Transport map Bayesian parameter estimation for dynamical systems

Jan Grashorn*[1], Jorge-Humberto Urrea-Quintero[2], Matteo Broggi[1], Ludovic Chamoin[3], and Michael Beer[1,4,5]

[1]Institute for Risk and Reliability, Leibniz University Hannover, Hannover, Germany
[2]Institute for Mechanics and Computational Mechanics, Leibniz University Hannover, Hannover, Germany
[3]Université Paris-Saclay, CentraleSupelec, ENS Paris-Saclay, CNRS, LMPS - Laboratoire de Mécanique Paris-Saclay, Gif-sur-Yvette, France
[4]Institute for Risk and Uncertainty, University of Liverpool, Liverpool, United Kingdom
[5]International Joint Research Center for Resilient Infrastructure & International Joint Research Center for Engineering Reliability and Stochastic Mechanics, Tongji University, Shanghai, China

15.01.2023

Abstract

Accurate online state and parameter estimation of uncertain non-linear dynamical systems is a demanding task that has been traditionally handled by adopting non-linear Kalman Filters or particle filters. However, in case of Kalman filters the system needs to be linearised and for particle filters the computational demand can be high. Recent advances in optimal transport theory and the application to Bayesian model updating pave the way for other approaches to system and parameter identification. They also provide a way of formulating the problem in such a way that efficient online estimation for complex systems is possible. In this work, we investigate the properties of the transport map approach when compared to standard Markov Chain Monte Carlo in an off-line setting as a first step towards on-line parameter estimation. We apply both approaches to an analytical exponential model and a dynamical system with seven unknown parameters subjected to ground displacement. Details on the theory of transport maps and on the used MCMC algorithm are also given.

1 Introduction

System identification and/or model parameters estimation is a daily task in any engineering discipline. In dynamical system applications, Kalman or particle filters proved to give great results and are therefore widely used. In certain situations however, these methods fail to accurately estimate the unknown parameters because of missing data, the system's complexity or unknown model properties. In the case of particle filters, the computational demand can sometimes be too high to apply them in an on-line fashion. For off-line estimation, there also exist a multitude of approaches, we focus here on the Bayesian formulation for which a recent interest has been ignited due to the development of very efficient sampling algorithms and the availability of higher computational resources. For the exploration of the posterior distribution usually algorithms based on Markov Chain Monte Carlo (MCMC) are used, since these approaches do not require knowledge about the posterior's topology. However, a downside of MCMC is that the convergence can not easily be assessed and sometimes many samples are needed in order to fully reach an adequate result. Some MCMC algorithms also suffer from burn-in.

Recently there have been advances in optimal transport theory [1] which were applied in the Bayesian updating context [2, 3]. This opens up the possibility of circumventing some of the issues of MCMC methods, since transport maps provide a means of formulating an analytical relationship between some chosen, easy to evaluate reference distribution and the posterior distribution. Integrals can thus be evaluated on the reference distribution and then be transported to the posterior. In addition, sampling from the posterior becomes a simple evaluation of the map. The problem of finding this map is solved by optimization. Previous

*Corresponding author: grashorn@irz.uni-hannover.de

works have implemented transport map (TM) approximation in various use-cases, using synergies of this approach with model order reduction techniques to speed up the process [4, 5].

2 Parameter Estimation

2.1 Bayesian Model Updating

Let $\theta \in \mathbb{R}^d$ be a d-dimensional random variable with probability $p(\theta)$ describing uncertain parameters of a model $\mathcal{M}(\theta)$. Given measured data $\mathcal{D}$ the probability of observing θ in $\mathcal{M}(\theta)$ under the condition of $\mathcal{D}$ can be calculated using Bayes' theorem [7]

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)\, p(\theta)}{p(\mathcal{D})} \tag{1}$$

where the likelihood $p(\mathcal{D}|\theta)$ describes the probability of observing the data under the assumption of θ and is usually modeled as a stochastic distance between $\mathcal{M}(\theta)$ and $\mathcal{D}$. One key difficulty in Bayesian model updating (BMU) is the evaluation of $p(\mathcal{D}|\theta)$ since its shape is generally irregular and unknown and it can only be evaluated point-wise. Therefore, MCMC methods are employed to explore the probability space [8]. The obtained posterior distribution $p(\theta|\mathcal{D})$ is an expression for the updated probability for θ constrained by the observation of $\mathcal{D}$. $p(\mathcal{D})$ is constant for any given set of model and data so Eq. (1) is also used in the non-normalized form

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)\, p(\theta). \tag{2}$$

This poses no issue for MCMC methods since the posterior's shape is not affected.

The MCMC algorithm used in this paper is the Transitional MCMC (TMCMC) method [6]. The main idea is to introduce an exponent $\alpha_j \in [0, 1]$ to the likelihood

$$p(\theta|\mathcal{D}) \propto p(\mathcal{D}|\theta)^{\alpha_j}\, p(\theta) \tag{3}$$

and increasing α_j with each level j starting from $\alpha_1 = 0$, which is equal to sampling from the prior density. For $\alpha_j = 1$ Eq. (3) becomes Eq. (2). Values for α_j for the intermediate levels are chosen such that the COV of $p(\mathcal{D}|\theta_j)^{\alpha_{j+1}-\alpha_j}$, where θ_j are the samples in the j-th level, equals a user-chosen threshold. This ensures that the intermediate levels converge towards the posterior with some chosen rate. In this paper, the COV was chosen to be equal to 1. After drawing samples from the prior density, the Adaptive Metropolis-Hastings algorithm is used to draw samples for the next level until $\alpha_j = 1$ is reached. The main motivation behind TMCMC is to avoid the problem of sampling from difficult target PDFs but sampling from a series of PDFs that converge to the target PDF and that are easier to sample [6].

2.2 Transport Maps

A transport map M is a deterministic coupling between a reference density ρ and the target density π

$$\int f(y)\pi(y)\mathrm{d}y = \int f(M(x))\rho(x)\mathrm{d}x \quad \text{with } Y = M(X) \tag{4}$$

where the target density in the case of BMU is the posterior distribution. The reference density can be chosen freely by the analyst. Common choices are standard normal or standard uniform distributions [3]. Any integrals on the target density can thus be calculated on the reference density by use of the map M. Moreover, samples from the target density Y can be drawn by drawing samples X from the reference density and then evaluating the map M. This makes it possible to find an analytical formulation for the posterior density in BMU, which is usually difficult or impossible. The task now becomes to find the map M. A map can be any invertible function $M : \mathbb{R}^d \to \mathbb{R}^d$, e.g. polynomials or even neural networks [2]. Using the notation $M_\#$ for the push-forward operation the mismatch of the approximation $\pi \approx M_\#\rho$ can be expressed with the Kullback-Leibler (KL) divergence

$$\mathcal{D}_{\text{KL}}(M_\#\rho\,||\,\pi) = \mathcal{D}_{\text{KL}}(\rho\,||\,M_\#^{-1}\pi) \tag{5}$$

$$= \mathbb{E}_\rho\Big[\log \frac{\rho}{M_\#^{-1}\pi}\Big] \tag{6}$$

where the invertibility of the map is used in Eq. (5). With $\mathbf{a}$ as map parameters Eq. (6) becomes

$$\mathcal{D}_{\mathrm{KL}}(M_{\#}\rho||\pi) = \int_X \Big[\log \rho(x) - \log \pi(M(\mathbf{a}, x)) - \log[|\det \nabla M(\mathbf{a}, x)|]\Big]\rho(x)\mathrm{d}x. \tag{7}$$

Due to optimality and uniqueness properties, Maps M were proposed to be monotonic, lower-triangular and constructed from components

$$M^k(\mathbf{a}_c^k, \mathbf{a}_e^k, \theta) = \Phi_c(\theta)\mathbf{a}_c^k + \int_0^{\theta_k} \left(\Phi_e(\theta_1, ..., \theta_{k-1}, \bar{\theta})\mathbf{a}_e^k\right)^2 \mathrm{d}\bar{\theta} \tag{8}$$

so that the resulting map has the structure

$$M(\theta) = \begin{bmatrix} M^1(\theta_1) \\ ... \\ M^d(\theta_1, ..., \theta_d) \end{bmatrix}. \tag{9}$$

A good approximation of the posterior density results in a small KL divergence, so that Eq. (7) can be transformed into a minimization problem. Note that $\rho(x)$ does not depend on the map parameters and instead of the full posterior $\pi = p(\theta|\mathcal{D})$ the non-normalized form

$$\tilde{\pi} = p(\mathcal{D}|\theta)\, p(\theta) \tag{10}$$

can be used. Furthermore, since M consists of analytical functions, the involved integrals can be computed by Gauss quadrature. The final minimization problem to obtain the needed map parameters $\mathbf{a}$ is then

$$\min_{\mathbf{a}} \sum_i \omega_i \Big[-\log\left(\tilde{\pi}(M(\mathbf{a}, \theta_i)) - \log(|\det \nabla M(\mathbf{a}, \theta_i)|)\right)\Big] \tag{11}$$

where ω_i and θ_i are weights and integration points for the quadrature.

For computations, a transport map framework (https://transportmaps.mit.edu) developed in Python was used. There is also a newer version available [1], however the calculations are based on the firstly mentioned framework. In early trials it was found that for optimal results a second-order optimization was needed, therefore all optimization was done using the Newton-CG algorithm and thus gradient and Hessian information of the likelihood are needed. Since in general the likelihood function is dependent on the model, the issue arises that also the derivative of the model with respect to θ is needed. In [4, 5] this was solved by using the TM approach in conjunction with model order reduction methods based on polynomial functions, which naturally allow for the calculation of gradient and Hessian.

3 Examples

3.1 Analytical Exponential Model

As a simple example we show the application of transport maps to an analytical problem of the form

$$c = A(1 - \mathrm{e}^{Bt}) + C + \zeta \tag{12}$$

with $A \sim U(0.4, 1.2)$, $B \sim U(0.01, 0.31)$, $C \sim U(-5, 5)$, $\zeta \sim \mathcal{N}(0, \sigma)$, where ζ is a zero-mean Gaussian noise with standard deviation σ. The model was taken from [2]. The parameters to estimate are thus A, B and C. The data is taken from evaluating Eq. (12) at times $t = \{1, 2, 3, 4, 5\}$ with parameters $A = 1$, $B = 0.21$ and $C = 3$. A result plot with indicated measurements and results taken from samples from transport maps and TMCMC can be seen in Figure 1.

To assess the differences in efficiency between the TMCMC and TM approaches in different scenarios we ran the TM estimation once with high and once with low optimization tolerances. Setting a higher tolerance decreases the accuracy but also reduces the amount of calculations needed. The resulting difference in samples is shown in Figures 2 and 3. The number of evaluations needed are summed up in Table 1. Clearly the selected optimization parameters play a role in the efficiency and the accuracy of the transport map method, since both TMCMC and TM give overlapping results for the low tolerance TM approximation, but for a higher tolerance the TM approach has some deficiencies in the covariance structure. However, the

[1] https://measuretransport.github.io/MParT/

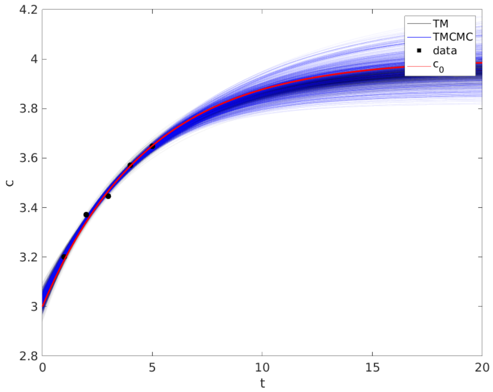

Figure 1: Plot of c over t for the analytical exponential model. The red curve marks the original solution, black dots are noisy measurements, black and blue lines are outputs of samples taken from TM and TMCMC approaches respectively.

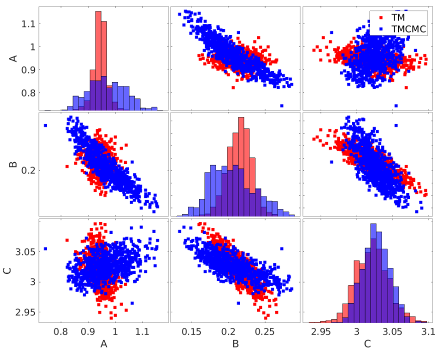

Figure 2: Samples from transport map and TMCMC estimated posterior with high optimization tolerance.

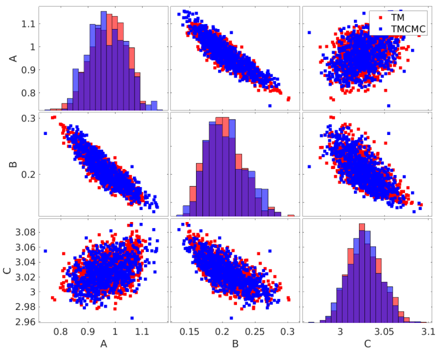

Figure 3: Samples from transport map and TMCMC estimated posterior with low optimization tolerance.

mean values of the parameters are captured very well, even with high tolerance. With regard to the number of model evaluations the high tolerance TM approach has less overall calculations than TMCMC, however if the tolerance is decreased the computational effort increases. Overall the accuracy of the obtained approximation from using transport maps is equal to using TMCMC. Note however that after calculation of the map coefficients, a fully analytical expression is obtained which allows for cheap drawing of new samples if needed. Doing the same with MCMC methods would require further model evaluations.

3.2 Non-linear Dynamical System

Following example shows how to identify a half-car model parameters for a suspension system of 8-th order using Bayesian updating and TM. Conventional passive suspensions use a spring and damper between the car body and wheel assembly. A schematic representation of the system is provided in Figure 4. The mass

Evaluations	Model	Gradient	Hessian
TMCMC	11000	0	0
TM high TOL	2160	4104	1944
TM low TOL	4104	7128	3024

Table 1: Number of model, gradient and hessian evaluations for TMCMC and both cases of TM approximations.

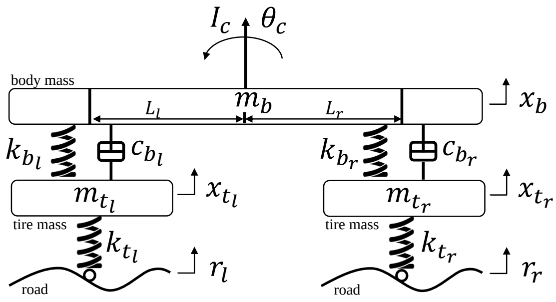

Figure 4: Schematic of the half-car suspension system.

m_b (in kilograms) represents the car chassis (body) and the masses m_{t_l} and m_{t_r} (in kilograms) represents the left and right wheel assembly, respectively. Spring-damper configurations k_{b_l}, c_{b_l} and k_{b_r}, c_{b_l} represent the left and right passive springs and shock absorbers placed between the car body and the wheel assembly. Springs k_{t_l} and k_{t_r} model the compressibility of the pneumatic tire. The variables x_b, x_{t_l}, x_{t_r}, r_l and r_r (all in meters) are the body travel, wheel left and right travel, and road disturbance, respectively. θ_c refers to the pitch (rotational) angle of the chassis. Notice that the non-linear system response comes from the coupling between the vehicle legs through the car body and becomes evident when the equations of motion are written explicitly down for the masses representing the left and right wheel assembly.

Parameters to update were chosen to be all masses and stiffnesses, so $\theta = [k_{t_r}, k_{t_l}, k_{b_r}, k_{b_l}, m_{t_r}, m_{t_l}, m_b]^T$, adding up to a total of seven parameters. Measurements were taken by simulation of the eight degrees of freedom (displacement and velocity of the three masses as well rotation of the top mass) as over a period of 30 s at a sampling rate of 4 Hz, afterwards Gaussian noise was added to all measurements. To help with regularizing the posterior the parameters were transformed to standard normal space. The measurements were then taken from a simulation with all model parameters set to 1. Priors were assumed to be uniformly distributed in the interval $[0.5\mu_i, 1.5\mu_i]$ where μ_i indicates the mean value of θ_i in physical space. For TMCMC 1000 samples were taken per level, for the TM approach a low tolerance (step size smaller than 10^{-6}) was chosen in the optimization to give an adequate approximation. Figure 5 shows the samples from the TM and TMCMC estimated posterior. Figure 6 shows the model response for x_{t_l} for the obtained samples as an example for the approximation.

Both estimation methods show similar results, with some deviation in θ_3 and θ_7, however the model response for samples obtained from TM and TMCMC are very similar. Moreover, the model response shows that both methods are able to capture the dynamics well. The close-up in figure 6 shows that both methods have a slight bias in different directions since samples from either method accumulate either above or below the true solution (dashed line). Evaluating the log-posterior at the sample mean of both approaches gives a value of -13.3 for TMCMC and -18.6 for TM, indicating that the TMCMC results are more probable. The differences between both methods possibly result from the more difficult optimization problem in the TM approach. Since transport maps approximate the target density by integration, the resulting posterior density captures as much probability mass as possible with the given map layout. A higher map order or different map structures would lead to different results in this example. Finding a suitable expression for the likelihood, e.g. one based on frequency-domain approaches, could also help mitigate some issues since it could overcome the inherent non-uniqueness, however due to the limited scope of this work there was no effort taken in finding an optimal formulation of the likelihood function. Moreover, the KL divergence which is used as minimization target can directly be used to assess the approximation quality without any further calculations. This also can help to analyze if the posterior was approximated sufficiently or if further steps need to be taken. Again, due to the limited scope the approximation was not analyzed quantitatively.

4 Conclusion and Outlook

In this contribution the transport map approach for estimation of the posterior in Bayesian parameter estimation was compared to standard MCMC. Both methods were used on noisy data from an analytical exponential model and a model of a non-linear dynamical system. The TM estimation shows great promise in circumventing some of the problems arising in MCMC sampling, however further research needs to be done for the application to dynamical systems. Special care needs to be taken in the problem setup. In

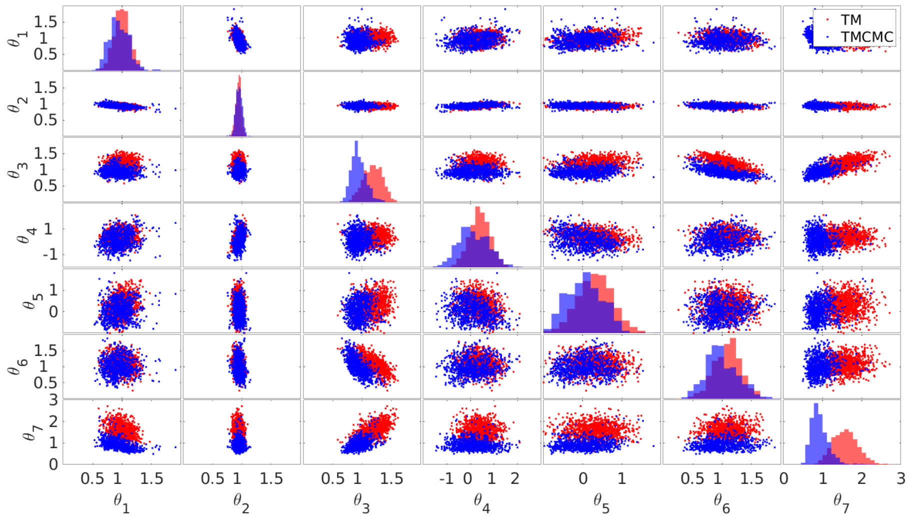

Figure 5: Samples from transport map and TMCMC estimated posterior.

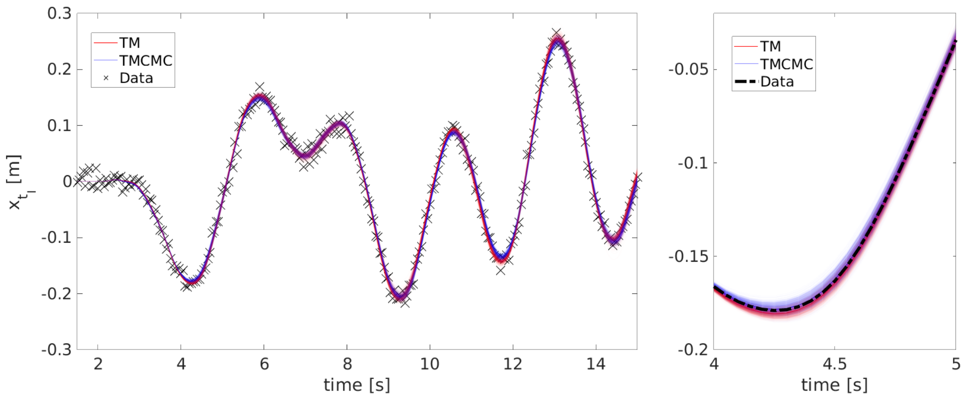

Figure 6: Left: Model output and data for x_{t_l} from obtained samples with TM and TMCMC for the first 15 s. Right: Close-up of model results and data from 4 to 5 s. The dashed line in this plot is the true model output without noise.

order to fully use the capabilities of the TM framework, first and second order derivatives of the posterior are needed, which in turn requires the model to be differentiable with respect to its parameters. It was found that the TM approach generally requires more care in the setup, since the optimization method and its parameters need to be chosen adequately. Further parameters that have an influence on the TM approximation quality but were not analyzed here are the used integration scheme for the maps themselves and the KL-divergence, as well as the order of the maps. It is also possible to use entirely different map layouts such as neural networks, which can have a large effect on the map accuracy and efficiency. The result of the transport map approximation is a fully analytic expression of the posterior distribution, allowing for integration and resampling in an efficient way, which is one of the main advantages over MCMC-based methods. Moreover, sequential updating, which was not covered here, is naturally possible by combining multiple maps. Issues with optimization convergence that arise due to the complicated shape of the posterior distribution could be reduced this way, since the change in the approximated posteriors is smaller in the sequential setting when compared to using all data points at once. The usage of transport maps for sequential updating is also interesting for on-line parameter estimation when combined with model-order reduction methods.

Acknowledgement

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

References

[1] C. Villani. Optimal transport: old and new. Berlin: Springer, 2009.

[2] Parno, Matthew D., and Youssef M. Marzouk. "Transport map accelerated markov chain monte carlo." SIAM/ASA Journal on Uncertainty Quantification 6.2 (2018): 645-682.

[3] Spantini, Alessio, Daniele Bigoni, and Youssef Marzouk. "Inference via low-dimensional couplings." The Journal of Machine Learning Research 19.1 (2018): 2639-2709.

[4] Rubio, Paul-Baptiste, François Louf, and Ludovic Chamoin. "Transport Map sampling with PGD model reduction for fast dynamical Bayesian data assimilation." International Journal for Numerical Methods in Engineering 120.4 (2019): 447-472.

[5] Rubio, Paul-Baptiste, Ludovic Chamoin, and François Louf. "Real-time Bayesian data assimilation with data selection, correction of model bias, and on-the-fly uncertainty propagation." Comptes Rendus Mécanique 347.11 (2019): 762-779.

[6] Ching, Jianye, and Yi-Chu Chen. "Transitional Markov chain Monte Carlo method for Bayesian model updating, model class selection, and model averaging." Journal of engineering mechanics 7 (2007): 816-832.

[7] Beck, James L., and Lambros S. Katafygiotis. "Updating models and their uncertainties. I: Bayesian statistical framework." Journal of Engineering Mechanics-Proceedings of the ASCE 124.4 (1998): 455-462.

[8] Beck, James L., and Siu-Kui Au. "Bayesian updating of structural models and reliability using Markov chain Monte Carlo simulation." Journal of engineering mechanics 128.4 (2002): 380-391.