Nonlinear Filtering with Homotopy Continuation

Marcel Hiltscher

Abstract—The filtering in a nonlinear context is a complex endeavour that in most cases does not have any closed form solution available. Thus, an approximation of the posterior density is necessary. This approximation is however not a trivial task to accomplish as many algorithms that are leveraged in nonlinear filtering have to deal with difficulties like local minima or particle degeneracy. In this context, the mathematical concept of homotopy is leveraged to circumvent these problems.

I. INTRODUCTION

The task of state estimation based on a given system emerges in a variety of fields such as tracking, computer vision, robotics, or navigation. In many cases these systems have a dynamic property. Consequently, the states can not be directly observed due to physical limitations. In order to circumvent this problem, noisy measurements are being taken from the system and used for the state estimation.

In order to estimate states in a dynamic system, the corresponding system and measurement model are used for the necessary inference. A system model describes the relation of the state changes whereas the measurement model connects the state and the measurement.

The general approach of state estimation leverages all information available and proceeds in two steps:

1) prediction using the state model
2) measurement update or filter step using the measurement model

First, the prediction step is concerned with the state estimation before a new measurement is obtained. Thus, the estimation is based on previous measurements and system inputs. In Bayesian terms, the resulting estimate is the so called prior estimate. Next, we include the new measurement in order to update our estimation which leads to the posterior estimate. Afterwards, the posterior is used to calculate the prior for the next time stamp and the procedure starts from the beginning. These estimates can be specified as probability densities conditioned on the measurements and the preceding system inputs.

Including all measurement and input information at every estimation cycle is not computationally beneficial. Thus, a recursive approach that sequentially leverages the current information is preferred.

An additional property of systems is the linearity or nonlinearity which respectively leads to different strategies regarding the previously mentioned estimation procedure. In the linear case, the well known Kalman Filter [17] leads to an optimal solution for the estimation problem and includes a recursive approach. However, the nonlinearity of a system impedes a closed form solution of the prior and posterior distribution. The reason for this ist that the calculation of the respective probability density is computational infeasible and in the worst

case not possible. Thus, an approximation of the respective estimates is necessary. Typical examples for approaching this problem are the Unscented Kalman Filter (UKF) [16] or the Extended Kalman Filter (EKF) [15].

Next to the Kalman based filters, many other approaches are introduced in scientific research [6]. A broad field of recursive methods are particle filters approximating the density by weighted samples or rather particles [1], [5]. The sampling follows different paradigms like the importance sampling or sequential Monte Carlo [3]. The advantage of such filters is that they do not depend on the posterior to be closely approximated by a Gaussian distribution [3] like the case in the UKF.

Beside the particle filter, the use of homotopy continuation methods is a further field of nonlinear state estimation. The core idea is to gradually approximate the desired density by a given initial density via a homotopy. It allows for the posterior density to have the same parameterization class as the prior density which is a positive effect because the posterior is used for the calculation of the prior distribution in the next prediction step [9]. The use of homotopy progressive methods are introduced in [12] and [6]. However, the latter combines it with particle flow whereas the former introduce a more general approach to nonlinear state estimation.

The focus of this seminar are homotopy based filters in a nonlinear setting. In the following sections, the usage of homotopy is discussed for two distinct use cases namely the approximation of the posterior parameters and the usage in importance sampling according to particle filters. First, section II will formally define the foundation of the estimation problem and then focuses on the problems of nonlinear filtering. This section also includes a short explanation of advantages that homotopy methodology imply with respect to the problems. Section III will derive the basics of homotopy filtering and explain its usage in context of posterior parametrization. In section IV, the intersection of homotopy and particle filter in context of importance sampling will be given which is then followed by the conclusion V.

II. PROBLEM FORMULATION

The estimation of a state depends on the underlying structure of the nonlinear system and measurement model. A general definition for discrete-time models is given by

$$\underline{x}_{k+1} = \underline{a}_k(\underline{x}_k, \hat{\underline{u}}_k, \underline{w}_k), \qquad (1)$$

$$\underline{y}_k = \underline{h}_k(\underline{x}_k, \underline{v}_k), \qquad (2)$$

where equation (1) defines the system model whereas equation (2) defines the measurement model. In equation (1), the $\underline{a}_k$ describes the nonlinear relation of state vector $\underline{x}_k$ and the

system input $\hat{\underline{u}}_k$ with regard to the state vector in the next time step – $\underline{x}_{k+1}$. System disturbances are modelled by a random vector $\underline{w}_k$ with a density of $f_k^w(\underline{w}_k)$. With respect to the state vector $\underline{x}_k$ and the measurement vector $\underline{y}_k$, the nonlinear relation is given by $\underline{h}_k$. The measurement noise is stated as a random vector $\underline{v}_k$ including its corresponding density $f_k^v(v_k)$. Additionally, let k denote a discrete time stamp. For $\underline{w}_k$ and $\underline{v}_k$ a white noise assumption is used.

Based on the equations above, the probability density of the new state, called transition probability, and the probability density for the measurement, referred to as the likelihood function, are defined as:

$$f^T \sim p(\underline{x}_{k+1}|\underline{x}_k, \hat{\underline{u}}_k)\,, \tag{3}$$
$$f^L \sim p(\underline{y}_k|\underline{x}_k)\,. \tag{4}$$

For simplicity reasons, the $\hat{\underline{u}}_k$ in (3) is omitted which leads to $p(\underline{x}_{k+1}|\underline{x}_k)$.

According to the probabilities outlined above, the two steps of a state estimation can be formulated. Firstly, the first order markov model for the state transition is assumed. The Markov property states that the current state vector only depends on the last state vector $p(\underline{x}_{k+1}|\underline{x}_{0:k}) = p(\underline{x}_{k+1}|\underline{x}_k)$ with $\underline{x}_{0:k} := \{\underline{x}_0, \dots, \underline{x}_k\}$. Furthermore, let Y_k denote a set of measurements from time one up to k $\underline{Y}_k := \{\underline{y}_1, \dots, \underline{y}_k\}$.

The prediction is stated as

$$\begin{aligned} p(\underline{x}_{k+1}|\underline{Y}_k) &= \int p(\underline{x}_{k+1}, \underline{x}_k|\underline{Y}_k)\,\mathrm{d}\underline{x}_k \\ &= \int p(\underline{x}_{k+1}|\underline{x}_k)p(\underline{x}_k|Y_k)\,\mathrm{d}\underline{x}_k\,, \end{aligned} \tag{5}$$

and is referred to as the prior density. It combines the transition density (3) and the previous measurements. On the contrary, a recursive measurement update can be derived with help of Bayes' theorem resulting in the desired posterior density:

$$p(\underline{x}_{k+1}|\underline{Y}_{k+1}) = \frac{p(\underline{y}_{k+1}|\underline{x}_{k+1})p(\underline{x}_{k+1}|\underline{Y}_k)}{p(\underline{y}_{k+1}|\underline{Y}_k)}\,. \tag{6}$$

The numerator of the the posterior density combines the likelihood (4) with the prior density (5) where the denominator only functions as a normalization factor. The denominator of the posterior density (6) can be left out which simplifies the expression to $p(\underline{x}_{k+1}|\underline{Y}_{k+1}) \propto p(\underline{y}_{k+1}|\underline{x}_{k+1})p(\underline{x}_{k+1}|\underline{Y}_k)$.

Assuming the models (1) and (2) to be a probabilistic model, the equations (5) and (6) can be rewritten to

$$f_{k+1}(\underline{x}_{k+1}|\underline{Y}_k) = \int f_{k+1}^T(\underline{x}_{k+1}|\underline{x}_k)f_k^e(\underline{x}_k|Y_k)\,\mathrm{d}\underline{x}_k \tag{7}$$

$$\begin{aligned} f_{k+1}(\underline{x}_{k+1}|\underline{Y}_{k+1}) &= \frac{f_{k+1}^L(\underline{y}_{k+1}|\underline{x}_{k+1})f_{k+1}^p(\underline{x}_{k+1}|\underline{Y}_k)}{f(\underline{y}_{k+1}|\underline{Y}_k)} \\ &= c\,f_{k+1}^L(\underline{y}_{k+1}|\underline{x}_{k+1})f_{k+1}^p(\underline{x}_{k+1}|\underline{Y}_k) \\ &\propto f_{k+1}^L(\underline{y}_{k+1}|\underline{x}_{k+1})f_{k+1}^p(\underline{x}_{k+1}|\underline{Y}_k)\,. \end{aligned} \tag{8}$$

For an easier notation, the prior density (7) is abbreviated to $f_{k+1}^p(\underline{x}_{k+1})$ and the posterior density (8) to $f_{k+1}^e(\underline{x}_{k+1})$. Furthermore, the likelihood function can also be reduced to $f_{k+1}^L(\underline{x}_{k+1})$.

As stated in section I a closed form solution of the prior and posterior density is generally not possible except in few special cases - e.g. additive Gaussian noise with initial Gaussian distribution. Thus, an approximation for the respective densities is needed.

In the following chapters, an approximation of a true density $\tilde{f}$ is given by f. Additionally, the formulations are given for a specific k^{th} measurement update or rather filter step. The goal now is to find an approximation of $\tilde{f}^e(\underline{x})$ denoted as $f^e(\underline{x})$. Keep in mind that the recursive filtering demand a re-approximation after each filtering step because the type of the density may change.

One method is to use a measure of deviation $G(\underline{\theta})$ between the approximation probability density $f^e(\underline{x}, \underline{\theta})$ – e.g. Gaussian Mixtures – and the desired true posterior $\tilde{f}^e(\underline{x})$. In order to get a good approximation, one need to minimize the deviation measure with respect to the parameter vector $\underline{\theta}$. Though, numerical solution of the optimization problem may lead to local minima which does not ensure an optimal parameter for the approximation density $f^e(\underline{x}, \underline{\theta})$.

Furthermore, it is possible to calculate f^e based on weighted samples using particle filters. The main problem here is the sample degeneracy. The sample degeneracy states that only few particles contribute to the approximation whereas most particles have weights close to zero [1] leading to an increase in the weight variance. Therefore, a large part of the computational power is used for updating particles having no effect in the approximation process.

III. HOMOTOPY CONTINUATION AND PARAMETER OPTIMIZATION

Homotopy continuation or short homotopy is a mathematical concept originated from topology and differential geometry. A homotopy can be used to continuously deform a function to a different function. For example, the use of a homotopy allows to continuously deform a circle to a an ellipse [18]. In the case of density approximation, one can use a homotopy in combination with a deviation measure to obtain a good approximation of the posterior. The measure of deviation is used to find such optimal approximation of the posterior but its solution may lead to the problem of local minima. The homotopy helps to circumvent this problem by tracking the optimum while the posterior is continuously approximated. Approaching the filtering problem with homotopy is also called progressive filtering [12], [13]. Additionally, homotopies are applied in different use cases such as obtaining optimal posterior parameters [9] or filtering in polynomial systems [14]. In section III-A the mathematical side is shortly introduced which is then followed by the general idea of using a homotopy for the purpose of filtering III-B. Lastly, section III-C will derive one use of homotopy in context of filtering.

A. Mathematical Background of Homotopy

The mathematical definition for a homotopy is a continuous function $H : X \times [0, 1] \to Y$ mapping a function $f(x)$ from the topological space X to the function $g(x)$ from topological

space Y in such way that $H(x,0) = g(x)$ and $H(x,1) = f(x)$ [18].

For example, let $f(x) = g(x)l(x)$ which is the posterior in filtering context. By using the homotopy $H(x,\lambda) = g(x)l(x)^\lambda$ the prior $g(x)$ is gradually transformed to the posterior. With $\lambda = 0$ the homotopy reduces to the prior whereas $\lambda = 1$, the homotopy is equal to the posterior [7]. The parameter λ is also called the homotopy-parameter [18].

Based on the example above, the advantage of using a homotopy is that given the known prior $H(x,0) = g(x)$, the continuous deformation of $g(x)$ by increasing the parameter λ results in the desired posterior $f(x)$.

B. General Procedure

The general procedure of using homotopy for the approximation of the posterior follows four steps [12]:

1) Formulate a homotopy $\tilde{f}^e(\underline{x}, \lambda)$ which is a parameterized version of the true density $\tilde{f}^e(\underline{x})$. As in the section of the mathematical background III-A the homotopy-parameter λ handles the progression between a starting density ($\lambda = 0$) and the posterior $\tilde{f}^e(\underline{x})$ in case for $\lambda = 1$.
2) Define a density $f^e(\underline{x}, \underline{\theta})$ with a λ dependent parameter vector $\underline{\theta}$ which will approximate $\tilde{f}^e(\underline{x}, \lambda)$.
3) Define a measure of deviation $G(\underline{\theta}, \lambda)$ between $f^e(\underline{x}, \underline{\theta})$ and $\tilde{f}^e(\underline{x}, \lambda)$.
4) An ordinary differential equation (ODE) is derived in order to calculate the parameter $\underline{\theta}(\lambda)$. The ODE form itself can vary but in a general case it is similar to

$$\underline{b}(\underline{\theta}, \lambda) = \mathbf{A}(\underline{\theta})\underline{\dot{\theta}}(\lambda). \qquad (9)$$

As λ approaches 1, the solution of the ODE (9) provides the optimal parameter $\underline{\theta}$ for the density $f^e(\underline{x}, \underline{\theta})$ and thus, an optimal approximation of $\tilde{f}(\underline{x})$ is found. The concrete result of the steps described may vary depending on the chosen densities and deviation measures.

C. Approximation of the Posterior Parameters via Homotopy

The first step in the approximation of the posterior density is the formulation of a homotopy $\tilde{f}^e(\underline{x}, \lambda)$. Its properties with regard to the λ-parameter are i) $f^e(\underline{x}, 0) = g(\underline{x})$ and ii) $\tilde{f}^e(\underline{x}, 1) = \tilde{f}^e(\underline{x}) \propto f^L(\underline{x})f^p(\underline{x})$. With respect to the choice of $g(\underline{x})$, different approaches are feasible. [12] states that $g(\underline{x})$ is an arbitrarily chosen and easy to approximate density. On the contrary, [9] specifies $g(\underline{x})$ to be the prior $f^p(\underline{x})$. The advantage for using the prior density as $g(\underline{x})$ is coupled with the choice of the starting approximation density $f^e(\underline{x}, \underline{\theta})$. If both densities have the same parameterization class, this is ensured by setting $f^p(\underline{x}) = f^e(\underline{x}, \underline{\theta})$, the posterior itself has the same parameterization class. The parameterization of the prior is yielded by the preceding prediction step in course of the estimation procedure. Additionally, the deviation measure for $\lambda = 0$ is fortunately minimized.

As far as the homotopy formulation concerned, the remaining part of this seminar uses the following equation [9]:

$$\tilde{f}^e(\underline{x}, \lambda) = f^p(\underline{x})f^L(\underline{x})^\lambda. \qquad (10)$$

However, it should be noted that two similar versions of homotopy equation could further be considered. In [7] the homotopy is a log-version of (10) resulting in $\log(\tilde{f}^e(\underline{x}, \lambda)) = \log(f^p(\underline{x})) + \lambda \log(f^L(\underline{x}))$. Thus, the multiplication of the prior and the corresponding likelihood is replaced with a sum of prior and likelihood which reduces the complexity of integral based deviation measures.

On the other hand, [10] and [12] use a parameterized version of the likelihood λ resulting in $\tilde{f}^e(\underline{x}, \lambda) = f^p(\underline{x})\tilde{f}^L(\underline{x}, \lambda)$ and the homotopy-parameter λ. The parameterized likelihood $\tilde{f}^L(\underline{x}, 0)$ is defined to be a simple approximable likelihood density and $\tilde{f}^L(\underline{x}, 1)$ the normal likelihood $\tilde{f}^L(\underline{x})$. However, an additive noise term of the measurement model $\underline{y} = \underline{h}(\underline{x}) + \underline{v}$ is assumed and as a consequence, the likelihood can be formulated with help of the noise term $\tilde{f}^L(\underline{x}, \lambda) = f^v(\underline{y} - \underline{h}(\underline{x}), \lambda)$. As a result of this, a more convenient representation for the homotopy is established. Similar approaches are found in [13] and [11].

In the following step, a $f^e(\underline{x}, \underline{\theta})$ needs to be specified for the approximation. As described above, [9] closely couples the choice of $f^e(\underline{x}, \underline{\theta})$ with the density type of the prior. For a more general case, it is possible to choose $f^e(\underline{x}, \underline{\theta})$ independent from the density type of the prior density.

The work in [12] and [10] use a Gaussian mixture but with a scalar and a two-dimensional state respectively. A general Gaussian Mixture with a state vector $\underline{x}$ is given by

$$f^e(\underline{x}, \underline{\theta}) = \sum_{i=1}^{L} w^i \mathcal{N}(\underline{x} - \underline{\hat{x}}^i, \mathbf{C}_{x,i}), \qquad (11)$$

where $\mathcal{N}(\underline{x}, \mathbf{C})$ is a Gaussian with mean vector $\underline{\hat{x}}$ and variance matrix $\mathbf{C}_{x,i}$. The weights fulfil $w_i > 0$ and $\sum_{i=0}^{L} w_i = 1$. An advantage of using Gaussian mixtures (11) is the possibility of structural adaptation of the mixture components. If the given normalized measure of deviation $G_N(\underline{\theta}, \lambda)$ is not in a given range of $G_N^L < G_N(\underline{\theta}, \lambda) < G_U^L$, the critical mixture components are adjusted [12].

A different approach for selecting $f^e(\underline{x}, \underline{\theta})$ is stated in [13]. It begins with a Gaussian representation of $f^e(\underline{x}, \underline{\theta})$ but later uses an approximation via Dirac mixtures given by

$$f^e(\underline{x}, \underline{\theta}) = \sum_{i=1}^{L} w^i \delta(\underline{x} - \underline{\hat{x}}^i), \qquad (12)$$

with the Dirac delta function $\delta(\cdot)$, weights $w^i > 0$ and $\sum_{i=0}^{L} w^i = 1$ The reason for using an alternative representation is that the solution for the derived ODE with a Gaussian density is not always feasible and by using Dirac mixtures, the integration in the ODE is transformed to a summation. Additionally, the Dirac Mixture $f^e(\underline{x}, \underline{\theta})$ incorporates intermediate representations of the true posterior due to the progressive homotopy approach. Thus, it covers only the important regions of the state space which leads to the use of all its Dirac components [13]. If instead the prior is approximated by a Dirac mixture, the calculation of the posterior would lead to sample degeneration for large λ.

The choice of a specific $f^e(\underline{x}, \underline{\theta})$, e.g a Gaussian Mixture (11) or a Dirac Mixture (12), lead to different approaches as described previously.

With regard to the measure of deviation $G(\underline{\theta}, \lambda)$, one option is to use the squared integral deviation

$$G_{SI}(\underline{\theta}, \lambda) = \frac{1}{2} \int (\tilde{f}^e(\underline{x}) - f^e(\underline{x}))^2 \, d\underline{x}, \qquad (13)$$

or the Kullback-Leibler-divergence

$$G_{KL}(\underline{\theta}, \lambda) = \int \tilde{f}^e(\underline{x}) \log\left(\frac{\tilde{f}^e(\underline{x})}{f^e(\underline{x})}\right) \, d\underline{x}. \qquad (14)$$

The first step in deriving the ODE is to take the derivative of $G(\underline{\theta}, \lambda)$ with respect to $\underline{\theta}$ as it is necessary for an optimal parameter $\underline{\theta}$

$$G_{\underline{\theta}}(\underline{\theta}, \lambda) = \frac{\partial G(\underline{\theta}, \lambda)}{\partial \underline{\theta}} = 0. \qquad (15)$$

Based on the tracking paradigm of the optimal parameter $\underline{\theta}$, one need to take the total derivative of the equation (15) with respect to λ

$$\begin{aligned} \frac{\partial G_{\underline{\theta}}(\underline{\theta}(\lambda), \lambda)}{\partial \lambda} &= \frac{\partial G_{\underline{\theta}}(\underline{\theta}(\lambda), \lambda)}{\partial \underline{\theta}} \frac{\partial \underline{\theta}(\lambda)}{\partial \lambda} + \frac{\partial G_{\underline{\theta}}(\underline{\theta}(\lambda), \lambda)}{\partial \lambda} \\ &= G_{\underline{\theta}\underline{\theta}}(\underline{\theta}(\lambda), \lambda) \, \dot{\underline{\theta}}(\lambda) + G_{\underline{\theta}\lambda}(\underline{\theta}(\lambda), \lambda). \end{aligned} \qquad (16)$$

After taking the derivative, the equation (16) can be rearranged to solve for $\dot{\underline{\theta}}(\lambda)$ which leads to the final ODE

$$\dot{\underline{\theta}}(\lambda) = -G_{\underline{\theta}\lambda}(\underline{\theta}(\lambda), \lambda) G_{\underline{\theta}\underline{\theta}}(\underline{\theta}(\lambda), \lambda)^{-1}, \qquad (17)$$

called homotopy differential equation (HDE) [9].

In context of a specific deviation measure, the two step derivation with regard to $\underline{\theta}$ and λ is imitated. The work in [12] formulates the ODE with the linearized deviation measure (13) whereas [9] uses slightly different procedure. It firstly approximates the HDE (17) with help of two different methods which is then further simplified by a given deviation measure such as (13) or (14).

Beside using special deviation measures for probability densities and its formulation into an ODE, the work in [13] utilizes moment matching for a scalar assumed Gaussian density $f^e(x)$. The moment matching for the first three moments is given by

$$\int \underline{m} f^e(x, \underline{\theta}) \, dx = \int \underline{m} \tilde{f}^e(x, \lambda) \, dx, \qquad (18)$$

with $\underline{m} = \begin{bmatrix} 1, x, x^2 \end{bmatrix}$. Based on (18), the derivative with respect to λ is taken. Once the terms are rearranged, a result similar to (17) is obtained.

The last step is to use an algorithm for solving the given ODE with respect to the homotopy-paramter λ. A typical method for solving an ODE is the Euler method.

IV. Homotopy Continuation and particle Filter

The field of particle filter is a further filter technique for recursive state estimation. Its basic idea is to sample weighted points called particles from the posterior and then use these points and the corresponding weights to compute $f^e(\underline{x})$ [1]. As mentioned in section I, importance sampling is one central paradigm in particle filters. Since the posterior density is most likely a complex density from which the samples can not be directly taken, a different but very similar density is leveraged for the sampling procedure. The different density is called importance density or proposal density [6], [4]. Compared to the approach described in section III-C, the focus of importance sampling is the usage of particles.

In order to solve the problem of sample degeneracy (see section II), either a good importance density or a smart resampling of the particles can be used. Besides these general approaches of mitigating the sample degeneracy, an homotopy-based approach is also possible [4], [2], [8]. These methods gradually incorporate the flow of particles into the approximation of the posterior density. This flow is controlled by the homotopy-parameter λ.

A. Sequential Importance Sampling for Filtering

The sequential importance sampling (SIS) is a fundamental filter approach on which further techniques are based on [1]. For the kth time step, the goal is to create particles $\hat{\underline{x}}^i$ from an importance density $q(\underline{x})$. With regard to the Markov property discussed in section II, one can recursively formulate the importance density as

$$q(\underline{x}_k|\underline{Y}_k) = q(\underline{x}_k|\underline{x}_{k-1}, \underline{y}_k) q(\underline{x}_{k-1}|\underline{Y}_{k-1}) \qquad (19)$$

with the particles $\hat{\underline{x}}_k^i \sim q(\underline{x}_k|\underline{Y}_k)$ sampled from the importance density [1]. The weights w_k^i for the particles $\hat{\underline{x}}_k^i$ are proportional to the quotient of the posterior and the importance density [3]. Consequently, the weights can be calculated via

$$\begin{aligned} w_k^i &\propto \frac{p(\hat{\underline{x}}_k^i|\underline{Y}_k)}{q(\hat{\underline{x}}_k^i|\underline{Y}_k)} \\ &\propto \frac{p(\underline{y}_k|\hat{\underline{x}}_k^i) p(\hat{\underline{x}}_k^i|\hat{\underline{x}}_{k-1}^i) p(\hat{\underline{x}}_{k-1}^i|\underline{y}_{k-1})}{q(\hat{\underline{x}}_k^i|\hat{\underline{x}}_{k-1}^i, \underline{y}_k) q(\hat{\underline{x}}_{k-1}^i|\underline{Y}_{k-1})} \\ &= w_{k-1}^i \frac{p(\underline{y}_k|\hat{\underline{x}}_k^i) p(\hat{\underline{x}}_k^i|\hat{\underline{x}}_{k-1}^i)}{q(\hat{\underline{x}}_k^i|\hat{\underline{x}}_{k-1}^i, \underline{y}_k)}. \end{aligned} \qquad (20)$$

Finally, one can formulate an approximation of the posterior $p(\underline{x}_k|\underline{Y}_k)$ via a Dirac mixture similar to the equation (12).

The SIS procedure begins with the initialization of equally weighted samples that are taken from the prior density. For the kth time step, one updates the particles with help of the importance density (19) and then adjusts the particle weights with equation (20). These predictions and updates of particles are then repeated for the subsequent time steps.

B. Homotopy and Importance Sampling

In comparison to the homotopy in the approximation of the posterior parameters III-C, the usage in importance sampling leverages a homotopy to create a well suited importance density. This importance density is used to approximate the posterior density. However, the basic idea of the homotopy remains the same: The closer the homotopy-parameter λ gets to one, the closer will be the approximation of the posterior to the true posterior which at the end leads to $\tilde{f}^e(\underline{x}, 1) = \tilde{f}^e(\underline{x})$.

Moreover, the method proposed in [2] and [4] describes a similar two step method derived in IV-A. Firstly, it predicts the particles and secondly, updates the particle weights. In addition, it combines the importance sampling with a homotopy.

The homotopy is leveraged in context of the optimal importance density (OID) which is the posterior density itself [2], [4]. An exact sampling of the posterior density is not possible and thus, an approximation of the OID is necessary which is achieved by introducing the homotopy. It is used to create interim importance densities at every progression step of the homotopy-parameter λ. The interim importance densities are an approximation of the interim posteriors $\tilde{f}^e(\underline{x}, \lambda)$. At the end of the procedure, the last interim importance density leads to the desired approximation of the posterior $\tilde{f}(\underline{x})$. With regard to the formulation of the homotopy that gradually forms the prior to the posterior, both methods state the homotopy as the equation (10).

As previously stated, the first step is concerned with updating the particles. In [2] the update of the particles is based on a homotopy importance density which is given by

$$q_{t,\lambda_\tau}(\underline{x}_{t,\lambda_\tau}|\underline{x}_{t-1}) \propto f^L(\underline{x}_{t,\lambda_\tau})^{\lambda_\tau} f(\underline{x}_{t,\lambda_\tau}|\underline{x}_{t-1}), \qquad (21)$$

with τ being the τth progression step. If (21) has not an exact analytical solution (e.g. a nonlinear model), [2] suggests to use a local approximation of the equation. On the contrary, the linear prediction model in [4] uses the preceding particles to update the particles:

$$\underline{x}_{\tau+1}^{(i)} = \mathbf{F}_\tau \begin{bmatrix} \underline{x}_\tau^{(i)} \\ \underline{x}_{\tau-1}^{(i)} \end{bmatrix} . \qquad (22)$$

The $\mathbf{F}$ is a transition matrix that adjusts the particles $\underline{x}_\tau^{(i)}$ and $\underline{x}_{\tau-1}^{(i)}$ via coefficients $T = \frac{\Delta\lambda_\tau}{\Delta\lambda_{\tau-1}}$ [4]. Based on the updated particles (22), the importance density needed for the next step is derived via moment matching [4]. This aspect of using the particles to derive the importance density instead of using a homotopy to track the importance density is the fundamental difference compared to the approach in [2].

With regard to adjusting the particle weights (see [2] and [4] for a detailed view), both approaches use the basic principle of setting the new weights proportional to the quotient of the posterior and the importance density similar to equation (20). The main difference is that the posterior is replaced with the interim posterior of the homotopy progression. After updating the particle weights, the weights are normalized and the $\tilde{f}^e(\underline{x}, \lambda_\tau)$ can be approximated with

$$f^e(\underline{x}, \lambda_\tau) = \sum_{i=1}^{L} w^{(i)} \delta(\underline{x} - \underline{x}_\tau^{(i)}) .$$

After the particles and the corresponding weights are updated, the two step procedure is repeated until $\lambda_\tau = 1$. The start of the overall method is given with the particles sampled from the prior density so that we gradually form the particles from the prior to the posterior.

V. Conclusion

This seminar elucidated the usage of homotopy in context of nonlinear filters by introducing it to two use cases. The overall goal is to approximate the posterior density. A homotopy is a mathematical method that continuously deform a given function $f(x)$ to a function $g(x)$ via the homotopy-parameter

λ. In case of filtering, homotopies are used to gradually include the newest measurement y_k into the posterior and by doing so obtain interim representations of the posterior.

The first case that leverages homotopy is the parameterized approximation of the posterior using a measure of deviation between the true posterior and its approximation density. Traditional optimization techniques regarding the minimization of the deviation measurements may lead to suboptimal parameters due to the local minima problem. The homotopy is used in such way that it allows to track the optimal parameter while the prior density is more and more transformed to the posterior by using an ODE.

The second case is concerned with the importance sampling in context of particle filters. In contrast to the first use case, the importance sampling results in a discrete approximation of the posterior with particles and corresponding weights. A traditional approach is the sequential importance sampling which however suffers from particle degeneracy. For this problem, a homotopy is leveraged to get an interim importance densities from which the particles that approximate the posterior are obtained. The specific exemplary method discussed in the seminar uses a Gaussian assumption for the prior and the posterior whereas the approach described in section III does not have such assumptions.

References

[1] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, February 2002.

[2] Pete Bunch and Simon Godsill. Particle filtering with progressive Gaussian approximations to the optimal importance density. In *2013 5th IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, pages 360–363, December 2013.

[3] Olivier Cappe, Simon J. Godsill, and Eric Moulines. An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo. *Proceedings of the IEEE*, 95(5):899–924, May 2007.

[4] Christof Chlebek, Jannik Steinbring, and Uwe D. Hanebeck. Progressive Gaussian filter using importance sampling and particle flow. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 2043–2049, July 2016.

[5] D. Crisan and A. Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50(3):736–746, March 2002.

[6] F. Daum. Nonlinear filters: Beyond the Kalman filter. *IEEE Aerospace and Electronic Systems Magazine*, 20(8):57–69, August 2005. Conference Name: IEEE Aerospace and Electronic Systems Magazine.

[7] Fred Daum and Jim Huang. Nonlinear filters with log-homotopy. In Oliver E. Drummond and Richard D. Teichgraeber, editors, *Optical Engineering + Applications*, pages 669918–669918–15, San Diego, CA, September 2007.

[8] Fred Daum and Jim Huang. Particle flow for nonlinear filters, Bayesian decisions and transport. In *Proceedings of the 16th International Conference on Information Fusion*, pages 1072–1079, July 2013.

[9] Jonas Hagmar, Mats Jirstrand, Lennart Svensson, and Mark Morelande. Optimal parameterization of posterior densities using homotopy. In *14th International Conference on Information Fusion*, pages 1–8, July 2011.

[10] U.D. Hanebeck and O. Feiermann. Progressive Bayesian estimation for nonlinear discrete-time systems:the filter step for scalar measurements and multidimensional states. In *42nd IEEE International Conference on Decision and Control (IEEE Cat. No.03CH37475)*, volume 5, pages 5366–5371 Vol.5, December 2003.

[11] Uwe D Hanebeck. PGF 42: Progressive Gaussian Filtering with a Twist. page 8.

[12] Uwe D. Hanebeck, Kai Briechle, and Andreas Rauh. Progressive Bayes: A new framework for nonlinear state estimation. In Belur V. Dasarathy, editor, *AeroSense 2003*, page 256, Orlando, FL, April 2003.

[13] Uwe D. Hanebeck and Jannik Steinbring. Progressive Gaussian filtering based on Dirac Mixture approximations. In *2012 15th International Conference on Information Fusion*, pages 1697–1704, July 2012.

[14] Marco F Huber and Uwe D Hanebeck. Gaussian Filtering for Polynomial Systems Based on Moment Homotopy. page 8.

[15] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[16] S.J. Julier and J.K. Uhlmann. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92(3):401–422, March 2004.

[17] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, mar 1960.

[18] Shijun Liao. *Homotopy Analysis Method in Nonlinear Differential Equations*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.