

Transport Map Coupling Filter For State-Parameter Estimation

Jan Grashorn^{a,b}, Matteo Broggi^a, Ludovic Chamoin^{b,c}, Michael Beer^{a,b,d,e,f}

^a*Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany*

^b*International Research Training Group 2657, Leibniz University Hannover, Hannover, Germany*

^c*Université Paris-Saclay, CentraleSupélec, ENS Paris-Saclay, CNRS,*

Laboratoire de Mécanique Paris-Saclay (LMPS), Gif-sur-Yvette, France

^d*Institute for Risk and Uncertainty, University of Liverpool, Liverpool, England*

^e*International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China*

Abstract

Many dynamical systems are subjected to stochastic influences, such as random excitations, noise and unmodeled behavior. Tracking the system's state and parameters based on a physical model is a common task for which usually filtering algorithms, e.g. Kalman filters and their non-linear extensions, are used. Many of these filters however use assumptions on the transition probabilities or the covariance model which can lead to inaccuracies in non-linear systems. We will show here the application of a stochastic coupling filter that is able to approximate arbitrary transition densities under non-Gaussian noise. The filter is based on so-called transport maps which couple the approximation densities to a user-chosen reference density and thus allow for straight-forward sampling and evaluation of probabilities.

Keywords: transport maps, Kalman filters, state estimation, parameter estimation, Bayesian updating

1. Introduction

Dynamical systems are at the heart of many engineering disciplines. Knowledge about the system's internal parameters is important in these to ensure safe operation conditions, to predict future behavior or to gain insight on the system's condition, i.e. for damage detection and health monitoring. This knowledge is obtained or updated through measurements from some system components, however, in many cases it is not possible to directly measure the system's state. Moreover, the measurements are polluted by inherent sensor noise which needs to be filtered out. In the past, many algorithms have been developed towards this goal. The most prominent is the Kalman filter, which, due to its optimality in linear systems subjected to Gaussian noise, still is rightfully used in many applications. For non-linear systems there have been many endeavors towards extending the Kalman filter's capacity, however, these often rely on the Gaussian noise assumption as an underlying uncertainty model. In the context of health and systems monitoring, Kalman filters and their derivatives are often used to track parameter changes over time or to extract knowledge about damage states from noisy measurements. In a more general sense this is also known as sequential updating (Vanik et al., 2000). As an example, (Ghanem and Ferro, 2006) used an ensemble Kalman filter (EnKF) to estimate damage in a four-story shear building by tracking the changes in system parameters over time. (Erazo et al., 2018) showed the application of Kalman filters to structural health monitoring using a decoupling between environmental changes and parameter changes. (Xie and Feng, 2011) applied an unscented Kalman filter (UKF) to non-linear structural identification. Similarly, (Lei et al., 2019) proposed an UKF to identify non-linear systems under unknown inputs and (Diaz et al., 2023) used an UKF and error estimation techniques to update a structure subjected to vibrations from a shaking table.

Sequential updating offers the possibility to also track a system's changes in real-time. Here the effort lies in fast and efficient sampling methods such that the data can be assimilated as soon as it is measured. Towards this goal, (Chatzi and Smyth, 2012) introduced mutation in particle filter applications to tackle the issue of the samples collapsing to only few states. Particle filters were also proposed for fault detection and isolation because

of their performance in highly non-linear settings by (Orchard and Vachtsevanos, 2009). (Lye et. al, 2023) proposed the sequential ensemble Monte Carlo sampler, which increases robustness and efficiency of sampling from complicated probability distributions.

We will show here a stochastic coupling filter that can be used as a replacement for general sequential updating settings where the goal is to estimate states and parameters under noisy measurements. The filter has no underlying assumption on the model or noise uncertainty. The rest of the paper is structured as follows: In section 2 we will introduce the general transport map formulation and the underlying theory. Section 3 introduces the filter methodology for state estimation with section 4 expanding it to joint state-parameter estimation. Section 5 shows the application to a Duffing oscillator and section 6 concludes this paper.

2. Transport maps

A transport map is an invertible deterministic coupling between two probability density functions (PDFs), one being the target density of interest π and the other one a user-chosen reference density ρ . This coupling allows for the direct sampling and calculations of integrals on the target PDF since calculations can be performed through change of variables (Spantini et al., 2018)

$$\int_Y f(\mathbf{y})\pi(\mathbf{y})d\mathbf{y} = \int_X f(M(\mathbf{x}))\rho(\mathbf{x})d\mathbf{x} \quad (1)$$

where $M(\cdot) : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is the transport map that transforms samples from the target density to the reference density. The reference density is usually chosen as standard normal or standard uniform, since this choice makes calculations straight forward once the map is found (Spantini et al., 2018). With the push-forward operator

$$M_{\#}\pi = \rho(M(\mathbf{x})) \cdot |\det \nabla M(\mathbf{x})| \quad (2)$$

and the pull-back operator for the inverted map M^{-1}

$$M^{\#}\rho = \pi(M^{-1}(\mathbf{x})) \cdot |\det \nabla M^{-1}(\mathbf{x})| \quad (3)$$

a distance for the approximation $M^{\#}\rho \approx \pi$ can be calculated with the Kullback-Leibler divergence

$$D_{\text{KL}}(M^{\#}\rho || \pi) = \mathbb{E}_{\rho} \left[\log \frac{M^{\#}\rho}{\pi} \right] \quad (4)$$

or, due to invertibility of the map,

$$D_{\text{KL}}(\rho || M_{\#}\pi) = \mathbb{E}_{\rho} \left[\log \frac{\rho}{M_{\#}\pi} \right] \quad (5)$$

The distance can thus either be defined in terms of the reference or the target density. Both options are viable, depending on the use case. If a function for the target density is known, i.e. in Bayesian model updating, where the goal is to approximate the posterior density, the expression in (4) can be used. If only samples from the posterior are known, which is the setting of this paper, the expression in (5) is more useful. Multiple formulations have been proposed for the maps M , i.e. polynomials (Marzouk et al., 2016), tensor trains (Dolgov et al., 2020) or neural networks (Ardizzone et al., 2019). In this context, the latter are also referred to as normalizing flows (Kobyzev et al., 2020). For uniqueness and ease of computation purposes, as well as to fulfil the invertibility criterion, the maps are constructed in a lower-triangular fashion, also known as Knothe-Rosenblatt rearrangement

$$M(\mathbf{x}) = \begin{pmatrix} M^1(x_1) \\ \vdots \\ M^n(x_1, \dots, x_n) \end{pmatrix} \quad (6)$$

where each component $M^i: \mathbb{R}^i \rightarrow \mathbb{R}$ pushes forward the first i components of \mathbf{x} . Moreover, each component is constructed as

$$M(\mathbf{a}^i, x_1, \dots, x_i) = f(\mathbf{a}^i, x_1, \dots, x_{i-1}, 0) + \int_0^{x_i} g(\partial_i f(\mathbf{a}^i, x_1, \dots, x_{i-1}, \bar{x})) d\bar{x} \quad (7)$$

with parameterized functions $f(\mathbf{a}^i, \cdot)$ and a rectifier function $g: \mathbb{R} \rightarrow \mathbb{R}_+$, e.g. the exponential function, that ensures monotonicity in the last argument. For computations we used the MParT framework (MParT Development Team, 2023). The functions $f(\mathbf{a}^i, \mathbf{x}_1, \dots, \mathbf{x}_i)$ are defined as linear combinations of Hermite polynomials $\Psi(\cdot)$

$$f(\mathbf{a}^i, \mathbf{x}_1, \dots, \mathbf{x}_i) = \sum_j \mathbf{a}_j^i \Psi_{\alpha^j}(\mathbf{x}_1, \dots, \mathbf{x}_i). \quad (8)$$

with

$$\Psi_{\alpha^j}(\mathbf{x}_1, \dots, \mathbf{x}_i) = \prod_{k=1}^i \Psi_{\alpha_k^j}(\mathbf{x}_k).$$

The subscript α^j denotes a vector of so-called multi-indices which define the order of the Hermite polynomials. For a more in-depth explanation and derivations the interested reader is referred to (Baptista, 2023) and the references therein.

In order to find the map parameters \mathbf{a} , (2) and (5) are written as

$$D_{KL}(\rho || M_{\#}\pi) = \int_{\mathbf{x}} [\log \rho(\mathbf{x}) - \log \rho(M(\mathbf{a}, \mathbf{x})) - \log(|\det \nabla M(\mathbf{a}, \mathbf{x})|)] \rho(\mathbf{x}) d\mathbf{x}. \quad (9)$$

A smaller distance indicates a better fit between the approximation and the target PDF. If a number of N samples $\hat{\mathbf{x}}$ from the target distance are available, (9) can be transformed into a minimization procedure

$$\arg \min_{\mathbf{a}} \frac{1}{N} \sum_i \left[-\log(\rho(M(\mathbf{a}, \hat{\mathbf{x}}_i))) - \log(|\det \nabla M(\mathbf{a}, \hat{\mathbf{x}}_i)|) \right]. \quad (10)$$

If ρ is chosen as standard normal density, the formulation of (9) is equivalent to normalizing the samples $\hat{\mathbf{x}}$ through the (non-linear) mapping to be distributed according to the standard normal density. The polynomial degree of Ψ in (8) can be changed in order to change the approximation quality, where higher order maps are able to better approximate the target, however, this naturally also increases the computational effort, as the size of \mathbf{a} increases. (10) can be solved by any optimization algorithm, especially those that use gradient information have proven to be effective here. The gradient of (10) is readily available, since all of the expressions are formulated analytically.

3. Coupling Filter for state estimation

As a first step we will summarize here the general approach and the formulation for state estimation of a dynamical system to motivate the application to combined state-parameter estimation. This approach was introduced by (Spantini et al., 2022), where more in-depth derivations and proofs can be found. Consider the state-space system with d states and m measurements at some time k

$$\begin{aligned} \dot{\mathbf{x}}_k &= \mathbf{A}\mathbf{x}_k + \mathbf{B}\mathbf{u}_k \\ \mathbf{y}_k &= \mathbf{C}\mathbf{x}_k + \mathbf{w}_k \end{aligned} \quad (11)$$

where $\mathbf{x} \in \mathbb{R}^d$ are the states (e.g. positions, velocities...) and their derivatives in time, $\mathbf{A} \in \mathbb{R}^{d \times d}$ is the state transition matrix, $\mathbf{B} \in \mathbb{R}^{d \times d}$ is the input matrix, $\mathbf{u} \in \mathbb{R}^d$ are inputs to the system, $\mathbf{y} \in \mathbb{R}^m$ are measurements, $\mathbf{C} \in \mathbb{R}^{m \times d}$ is the measurement matrix and $\mathbf{w} \in \mathbb{R}^m$ is additive random noise that corrupts the measurements, i.e. $\mathbf{w} \sim \pi_{\mathbf{w}}$, where $\pi_{\mathbf{w}}$ is the PDF of \mathbf{w} . Gaussian noise is often used here. The subscript k indicates the discrete timestep at time t_k . Since entries of the state \mathbf{x} are only indirectly observable through measurements the goal is to update the system's state \mathbf{x}_{k+1} at time t_{k+1} based on the information obtained through measurements $\mathbf{y}_{1:k+1}$ up until this point. In most cases, the measurements are also incomplete, i.e. $m < n$. Given some samples $\{\mathbf{x}_k^1, \dots, \mathbf{x}_k^n\}$ at time t_k distributed according to the (assumed known) PDF $\pi_k(\mathbf{x}_k | \mathbf{y}_{1:k})$, samples from the prior distribution $\pi_k(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})$ at time step t_{k+1} can be obtained by propagating \mathbf{x}_k forward in time with integration of (11). By constructing a joint probability for measurements and states $\pi_{\mathbf{y}, \mathbf{x}}$ and letting the reference PDF ρ be the standard normal density, a map $S(\mathbf{y}, \mathbf{x}): \mathbb{R}^{m+d} \rightarrow \mathbb{R}^{m+d}$ can be defined using the push-forward operation from (2) as

$$S_{\#}\pi_{\mathbf{y},\mathbf{x}} = \rho. \quad (12)$$

The map S has the structure

$$S(\mathbf{y}, \mathbf{x}) = \begin{pmatrix} S^1(y_1) \\ \vdots \\ S^{m+d}(y_1, \dots, y_m, x_1, \dots, x_d) \end{pmatrix}. \quad (13)$$

Each component $S^i(\mathbf{y}, \mathbf{x})$ in (13) pushes forward the first i inputs to a marginal distribution of the reference density. Therefore, by definition of the standard normal density, the components are independent and can be computed in parallel. Moreover, since we are only interested in an expression for \mathbf{x} , the first m components can be ignored, giving the final form of $S(\mathbf{y}, \mathbf{x})$

$$S(\mathbf{y}, \mathbf{x}) = \begin{pmatrix} S^1(\mathbf{y}, x_1) \\ \vdots \\ S^d(\mathbf{y}, x_1, \dots, x_d) \end{pmatrix} \quad (14)$$

where each component is a function $S^i(\mathbf{y}, \mathbf{x}_{1:i}): \mathbb{R}^{m+i} \rightarrow \mathbb{R}$.

Samples from $\pi_{\mathbf{y},\mathbf{x}}$ can be obtained by sampling \mathbf{y}_{k+1} from the noise distribution π_W after obtaining prior samples $\hat{\mathbf{x}}_{k+1}$. Each simulated state $\hat{\mathbf{x}}_{k+1}$ is thereby augmented by a simulated measurement $\hat{\mathbf{y}}_{k+1}$ to obtain pairs $(\hat{\mathbf{y}}_{k+1}^i, \hat{\mathbf{x}}_{k+1}^i)$, $i \in [1, \dots, n]$. Note that the posterior is obtained from defining the pairs $(\mathbf{y}_{k+1}, \mathbf{x}_{k+1}^i)$ where \mathbf{y}_{k+1} is the actual measurement at time t_{k+1} and \mathbf{x}_{k+1}^i are the unknown samples from the posterior density $\pi_{k+1}(\mathbf{x}_{k+1} | \mathbf{y}_{1:k+1})$. By definition, these pairs can be pushed to ρ by (12), therefore \mathbf{x}_{k+1}^i can be obtained after finding the map S by first pushing the pairs $(\hat{\mathbf{y}}_{k+1}^i, \hat{\mathbf{x}}_{k+1}^i)$ to the standard normal density, then setting $\hat{\mathbf{y}}_{k+1} = \mathbf{y}_{k+1}$ and inverting the map, i.e.

$$\mathbf{x}_{k+1} = S^{-1}(\mathbf{y}_{k+1}, S(\hat{\mathbf{y}}_{k+1}, \hat{\mathbf{x}}_{k+1})) \quad (15)$$

where the inverse map $S^{-1}(\mathbf{y}_{k+1}, \cdot)$ is defined by finding the input $\tilde{\mathbf{x}}$ to S such that

$$S(\mathbf{y}_{k+1}, \tilde{\mathbf{x}}) = S(\hat{\mathbf{y}}_{k+1}, \hat{\mathbf{x}}_{k+1}) \quad (16)$$

for the known inputs $\mathbf{y}_{k+1}, \hat{\mathbf{y}}_{k+1}, \hat{\mathbf{x}}_{k+1}$. Because of the components' independence, finding the inverse of the map is equivalent to finding the d roots of each scalar function in S^{-1} , this can again be done in parallel. The obtained samples \mathbf{x}_{k+1} from (15) can then be used to update the states for the next time steps in a recursive manner.

An illustration for the filtering process between two time steps is shown in Fig. 1. The blue samples represent the propagated states from time k to time $k+1$, i.e. they are sampled according to $\pi_k(\mathbf{x}_{k+1} | \mathbf{x}_k)$. Red samples indicate the likelihood samples that were drawn from $\pi_k(\hat{\mathbf{y}}_{k+1} | \hat{\mathbf{x}}_{k+1})$. Together, these represent samples (magenta) from the joint PDF $\pi_{\mathbf{y},\mathbf{x}}$ (indicated as contour plot), which can be approximated by the transport map formulation to an arbitrary degree. After obtaining this approximation it can be used to condition the joint PDF on a measurement \mathbf{y}_{k+1} by simply inverting the map, thus reaching the desired posterior PDF $\pi(\hat{\mathbf{x}}_{k+1} | \mathbf{y}_{k+1})$ (red line in Fig. 1).

The advantage of using transport maps to approximate the joint density is that there is virtually no assumption on the structure of the involved probability functions, especially on $\pi_{\mathbf{y},\mathbf{x}}$. As long as the map parameterization is rich enough to describe the involved densities, neither the state transition nor the noise needs to follow Gaussian or linear assumptions, as is the case with many other filtering algorithms.

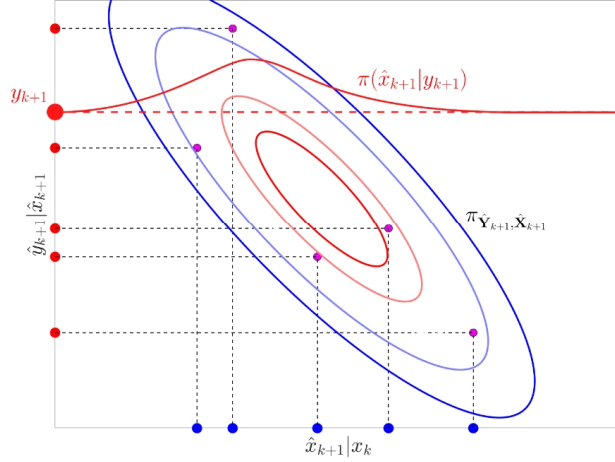


Fig. 1. Illustration of the filtering process from time k to time $k + 1$ through the joint PDF $\pi_{y,x}$.

4. Joint state-parameter estimation using the coupling filter

Many dynamical systems are dependent on some parameters that define their behavior in time. Let $\theta \in \mathbb{R}^{d_\theta}$ be a vector of parameters for the system described in (11), then it can be rewritten as

$$\begin{aligned} \dot{\mathbf{x}}_k &= \mathbf{A}(\theta) \mathbf{x}_k + \mathbf{B}(\theta) \mathbf{u}_k \\ \mathbf{y}_k &= \mathbf{C}(\theta) \mathbf{x}_k + \mathbf{w}_k \end{aligned} \quad (17)$$

where the system's matrices are now dependent on θ . If θ is unknown and subject to identification, the state vector \mathbf{x} can be augmented by simply adding θ as entries in \mathbf{x} , i.e.

$$\mathbf{x} = [x_1 \quad \dots \quad x_d \quad \theta_1 \quad \dots \quad \theta_{d_\theta}]^T. \quad (18)$$

Estimating both states and parameters together follows the same procedure as described in the previous chapter, however the orders of magnitude in each dimension of \mathbf{x} can vary drastically which makes solving the optimization procedure in (10) inefficient. We therefore introduce an additional regularization step before building the map S , which decreases the computational effort without adding too much complexity. Since the standard normal density is used as a reference, the prior pairs $\hat{\mathbf{y}}_{k+1} = (\hat{\mathbf{y}}_{k+1}, \hat{\mathbf{x}}_{k+1})$ and $\tilde{\mathbf{y}}_{k+1} = (\mathbf{y}_{k+1}, \hat{\mathbf{x}}_{k+1})$ can be normalized to a zero-mean distribution with unit covariance by shifting and scaling the samples with

$$\begin{aligned} \hat{\chi}_{k+1} &= \Sigma_{k+1}^{-1} \cdot (\hat{\mathbf{y}}_{k+1} - \tilde{\mathbf{y}}_{k+1}) \\ \tilde{\chi}_{k+1} &= \Sigma_{k+1}^{-1} \cdot (\tilde{\mathbf{y}}_{k+1} - \tilde{\mathbf{y}}_{k+1}) \end{aligned} \quad (19)$$

where $\hat{\chi}_{k+1}$ and $\tilde{\chi}_{k+1}$ denote the normalized samples, Σ_{k+1}^{-1} is the inverse lower-triangular Cholesky factor of the covariance matrix of the prior samples and $\tilde{\mathbf{y}}_{k+1}$ is the mean of the prior samples. Note that the pair $\hat{\mathbf{y}}_{k+1}$ contains the simulated measurements, while the pair $\tilde{\mathbf{y}}_{k+1}$ contains the actual measurements. Both are normalized using the same mean and covariance. The resulting normalized samples more closely resemble the target distribution which decreases the needed effort to find the map, since the distance to the target is smaller. Moreover, the transformation in (19) is linear and easily available, because it only involves calculating the mean and covariance of the prior samples. The objective now is to find normalized posterior samples χ_{k+1} , which can be transformed back to the non-normalized posterior pairs $\mathbf{y}_{k+1} = (\mathbf{y}_{k+1}, \mathbf{x}_{k+1})$ by

$$\mathbf{y}_{k+1} = \Sigma_{k+1} \cdot (\tilde{\chi}_{k+1,1:m}, \chi_{k+1}) + \tilde{\mathbf{y}}_{k+1} \quad (20)$$

where χ_{k+1} is the output of the map inversion from (15). Since $\chi_{k+1} \in \mathbb{R}^d$ and $\mathbf{y}_{k+1} \in \mathbb{R}^{m+d}$ the first m normalized measurements from $\tilde{\chi}_{k+1}$ are used in the back-transformation in (20). The full procedure to estimate states and parameters is summarized below:

- at time t_{k+1} propagate \mathbf{x}_k to \mathbf{x}_{k+1} through the (non-linear) system [equation (17)]
- sample pairs $\hat{\mathbf{y}}_{k+1}$ from $\pi_{\mathbf{y},\mathbf{x}}$
- normalize obtained samples to obtain $\hat{\mathbf{x}}_{k+1}$ and $\tilde{\mathbf{x}}_{k+1}$ [equation (19)]
- build and optimize $S(\hat{\mathbf{x}}_{k+1})$ [equation (14) and (10)]
- calculate inverted map output fixed on the measurements [equation (15)]
- calculate non-normalized posterior samples \mathbf{y}_{k+1} [equation (19)]

4.1. Likelihood oversampling

In order to enhance the efficiency of the filter we propose to oversample the likelihood, i.e. draw more than one sample per propagated state. This increases the information content in the approximation of the joint distribution $\pi_{\mathbf{y},\mathbf{x}}$ and makes it more robust, since the maps are able to better approximate the joint PDF. While drawing samples from the likelihood is computationally inexpensive, this slightly increases the time it takes to compute the transport maps since more samples need to be evaluated. This is however favorable over simulating more system states since this would require more evaluations of the system's equations, which usually is more time-consuming.

5. Numerical example

In order to show the applicability of the coupling filter to non-linear state-parameter estimation, we will show the procedure here on a Duffing oscillator. The equation of motion is

$$\ddot{x} + \delta \dot{x} + \alpha x + \beta x^3 = \kappa \cos(\omega_0 t) \quad (21)$$

with scalar states \ddot{x}, \dot{x}, x that denote acceleration, velocity and position, linear stiffness α , linear damping δ and non-linear stiffness β . In this example we use a forced Duffing oscillator with force amplitude κ and frequency ω_0 . (21) can be written in state-space

$$\begin{bmatrix} \dot{\hat{x}} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} A \\ -\alpha & 1 \\ -\delta & \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix} + \begin{bmatrix} B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ \kappa \cos(\omega_0 t) \end{bmatrix} + \begin{bmatrix} A_{nl} \\ 0 & 0 \\ -\beta & 0 \end{bmatrix} \begin{bmatrix} x^3 \\ \dot{x} \end{bmatrix} \quad (22)$$

with measurements

$$\begin{bmatrix} y \\ \dot{y} \end{bmatrix} = \mathbf{C} \mathbf{x} + w \quad (23)$$

For estimation of the parameters, \mathbf{x} is augmented as $\mathbf{x} = [x, \dot{x}, \alpha, \delta, \beta]^T$ and $\mathbf{C} = [1 \ 0 \ 0 \ 0 \ 0]$, i.e. only the position is measured directly. The noise w is chosen to follow a zero-mean Laplace distribution, since this exhibits much longer tails than a Gaussian distribution. The equation for the Laplace distribution (omitting the mean value) is

$$\pi_w(x) = \frac{1}{2\sigma_w} \exp\left(-\frac{|x|}{\sigma_w}\right) \quad (24)$$

with $\sigma_w = 0.3^2$. We use here $\alpha = -1, \delta = 0.3$ and $\beta = 2$ for the system parameters to be identified, as well as $\kappa = 0.5$ and $\omega_0 = 1.2$ for the input parameters. To show the effect of likelihood oversampling we calculated results for a single and three likelihood samples per state. As initial conditions we set the position and velocity to zero and set the initial guesses for the parameters to be twice the value of the actual parameters to verify that the filter is able to deal with non-optimal initialization. Further, we use a sample size of 20 systems, with the oversampling described as above this leads to 20 and 60 samples to approximate $\pi_{\mathbf{y},\mathbf{x}}$, respectively. We simulate the system in $T = [0,100]$ with time step size $\Delta t = 0.1$. An example for the system behavior and the histogram of the noise is shown in

Fig. 2. The figure shows the system displacement (black line) from $t = 20$ to $t = 40$ and the measurements (red dots) on the left. The right diagram shows the relative occurrence of noise values over the full simulated time.

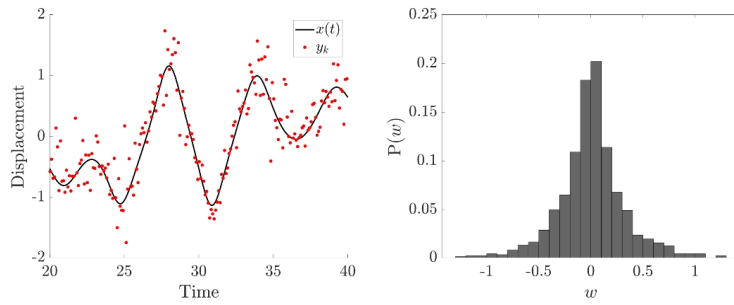


Fig. 2. Example for system behavior and noise

Results of the identification are shown in Fig. 3 and Fig. 4, where the top two graphs show the results for position x and velocity \dot{x} and the bottom row shows the estimations for the three parameters α , δ and β . True values are given as a blue line, all graphs show the estimated posterior PDFs in time shaded in white. In the case without oversampling (Fig. 3) it can be observed that the state and velocity are estimated quite well, although with greater variance than if oversampling is used (Fig. 4). Especially the uncertainty in the non-linear stiffness parameter β is very large and it deviates quite far from the true value around $t = 40$. A similar situation can be observed for the linear stiffness α . The damping δ is estimated with less variance. The estimations for x and \dot{x} also deviate from the true state, however they do not become unstable, which could be a concern in this example since the system behaves slightly chaotic.

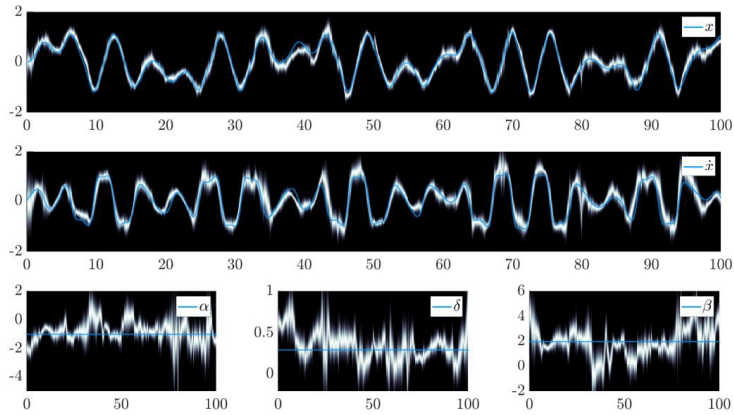


Fig. 3. Results for no oversampling for updating of Duffing-oscillator, figures show the PDF of the states and parameters.

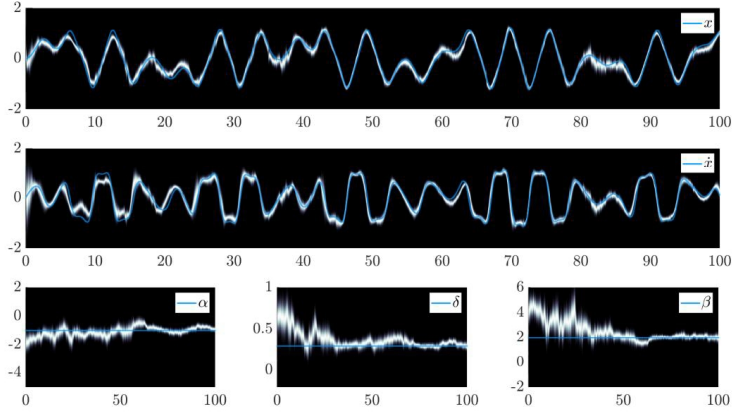


Fig. 4. Results for an oversampling factor of 3 for updating of Duffing-oscillator, figures show the PDF of the states and parameters.

Oversampling the likelihood with three samples per propagated state (Fig. 4) overall results in less variance in the estimations. Position and velocity closely follow the true values and the three parameters converge to the prescribed values. From this example it is evident that an oversampling of the likelihood provides more information about $\pi_{Y,X}$ without drastically increasing the computational burden. To illustrate this further, Fig. 5 shows the computation time for the maps at each time step, with the no oversampling case in red and the oversampling in black. Note that these times do not include the time for the state propagation from step k to step $k + 1$. Since we used 20 samples from the system in both cases the time to propagate the system is identical. It can be seen that, on average, the time for the oversampled maps are longer than the times for maps without oversampling. The overall computation times, including model propagation, were 36.55 s without oversampling and 39.60 s with oversampling. Also note that the computation time in each time step roughly stays the same, which is an important criterion for real-time state-parameter estimation.

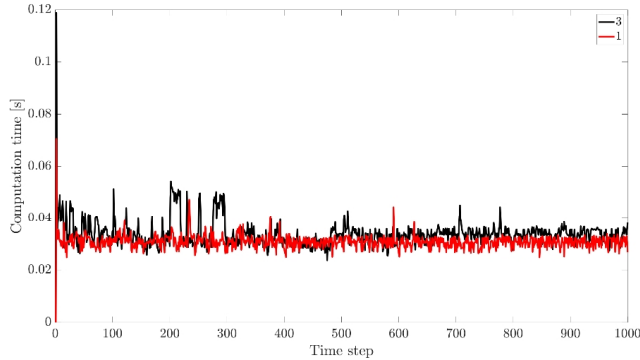


Fig. 5. Map computation time at each time step for oversampling factors of 1 (red) and 3 (black).

The difference between the two cases depends on multiple factors like the map dimensionality, map order, the oversampling factor or the topology of $\pi_{Y,X}$. Increasing the map dimensionality, i.e. increasing the number of estimated states and parameters, or increasing the map order increase the map complexity and therefore the optimization time. If oversampling is used, the maps need to be evaluated more often. The longer computation times in the maps themselves then lead to a larger difference in the evaluation times when the likelihood is oversampled. The same argument is true for the oversampling factor, i.e. more samples from the likelihood lead to more computation time, as we showed in the example in this paper. Generally, a balance should be found between the number of simulated system states \hat{x}_{k+1} and the number of samples from the likelihood \hat{y}_{k+1} . Fig. 1 gives some intuition for this process. In this two-dimensional example a larger number of samples \hat{x}_{k+1} directly correspond to a larger number of samples from $\pi_{Y,X}$ (if one likelihood sample per propagated state is drawn, i.e.

no "undersampling" is performed) which adds information about the uncertainty of parameters and measurement noise at the same time. On the other hand, if the likelihood is oversampled, only information about the noise is added. Oversampling thus increases the robustness against measurement noise to some extent but should also not be exaggerated. At some point there is little gain from increasing the number of samples from the likelihood, however, this would still increase the computation time.

6. Conclusion and outlook

In this contribution a coupling filter for joint state-parameter estimation for non-linear dynamical systems based on transport maps was introduced and evaluated. The filter works by approximating the joint density of states and measurements with so-called transport maps, which give an analytic formulation of the density. This analytic formulation can then be used to condition the joint PDF at any assimilation time step on a measurement, giving an approximation for the posterior density. In the used transport map framework this operation is done by simply inverting polynomial expressions which make the process very efficient. Moreover, since there are no assumptions on the noise distributions or the model, non-Gaussian and non-linear settings can be treated in a straight-forward manner. The presented filter performed well for the state and parameter estimation of a Duffing oscillator with non-optimal initial conditions under non-Gaussian noise. We also evaluated likelihood oversampling, which can help to increase the available information about the measurement noise and make the filter more robust against it. We found that oversampling the likelihood function decreases the overall variance in the estimations.

Acknowledgements

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) for the GRK2657 (grant reference number 433082294) is greatly appreciated.

References

- Ardizzone, L., Kruse, J., Wirkert, S., Rahner, D., Pellegrini, E., Klessen, R., Maier-Hein, L., Rother, C., Köthe, U. 2019. Analyzing Inverse Problems with Invertible Neural Networks. arXiv:1808.04730 [cs.LG]
- Baptista, R., Marzouk, Y., Zahm, O. 2023. On the representation and learning of monotone triangular transport maps. *Foundations of Computational Mathematics*, 1-46.
- Chatzi, E. N., Smyth, A. W. 2013. Particle filter scheme with mutation for the estimation of time-invariant parameters in structural health monitoring applications. *Structural Control and Health Monitoring* 20(7), 1081-1095.
- Diaz, M., Charbonnel, P. É., Chamoïn, L. 2023. A new Kalman filter approach for structural parameter tracking: application to the monitoring of damaging structures tested on shaking-tables. *Mechanical Systems and Signal Processing* 182, 109529.
- Dolgov, S., Anaya-Izquierdo, K., Fox, C., Scheichl, R. 2020. Approximation and sampling of multivariate probability distributions in the tensor train decomposition. *Statistics and Computing* 30, 603-625.
- Erazo, K., Sen, D., Nagarajaiah, S., Sun, L. 2019. Vibration-based structural health monitoring under changing environmental conditions using Kalman filtering. *Mechanical systems and signal processing* 117, 1-15.
- Ghanem, R., Ferro, G. 2006. Health monitoring for strongly non-linear systems using the Ensemble Kalman filter. *Structural Control and Health Monitoring: The Official Journal of the International Association for Structural Control and Monitoring and of the European Association for the Control of Structures* 13(1), 245-259.
- Kobyzev, I., Prince, S. J., Brubaker, M. A. 2020. Normalizing flows: An introduction and review of current methods. *IEEE transactions on pattern analysis and machine intelligence* 43(11), 3964-3979.
- Lei, Y., Xia, D., Erazo, K., Nagarajaiah, S. 2019. A novel unscented Kalman filter for recursive state-input-system identification of nonlinear systems. *Mechanical Systems and Signal Processing* 127, 120-135.
- Lye, A., Marino, L., Cicirello, A., Patelli, E. 2023. Sequential ensemble Monte Carlo sampler for on-line Bayesian inference of time-varying parameter In engineering applications. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering* 9(3), 031202.
- Marzouk, Y., Moselhy, T., Parno, M., Spantini, A. 2016. Sampling via measure transport: An introduction. In: Ghanem, R., Higdon, D., Owhadi, H. (Eds.), *Handbook of Uncertainty Quantification*, Springer International Publishing, Cham, 1-41.
- MParT Development Team. 2023. Monotone Parameterization Toolkit (MParT), Version 2.2.0. <https://measuretransport.github.io/MParT/>
- Orchard, M. E., Vachtsevanos, G. J., 2009. A particle-filtering approach for on-line fault diagnosis and failure prognosis. *Transactions of the Institute of Measurement and Control* 31(3-4), 221-246.
- Spantini, A., Bigoni, D., Marzouk, Y. 2018. Inference via low-dimensional couplings. *The Journal of Machine Learning Research* 19(1), 2639-2709.
- Spantini, A., Baptista, R., Marzouk, Y. 2022. Coupling techniques for nonlinear ensemble filtering. *SIAM Review* 64(4), 921-953.
- Vanik, M. W., Beck, J. L., Au, S. K. 2000. Bayesian probabilistic approach to structural health monitoring. *Journal of Engineering Mechanics* 126(7), 738-745.

Xie, Z., Feng, J. 2012. Real-time nonlinear structural system identification via iterated unscented Kalman filter. *Mechanical Systems and Signal Processing* 28, 309-322.