can be computed exactly (e. g. when $p(y_t \mid y_t)$ is linear and the posterior of the system is Gaussian [1, p. 175] or when the underlying state space of the Markov model is finite, cf. [3, Example 1]). In a more general nonlinear non-Gaussian setting, approximative methods such as particle filters are necessary. Therefore, in practice a different approach is used.

(Sequential) importance sampling

The central idea of Particle filters is to represent the posterior of the system $p(x_k \mid y_{1:k})$ at some time k as a weighted set of samples, so called `particles`, denoted by $\{x_k^{(i)}, w_k^{(i)}\}$. If

[2] With abuse of notation we denote by $p(x)$ the pdf of the random variable X. For two random variables X and Y the corresponding (possibly different) density functions are denoted by $p(x)$ and $p(y)$ respectively; $p(x, y)$ denotes the joint pdf and $p(x \mid y)$ is the conditional pdf of X given $Y = y$.

[3] Note that Docuet and Johansen *do not* call this the filtering problem [3]. They reserve this term for the estimation of the joint distributions $p(x_{1:n} \mid y_{1:n})$. Since we are only concerned with estimating the marginal distribution $p(x_n \mid y_{1:n})$ we will still refer to this problem as filtering.

we ignore for a moment the weights and assume that the samples are from the desired distribution, i. e.

$$x_k^{(i)} \sim p(x_k^{(i)} \mid y_{1:k}), \quad i = 1, \ldots, N$$

the Monte carlo method approximates $p(x_k \mid y_{1:k})$ by the empirical measure[4]

$$\hat{p}(x_k \mid y_{1:k}) = \frac{1}{N} \sum_{i=1}^{N} \delta_{x_k^{(i)}}(x_k), \tag{1}$$ {eq:empirical_m

where $\delta_x(\cdot)$ denotes the Dirac delta centered at x. The expectation of a test function $f : \mathbb{R}^{d_x} \to \mathbb{R}$ given by

$$\mathbb{E}[f(x_k) \mid y_{1:k}] = \int f(x_k) p(x_k \mid y_{1_k}) \, \mathrm{d}x_k$$

is then estimated by

$$\mathbb{E}^{\mathrm{MC}}[f(x_k) \mid y_{1:k}] = \int f(x_k) \hat{p}(x_k \mid y_{1_k}) \, \mathrm{d}x_k = \frac{1}{N} \sum_{i=1}^{N} f(x_k^{(i)}).$$

It is well-known that the variance of the approximation error using this estimator decreases *independent of* d_x with a rate of $\mathcal{O}(N^{-1})$. However, often it is either impossible or practically intractable to sample from the posterior directly and thus, in practice one often relies on a technique called *importance sampling*.

We start by choosing an *importance density* $q(x_k \mid y_{1:k})$ and draw N samples $x_k^{(i)}$ from it. If we would use these samples to approximate $p(x_k \mid y_{1:k})$ as in (1) the result would obviously not be accurate in general. To correct this bias we introduce *importance weights*

$$w_k^{(i)} \propto \frac{p(x_k^{(i)} \mid y_{1:k})}{q(x_k^{(i)} \mid y_{1:k})}, \tag{2}$$ {eq:importance_

that we require to be normalised such that $\sum_i w_k^{(i)} = 1$. We can now approximate the target density by

$$p(x_k \mid y_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta_{x_k^{(i)}}(x_k). \tag{3}$$ {eq:karget:appr

This technique is called *importance sampling*. Expectations of test functions can then be

[4] Again, we slightly abuse the notation for simplicity; the alternations required for a rigorous measure-theoretic formulation are straightforward.

estimated by

$$\mathbb{E}^{\mathrm{MC}}[f(x_k) \mid y_{1:k}] = \sum_{i=1}^{N} w_k^{(i)} f(x_k^{(i)}).$$

Due to assumption (A3) ideally we would like a recursive formula to update the weights a each step. To obtain such a formula we consider the full posterior $p(x_{0:k} \mid y_{1_k})$ and express it in terms of the posterior at the previous time step and the known pdf's $p(y_k \mid x_k)$ and $p(x_k \mid x_{k-1})$:

$$\begin{aligned} p(x_{0:k} \mid y_{1:k}) &\propto p(y_k \mid x_{0:k}, y_{1:k-1}) p(x_{0:k} \mid y_{1:k-1}) \\ &= p(y_k \mid x_k) p(x_k \mid x_{0:k-1}, y_{1:k-1}) p(x_{0:k-1} \mid y_{1:k-1}) \\ &= p(y_k \mid x_k) p(x_k \mid x_{k-1}) p(x_{0:k-1} \mid y_{1:k-1}), \end{aligned}$$

where we used Bayes' theorem and the properties of the system described earlier. If in addition we choose an importance density that factorizes such that

$$q(x_{0:k} \mid y_{1:k}) = q(x_k \mid x_{0:k-1}, y_{1:k}) q(x_{0:k-1} \mid y_{1:k-1})$$

the weights (2) can be written as

$$\begin{aligned} w_k^{(i)} &\propto \frac{p(y_k \mid x_k^{(i)}) p(x_k^{(i)} \mid x_{k-1}^{(i)}) p(x_{0:k-1}^{(i)} \mid y_{1:k-1})}{q(x_k^{(i)} \mid x_{0:k-1}^{(i)}, y_{1:k}) q(x_{0:k-1}^{(i)} \mid y_{1:k-1})} \\ &= \frac{p(y_k \mid x_k^{(i)}) p(x_k^{(i)} \mid x_{k-1}^{(i)})}{q(x_k^{(i)} \mid x_{0:k-1}^{(i)}, y_{1:k})} w_k^{(i-1)}. \end{aligned}$$

Since in our case, we are only interested in estimating the filtered posterior $p(x_k \mid y_{1:k})$ we choose an importance density $q(x_k \mid x_{0:k-1}, y_{1:k}) = q(x_k \mid x_{k-1}, y_k)$ that only depends on x_{k-1} and y_k. Then, merely $x_k^{(i)}$ is held in memory and the path $x_{0:k-1}^{(i)}$ and history of observations $y_{1:k-1}$ need not be stored. The weights can the recursively be computed by

$$w_k^{(i)} \propto \frac{p(y_k \mid x_k^{(i)}) p(x_k^{(i)} \mid x_{k-1}^{(i)})}{q(x_k^{(i)} \mid x_{k-1}^{(i)}, y_{1:k})} w_k^{(i-1)}. \tag{4}$$ {eq:weight_upda

This is usually referred to as *sequential* importance sampling. We summarise the results up to this point in Algorithm 1. Note, that since at time $k = 0$ no observation or previous state is available, we sample from the (known) prior and weigh all particles equally.

Using this approach alone, however, leads to *degeneracy* of the particles. It can be shown

Algorithm 1: Sequential importance sampling

Input: n observations $y_1, y_2, \ldots, y_n$; number of particles N

Sample $x_0^{(i)} \sim \mu(x_0^{(i)})$;

Set weights $w_0^{(i)} = 1/N$;

for $k = 1, 2, \ldots, n$ **do**

  Sample $x_k^{(i)} \sim q(x_k^{(i)} \mid x_{k-1}^{(i)}, y_k)$. ;

  Compute weights according to (4). ;

  Normalise $w_k^{(i)} = \tilde{w}_k^{(i)} / \sum_j \tilde{w}_k^{(j)}$

end

{alg:sis}

that the variance of the weights can only increase at every step, which implies that the algorithm will eventually produce a single non-zero weight $w^{(i)} \approx 1$, carrying all the statistical information with the rest of the weights converging to zero. This is visualised in Figure 1 where we plotted a histogram of the weights after the first few time steps. One can clearly see that after a few steps almost all the weights are zero. To account for this problem, we introduce another technique called *resampling*.

Resampling

We are going to present two resampling strategies in this section. The overall goal of all resampling methods is to remove particles with negligible weights with a high probability and replicate those with high weights. After resampling, the future particles are more concentrated in domains of higher posterior probability, which entails improved estimates. It can, of course, happen that a particle with a low weight at time t has a high weight at time $t + 1$, in which case resampling could be wasteful. It should also be mentioned that if particles have (unnormalised) weights with a small variance, resampling might be unnecessary. This is discussed briefly at the end of this section. As above, we denote by $\{x^{(i)}, w^{(i)}\}_{1 \le i \le N}$ the set of particles with their associated weights at some time k (which is omitted in the notation). We assume that the weights have already been normalised, i. e. $\sum_i w^{(i)} = 1$. We further denote by $\{\tilde{x}^{(i)}, \tilde{w}^{(i)}\}_{1 \le i \le N}$ the particles and weights after resampling took place. We require the particles $\tilde{x}^{(i)}$ to be weighted equally which implies, since we also require the weights to be normalised, that $\tilde{w}^{(i)} = 1/N$.

The use of resampling to improve importance sampling was originally introduced by Gordon et al, see [4], laying the ground for Particle filters and SMC methods in general.

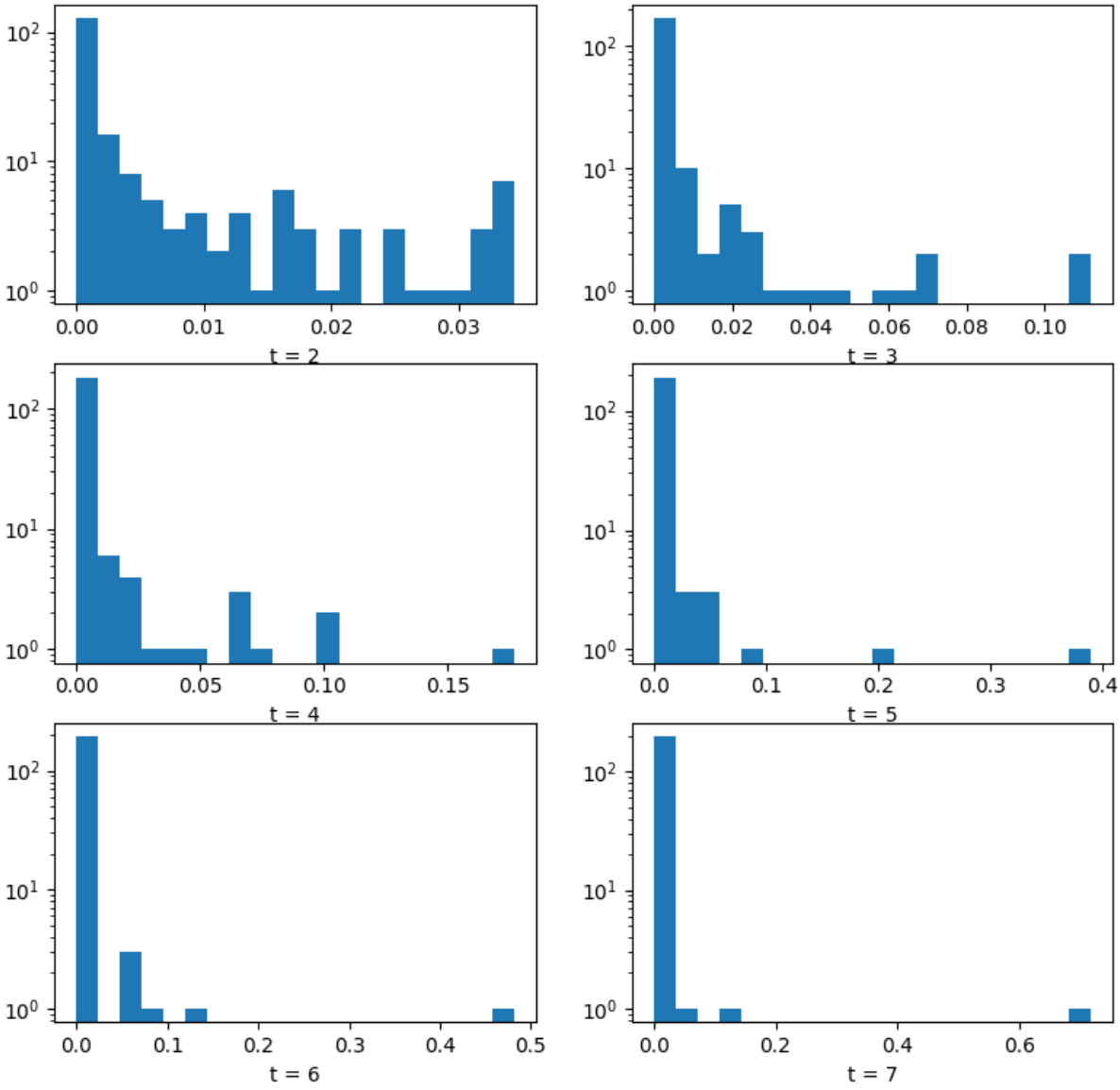

Figure 1: A histogram of 200 weights after just a few iterations. Almost all the weights are zero at $k = 7$ which demonstrates the degeneracy of the particles. {fig:weights}

The resampling methods presented here are two of the most popular amongst the literature, see [2]. The most simple resampling strategy, called *multinomial resampling* is not discussed here due to its poor performance compared to other techniques. It only is mentioned because it is the method introduced by Gordon et al. in [4] as part of the *bootstrap filter*, that uses the prior as the proposal density (c. f. Example 4).

Both the methods presented in the following are based on drawing samples from the point mass distribution $\sum_{j=1}^{N} w^{(j)} \delta_{X^{(j)}}$. In practice, this is achieved by repeated uses of the inversion method, which itself uses the empirical cumulative distribution function (cdf) associated with the weights. This is based on the following fact:

Claim. If U is a uniform random variable on $(0,1]$ then $X = F^{-1}(U)$ has distribution F, where F is the cumulative distribution function of X and $F^{-1}(t) = \min\{x \mid F(x) = t\}$ is the inverse cdf.

Proof. Let $U \sim \mathcal{U}(0,1]$. Then

$$\begin{aligned} P(F^{-1}(U) \leq x) &= P(\min\{x \mid F(x) = U\} \leq x) && \text{(definition of } F^{-1}\text{)} \\ &= P(U \leq F(x)) \\ &= F(x) && \text{(definition of distribution of } U\text{)}. \end{aligned}$$

□

The inversion method can be explained visually as follows. We plot the empirical cdf of the weights and sample from $U \sim \mathcal{U}(0,1]$. We denote the actual value of the sample by u. We then draw a horizontal line from the coordinate $(0,u)$ to the right until it intersects one of the bars, see Figure 2. The index of the bar that is intersected determines the new sample.

In our case, we do not draw just one sample U but we generate N different samples $\{U_i\}_{1 \leq i \leq N}$ in such a way that they are sorted in ascending order. For every of these samples (from lowest to highest) we look for the intersected bar and add its index to a list. This list then corresponds to the indices of the particles that should be resampled. Consider the following example: Suppose we had five particles and the list of indices after the resampling reads $\{1,3,4,4,5\}$. Then, particles $X^{(1)}, X^{(3)}, X^{(5)}$ should be resampled, particle $X^{(4)}$ should even be duplicated. Particle $X^{(2)}$, however, will be dropped. In other

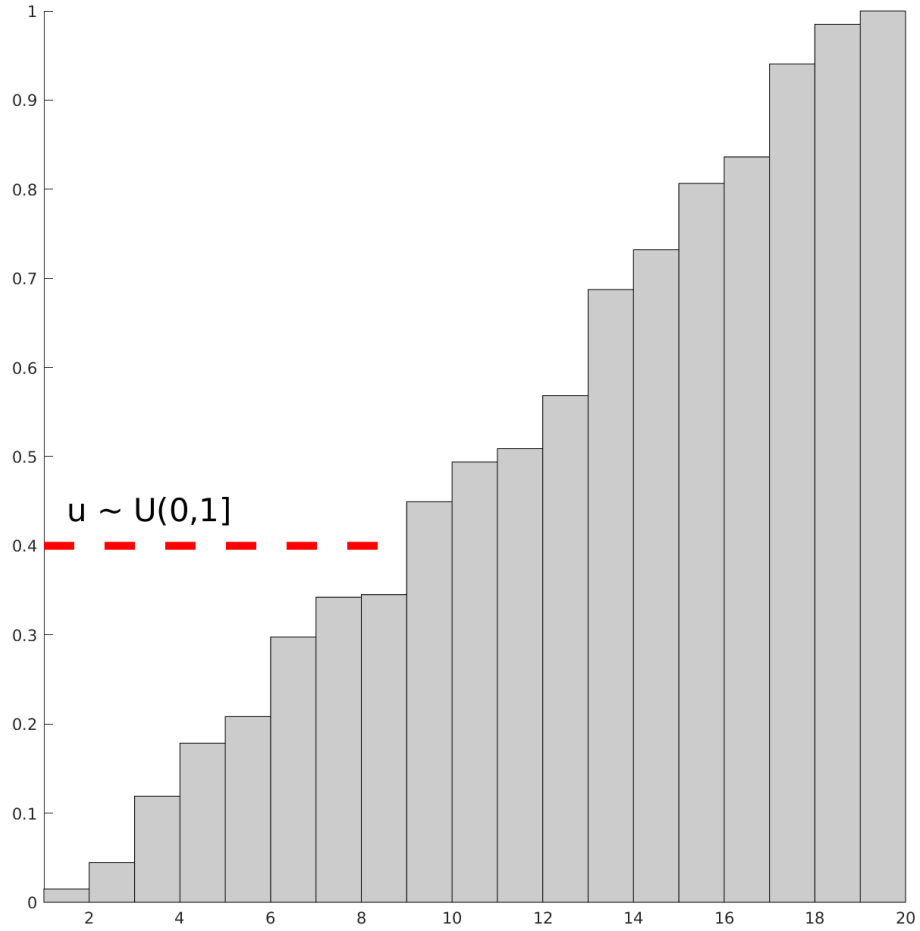

Figure 2: Visualisation of the inversion method. The bars represent the empirical cumulative distribution function associated with a set of 19 weights. The red line is horizontally drawn from the value of u at the vertical axis until it "hits" a bar. The index of the bar yields the generated sample (in this case the new sample is therefore 9).

{fig:ecdf}

words, the particles after the resampling are

$$\tilde{X}^{(1)} = X^{(1)},\ \tilde{X}^{(2)} = X^{(3)},\ \tilde{X}^{(3)} = X^{(4)},$$
$$\tilde{X}^{(4)} = X^{(4)},\ \tilde{X}^{(5)} = X^{(5)}.$$

The two strategies presented in the following only differ in the way the U_is are generated. We summarise the results in Algorithm 2.

Algorithm 2: Resampling using the empirical cdf

Input: N samples $U_i \sim \mathcal{U}(0,1]$ sorted in ascending order; list of weights
Result: List of indices I that represent that particles to be resampled
C = `cumsum`(weights) ; `// Generate the empirical cdf as list of cumulated sums`
I = zeros(N);
$i, j = 0$;
while $i < N$ **do**
  if $U_i < C_j$ **then** `// Found intersecting bar`
    $I_i = j$;
    $i = i + 1$;
  else
    $j = j + 1$;
  end
end

{alg:resampling}

Here, the function `cumsum` is assumed to work in the same way as, for example, MATLAB's or NumPy's `cumsum` function. That is, `cumsum([1,2,3,4])` should return `[1,3,6,10]`.

Systematic Resampling

This algorithm separates the sample space into N divisions. One random offset, drawn from a $\mathcal{U}[0,1)$ distribution, is used to choose where to sample from for all divisions. This guarantees that every sample is exactly $1/N$ apart.

In other words, the U_is are generated by sampling $\tilde{U} \sim \mathcal{U}[0,1)$ and defining

$$U_i = \frac{\tilde{U} + i - 1}{N} \quad \text{for } i = 1, \ldots, N.$$

Stratified Resampling

This algorithm is similar to the previous one, its aim is to make selections relatively uniformly across the particles. We start by partitioning the $(0,1]$ interval into N disjoint sets, $(0,1] = (0,1/N] \cup (1/N, 2/N] \cup \cdots \cup ((N-1)/N, 1]$. The U_is are then drawn independently in each of the sub-intervals:

$$U_i \sim \mathcal{U}((i-1)/N, i/N].$$

We mentioned earlier that resampling might be unnecessary if the weights are sufficiently uniform and we would like to have a criterion allowing us to check whether resampling should be performed. To that end the effective sampling size (ESS) is often used which can be estimated using

$$ESS \approx \left(\sum_{i=1}^{N} \left(w_t^{(i)} \right)^2 \right)^{-1},$$

where the weights $w_t^{(i)}$ are assumed to be already normalised (for more details, see [1, p. 179]). If the variance of the weights is maximal, i. e. if all but one of the weights are zero, the value of ESS is 1. If, however, the weights all have the same value $w_t^{(i)} = 1/N$ the value of ESS is N, since

$$\left(\sum_{i=1}^{N} \left(\frac{1}{N} \right)^2 \right)^{-1} = \left(N \frac{1}{N^2} \right)^{-1} = N.$$

Therefore, we will only resample if ESS is below a certain threshold, e. g. $N/2$.

We have now gathered everything we need to implement a particle filter, since essentially particle filters are simply a combination of sequential importance sampling and some resampling strategies. Therefore, sometimes these methods are also called as*Sequential Importance Sampling with Resampling* abbreviated by SIS/R.

Implementation

In this section we discuss an implemenation of a particle filter in C++. The particle filter itself is implemented as a dependency-free header-only generic library that, while being easy to set up and use, is versatile and can be used with a wide variety of problems. This is demonstrated by three examples, two of which are based on the same problem but are using different prior and proposal distributions.

The library consists of the following classes (and their respective header files)

- Particle (`particle.hh`)
- ParticleFilter (`particlefilter.hh`)
- Model (`model.hh`)
- History (`history.hh`)

All of these classes are templated to allow for arbitrary particle types, e. g. the `Particle` class, that holds the value and weight of a single particle, is of the following form

```
template <class PT>
class Particle {
private:
  PT m_value;
  double m_weight;
...
};
```

where the particle type `PT` could take values of some finite set, be a real number (i. e. a `double`) or a n-dimensional vector etc. The `ParticleFilter` class implements the algorithms introduced above. It takes the following template parameters

```
template <class PT, class OT, size_t N,
          bool enable_history = false,
          bool parallel = false>
class ParticleFilter
```

where PT and OT denote the type of particle and observation, respectively, N is the number of particles, enable_history specifies whether all particles from every timestep should be held in memory (e. g. for debugging or plotting) and parallel specifies whether certain functions should run in parallel (see below). Apart from these compile-time parameters, to construct a ParticleFilter one also needs to provide an instance of a Model, a resampling strategy[5], a resampling treshhold and an initial seed for the random number generator (rng). Only the first parameter is mandatory, i. e. the signature of the constructor of the ParticleFilter class is given by

```
ParticleFilter(
      Model<PT, OT> *t_model,
      double t_seed = 0,
      ResamplingStrategy t_strategy = ResamplingStrategy::RESAMPLING_SYSTEMATIC,
      double t_treshhold = 0.5)
```

The Model class is implemented as an abstract base class (sometimes called interface), meaning that the class itself cannot be instaniated. Therefore, in order to define a model, a class that is derived from Model has to be implemented. The Model class defines the following four pure virtual functions that every model that derives from this class needs to override.

```
virtual void sample_prior(Particle<PT> &t_particle) = 0;

virtual PT zero_particle() = 0;

virtual double update_weight(const Particle<PT> &t_particle_before_sampling,
                             const Particle<PT> &t_particle_after_sampling,
                             const OT &t_observation,
                             double t_time) = 0;

virtual PT sample_proposal(const Particle<PT> &t_particle,
                           const OT &t_observation,
                           double t_time) = 0;
```

All these methods are automatically called by the ParticleFilter. The method sample_prior is used to initialise the set of particles. It has to set the value of the given particle t_particle using the set_value method of the particle. This is different from

[5] At the moment, only systematic resampling is implemented.

the `update_weight` and `sample_proposal` methods that *do not* alter the particle themselves. Rather they should return the updated value of the particle's weight and value, respectively (see examples below).

The first paramter of the `update_weight` method is the particle before `sample_proposal` is called and the second after it is called. This is useful, since the developer cannot specify the order in which these two methods are called. However, some models require the value of the particle before it has been updated (cf. Example 4) and some after the sampling step (cf. Example 5). Also, it is important to note that

Example 3. Consider the following model (this example has been studied in a number of publications before, see for example [1, 4, 5]). Let

$$f(x_n \mid x_{n-1}) = \mathcal{N}(x_n; h_n(x_{n-1}), \sigma_{\text{sys}})$$
$$g(y_n \mid x_n) = \mathcal{N}(y_n; \frac{{x_n}^2}{20}, \sigma_{\text{obs}})$$

where

$$h_n(x_{n-1}) = \frac{1}{2}x_{n-1} + \frac{25x_{n-1}}{1+x_{n-1}^2} + 8\cos(1.2n)$$

{ex:1v1}

Example 4 (Lotka Volterra using bootstrap filter).

{ex:1v2}

Example 5 (Lotka Volterra using optimal proposal).

References

[1] M Sanjeev Arulampalam et al. "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking". In: *IEEE Transactions on signal processing* 50.2 (2002), pp. 174–188.

[2] R. Douc and O. Cappe. "Comparison of resampling schemes for particle filtering". In: *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis, 2005.* Sept. 2005, pp. 64–69.

[3] Arnaud Doucet and Adam M. Johansen. "A Tutorial on Particle Filtering and Smoothing: Fifteen years later". In: *The Oxford handbook of nonlinear filtering* (2011). Ed. by Dan Crisan and Boris Rozovskii, pp. 656–705.

[4] Neil J Gordon, David J Salmond, and Adrian FM Smith. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". In: *IEE proceedings F (radar and signal processing)*. Vol. 140. 2. IET. 1993, pp. 107–113.

[5] Genshiro Kitagawa. "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models". In: *Journal of Computational and Graphical Statistics* 5.1 (1996), pp. 1–25.