

Nonlinear Filtering with Homotopy Continuation

Marcel Hiltcher

Abstract—Non-linear filtering with use of ordinary differential equation.

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 limitation. 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 systems, the corresponding system and measurement model are used for the necessary inference. A system model describes the relation of the state changes where as 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 priori 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 a probability densities conditioned on the measurements and the system inputs in the former case.

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

An additional property of systems is the linearity or non-linearity which respectively leads to different strategies regarding the previous mentioned estimation procedure. In the linear case, the well known Kalman Filter [14] leads to an optimal solution for the estimation problem and includes a recursive approach. However, the non-linearity of a system impede a closed form solution of the priori and posterior distribution. The reasons 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) [13] or the Extended Kalman Filter (EKF) [12].

Next to the Kalman based filters, many other approaches are introduced in scientific research [4]. A broad field of recursive methods are particle filters approximating the density by weighted samples or rather particles [1], [3]. The sampling follows different paradigm like the importance sampling or sequential Monte Carlo [2]. The advantage of such filters is that they do not depend the posterior to be closely approximated by a gaussian distribution [2] like the case in the UKF.

Beside the particle filter, the use of homotopy continuation methods is a further field of non-linear 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 [6]. The use of homotopy progressive methods are introduced in [9] and [4]. However, the latter combines it with particle flow where as the former introduce a more general approach to non-linear state estimation.

The focus of this seminar are homotopy based filters in a non-linear setting. Section II will formerly define the foundation of the estimation problem followed by the general approach of homotopy based filtering in section III. Next, Section IV will exemplary present the intersection of homotopy and particle filter and lastly, section V will draw a conclusion.

II. PROBLEM FORMULATION

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

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

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

where equation (1) defines the system model where as equation (2) defines the measurement model. In equation (1), the \underline{a}_k describes the non-linear relation of state vector \underline{x}_k and the system input \hat{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 $f_k^w(w_k)$. With respect to the state vector \underline{x}_k and the measurement vector \underline{y}_k , the non-linear 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 donates 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, refereed to as the likelihood probability, are defined as:

$$\underline{x}_{k+1} \sim p(\underline{x}_{k+1} | \underline{x}_k, \hat{u}_k), \quad (3)$$

$$\underline{y}_k \sim p(\underline{y}_k | \underline{x}_k). \quad (4)$$

For simplicity reasons, the \hat{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 property 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 donate 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

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

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:

$$\begin{aligned} p(\underline{x}_{k+1} | \underline{Y}_{k+1}) &= p(\underline{x}_{k+1} | \underline{Y}_k, \underline{y}_{k+1}) \\ &= \frac{p(\underline{y}_{k+1} | \underline{x}_{k+1}, \underline{Y}_k)}{p(\underline{y}_{k+1} | \underline{Y}_k)} \\ &= \frac{p(\underline{y}_{k+1} | \underline{x}_{k+1})}{p(\underline{y}_{k+1} | \underline{Y}_k)} \\ &= \frac{p(\underline{y}_{k+1} | \underline{x}_{k+1}) p(\underline{x}_{k+1} | \underline{Y}_k)}{p(\underline{y}_{k+1} | \underline{Y}_k)}. \end{aligned} \quad (6)$$

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 $p(\underline{x}_{k+1} | \underline{Y}_{k+1}) \approx p(\underline{y}_{k+1} | \underline{x}_{k+1}) p(\underline{x}_{k+1} | \underline{Y}_k)$.

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

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

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

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 transitional density and the likelihood can also be reduced to $f_{k+1}^T(\underline{x}_{k+1})$ and $f_{k+1}^L(\underline{x}_{k+1})$.

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 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 an re-approximation after each filtering step because the type of the density may change.

One method is to use a measure of deviation $G(\theta)$ between the approximation probability density $f^e(\underline{x}, \theta)$ - 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 θ . Though, numerical solution of the optimization problem may lead to local minima which in return does not ensure an optimal parameter for the approximation density $f^e(\underline{x}, \theta)$.

III. NON-LINEAR FILTERING WITH HOMOTOPY CONTINUATION

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 an ellipse [15]. In the case of density approximation, one can use a homotopy in combination of 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 [9], [10]. Additionally, homotopies are applied in different use cases such as obtaining optimal posterior parameters [6] or filtering in polynomial systems [11]. 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 the measurement update with the use of a homotopy.

A. Mathematical background

The mathematical definition for a homotopy is a continuous function $H : X \times [0, 1] \rightarrow Y$ 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 $H(x, 1) = f(x)$ [15].

For example, let the function be $f(x) = g(x)l(x)$ - 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 where as $\lambda = 1$, the homotopy is equal to the posterior [5]. The parameter λ is also called the homotopy-parameter [15].

Based on the example above, the advantage of using a homotopy is that given the known prior $H(x, 0) = g(x)$, the continuously 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 [9]:

- 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 $\underline{\theta}(\lambda)$:

$$\dot{\underline{\theta}}(\lambda) = \mathbf{A}\underline{\theta}(\lambda) \quad (9)$$

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

C. Approximation of the Posterior Density 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 firstly, $\tilde{f}^e(\underline{x}, 0) = g(\underline{x})$ and secondly, $\tilde{f}^e(\underline{x}, 1) = \tilde{f}^e(\underline{x}) \approx f^L(\underline{x})f^p(\underline{x})$. With respect to the choice of $g(\underline{x})$, different approaches are feasible. [9] states that $g(\underline{x})$ is an arbitrary chosen and easy approximable density. On the contrary, [6] 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 [6]:

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

However, it should be noted that two similar versions of homotopy equation could further be considered. In [5] 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, [7] and [9] 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 [10] and [8].

In the following step, a $f^e(\underline{x}, \underline{\theta})$ needs to be specified for the approximation. As described above, [6] 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 [9] and [7] use a Gaussian Mixture but with a singular and a two-dimensional state \underline{x} respectively. A general Gaussian Mixture with a state vector \underline{x} is given by

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

where \mathcal{N} is a Gaussian with mean vector $\hat{\underline{x}}$ and variance matrix $\mathbf{C}_{x,i}$. The weights fulfill $w_i > 0$ 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 [9].

A different approach for selecting $f^e(\underline{x}, \underline{\theta})$ is stated in [10]. It begins with a gaussian representation of $f^e(\underline{x}, \underline{\theta})$ but later uses an approximation via Dirac Mixtures (see (12)). 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 representation of the 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 component [10]. If instead the prior is approximated by a Dirac Mixtures, it would lead to sample degeneration for large λ . A Dirac Mixture is stated by

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

with weights $w_i > 0$ and $\sum_{i=0}^L w_i = 1$. 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 options is the squared integral deviation

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

or the Kullback-Leibler-divergence

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

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

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

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

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

However, the equation (16) does not quite resembles the ODE given in (9) but a small rearrangement of the variables leads to

$$G_{\underline{\theta}\lambda}(\underline{\theta}(\lambda), \lambda) = G_{\underline{\theta}\underline{\theta}}(\underline{\theta}(\lambda), \lambda) \dot{\underline{\theta}}(\lambda), \quad (17)$$

which is equal to (9) except of the variable names.

In context of a specific deviation measure, the two step derivation targeting $\underline{\theta}$ and λ is imitated. The work in [9] derives the ODE with the linearized deviation measure (13) whereas [6] uses slightly different procedure. It first rearranges (17) to solve for $\dot{\underline{\theta}}(\lambda)$ which is called a homotopy differential equation (HDE). Afterwards, it approximates the HDE 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 [10] utilizes moment matching for a scalar assumed gaussian density $f^e(x)$ in order derive an ODE like (9). In case of a state vector \underline{x} the moment matching for the first three moments is given by

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

with $\underline{m} = [1, \underline{x}, \underline{x}^2]$.

After all necessary equations are defined, the last step is to use an algorithm for solving the given ODE with respect to the homotopy-paramter λ - e.g Euler method.

- what is the homotopy, compare to daum 2007 with log
- measures of deviation

IV. HOMOTOPY CONTINUATION AND PARTICLE FILTER

V. CONCLUSION

- Compare findings of different methods

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] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.

- [7] 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.
- [8] Uwe D Hanebeck. PGF 42: Progressive Gaussian Filtering with a Twist. page 8.
- [9] 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.
- [10] 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.
- [11] Marco F Huber and Uwe D Hanebeck. Gaussian Filtering for Polynomial Systems Based on Moment Homotopy. page 8.
- [12] Andrew H Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.
- [13] S.J. Julier and J.K. Uhlmann. Unscented Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92(3):401–422, March 2004.
- [14] R. E. Kalman. A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(1):35–45, mar 1960.
- [15] Shijun Liao. *Homotopy Analysis Method in Nonlinear Differential Equations*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012.