Fast Bayesian Inference for an Inverse Heat Transfer Problem using Approximations

M. Neumayer and D. Watzenig
Institute of Electrical Measurement and Measurement Signal Processing,
Kopernikusgasse 24/IV, 8010 Graz,
Graz University of Technology, Austria,
Email: neumayer@tugraz.at

H. R. B. Orlande and M. J. Colaco
Department of Mechanical Engineering
POLI/COPPE
Federal University of Rio de Janeiro, UFRJ
Rio de Janeiro, RJ, Brazil

G. S. Dulikravich
Department of Mechanical and Materials Engineering
Florida International University
Miami, Florida, U.S.A.

***Abstract*—The Bayesian inversion of measured data forms an attractive approach to gain statistical knowledge like confidential intervals about the unknown variables given measured data and a model. Out of the class of Markov chain Monte Carlo (MCMC) methods, the Metropolis Hastings (MH) algorithm is a commonly used algorithm to generate samples from the posterior distribution for computational inference. Though easy to implement, the MH algorithm offers drawbacks in terms of computation time and greater modeling costs. In this paper we present an acceleration approach to speed up MCMC with the MH algorithm for an inverse heat transfer problem using two different types of approximations. We will demonstrate the possibility to decrease computation times while maintaining the same estimation accuracy.**

I. Introduction

Inverse and parameter estimation problems form a class of estimation problems, where one soughs to determine a vector $\boldsymbol{x} \in \mathbb{R}^N$ from measurements $\tilde{\boldsymbol{d}} \in \mathbb{R}^M$, given knowledge about the physical process $P : \boldsymbol{x} \mapsto \tilde{\boldsymbol{d}}$ [1]. For the case of complex interactions between $\boldsymbol{x}$ and $\tilde{\boldsymbol{d}}$ the determination of $\boldsymbol{x}$ given $\tilde{\boldsymbol{d}}$ can become a considerable hard or even an intractable indirect measurement problem. This especially holds if P is determined by an underlying partial differential equation (PDE). Often ill-posed in their nature, these problems form a hard class of estimation problems as a stable solution can only be found approximately. The model based treatment of these problems requires a (computer) model $F : \boldsymbol{x} \mapsto \boldsymbol{y}, \boldsymbol{y} \in \mathbb{R}^M$, of the process P which typically has to be executed several times during the solution process. Classical deterministic inversion approaches often rely on the minimization of some meaningful objective function. The result provided by these methods is a point estimate and can usually be obtained after a few number of evaluations of F due to powerful optimization methods.

In opposition to deterministic methods are Bayesian methods. Bayesian inference methods utilize all available information to form the posterior density function $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ [2]. By this, any statistic about $\boldsymbol{x}$ can be computed in order to quantify the quality of the result or to perform uncertainty quantification (UQ) about $\boldsymbol{x}$. This immense benefit comes at the coast of highly increased computation times, especially when using Markov chain Monte Carlo (MCMC) methods.

Recently the application of approximation or surrogate models F^* has seen increased grow in engineering applications like optimization [3]. The idea is to use an approximation F^* of the accurate model F to predict its behavior for changes like the updates during the optimization process. Hereby numerous costly evaluations of the accurate model F can be avoided, as in the ideal case F has only to be evaluated for a verification of F^*. Thus, the computational load of the design process of computational complex problems can be reduced. An approach for using approximation techniques for the solution of inverse problems has been presented in [4], where the behavior of the posterior distribution is approximated. In this paper we will present the application of surrogate techniques to speed up Bayesian inversion by approximating the forward map F for an inverse heat transfer problem, which is the indirect measurement problem under investigation in this work. Several different acceleration approaches are presented together with techniques to incorporate knowledge about the approximation error or even to adapt the approximation.

This paper is structured as follows. In section II we give an introduction to the Bayesian inference approach to solve inverse problems and present the used algorithm. Section III describes the inverse heat transfer problem and the approximation approach to accelerate the MCMC algorithm. Finally section IV contains a simulation study about the proposed approach.

II. Inference and MCMC Methods for Inverse Problems

Bayesian inference approaches are marked by the use of Bayes law $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}}) \propto \pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})\pi(\boldsymbol{x})$, where $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})$ is the likelihood function and $\pi(\boldsymbol{x})$ is the prior [2]. Hereby $\pi(\cdot)$ denotes probability density functions (pdfs). By the prior distribution $\pi(\boldsymbol{x})$ expert knowledge about the expected solution can be incorporated. The likelihood function combines knowledge about the model and the measurement noise. Assuming that $\tilde{\boldsymbol{d}}$ is the sum of noise free measurements $\boldsymbol{d}$ and white Gaussian (measurement) noise $\boldsymbol{v}$, $\boldsymbol{v} \in \mathbb{R}^M$. In this case $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})$ is of

form

$$\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x}) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{y}-\tilde{\boldsymbol{d}}\right)^T \boldsymbol{\Sigma}^{-1}\left(\boldsymbol{y}-\tilde{\boldsymbol{d}}\right)\right), \quad (1)$$

where $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_v$ is the covariance matrix of the measurement noise [1]. As can be seen, the evaluation of $\pi(\tilde{\boldsymbol{d}}|\boldsymbol{x})$ requires one evaluation of the forward map F. Algorithms for practical computational Bayesian inference are typically sampling algorithms [5]. Sampling algorithms can be seen as random number generators which ideally compute independent samples from the target distribution. For inverse problems the target distribution is the posterior. This fact results in generally higher computational cost when using sampling algorithms for computational inference.

A class of algorithms for computational inference are so called Markov chain Monte Carlo (MCMC) methods, as they use an underlying Markov chain $\boldsymbol{X}$ for the generation of samples from the posterior. The Metropolis Hastings (MH) algorithm [6] is one prominent example of this class of methods. Due to its working principle the MH algorithm requires one evaluation of F in every iteration. With respect to its statistical efficiency, the MH algorithm is considerable expensive in terms of the computational costs versus the generation of independent samples. In this work we will use a variant of the MH algorithm termed the delayed acceptance MH algorithm (DAMH) [7]. The algorithm takes its computational advantage out of the use of an approximated forward map $F^*(\boldsymbol{x})$ and works as follows:

1) Pick the actual state $\boldsymbol{x} = \boldsymbol{X}_n$ from the Markov chain.
2) With proposal density $q(\boldsymbol{x}, \boldsymbol{x}')$ generate a new state $\boldsymbol{x}'$.
3) Compute the acceptance ratio $\alpha = \min\left[1, \frac{\pi^*(\boldsymbol{x}'|\tilde{\boldsymbol{d}})q(\boldsymbol{x}',\boldsymbol{x})}{\pi^*(\boldsymbol{x}|\tilde{\boldsymbol{d}})q(\boldsymbol{x},\boldsymbol{x}')}\right]$.
4) With probability α accept $\boldsymbol{x}'$ to be a proposal for the standard MH algorithm. Otherwise set $\boldsymbol{x}' = \boldsymbol{x}$ and return to 2.
5) Compute the acceptance ratio $\beta = \min\left[1, \frac{\pi(\boldsymbol{x}'|\tilde{\boldsymbol{d}})q(\boldsymbol{x}',\boldsymbol{x})}{\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})q(\boldsymbol{x},\boldsymbol{x}')}\right]$.
6) With probability β accept $\boldsymbol{x}'$ and $\boldsymbol{X}_{n+1} = \boldsymbol{x}'$, otherwise reject $\boldsymbol{x}'$ and set $\boldsymbol{X}_{n+1} = \boldsymbol{x}$.

The DAMH algorithm can be seen as two nested MH algorithms (in the original MH algorithm line 3 and 4 do not exist). Starting from the actual state $\boldsymbol{x}$ of the Markov chain $\boldsymbol{X}$, a proposal candidate $\boldsymbol{x}'$ is generated in line 2. Then the DAMH evaluates the acceptance ration α of the proposal candidate $\boldsymbol{x}'$ using the approximation F^* (line 3). If the proposal is accepted (line 4) the DAMH runs the inner MH algorithm where the acceptance ratio β is evaluated using the exact forward map F (line 5). Hereby the DAMH gains its advantage from the computational cheap pre-evaluation of the proposal candidates $\boldsymbol{x}'$ in line 3. For the MH algorithm a high rejection rate (a large number of candidates $\boldsymbol{x}^*$ are rejected) results in a huge amount of unprofitable computation time. For the DAMH an evaluation of F only occurs, if the proposal is accepted using F^*. In this sense the outer MH algorithm of the DAMH can be seen as a filter for bad proposals or in combination with the proposal generation as an improved proposal generator for the inner MH. It becomes obvious that the performance gain of the DAMH depends on the quality of the approximation F^*, as a deterministic approximation error $\boldsymbol{e} = F(\boldsymbol{x}) - F^*(\boldsymbol{x})$ is present. Thus, $\pi^*(\boldsymbol{x}|\tilde{\boldsymbol{d}})$ has to be close to $\pi(\boldsymbol{x}|\tilde{\boldsymbol{d}})$, meaning that $\boldsymbol{e}$ has to be small.

A remarkable point in the Bayesian framework is given by the fact, that information about the deterministic approximation error can be incorporated as well. Within the DAMH algorithm the deterministic approximation error $\boldsymbol{e}_n = F(\boldsymbol{x}) - F^*(\boldsymbol{x})$ is available after the evaluation of line 5. This can be used to build a Gaussian model for the distribution of $\boldsymbol{e}_n$ by [8]

$$\boldsymbol{\mu}_{\boldsymbol{e},n} = \frac{1}{n}\left((n-1)\boldsymbol{\mu}_{\boldsymbol{e},n-1} + \boldsymbol{e}_n\right), \quad (2)$$

$$\boldsymbol{C}_{\boldsymbol{e},n} = \boldsymbol{C}_{\boldsymbol{e},n-1} + \boldsymbol{e}_n\boldsymbol{e}_n^T, \quad (3)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{e},n} = \frac{1}{n-1}\left((n-1)\boldsymbol{C}_{\boldsymbol{e},n} - n\boldsymbol{\mu}_{\boldsymbol{e},n}\boldsymbol{\mu}_{\boldsymbol{e},n}^T\right). \quad (4)$$

Then the likelihood (1) of $\pi^*(\boldsymbol{x}'|\tilde{\boldsymbol{d}})$ is modified by $\boldsymbol{\mu}_{\boldsymbol{e}}$ and the covariance matrix $\boldsymbol{\Sigma}$ becomes the sum of $\boldsymbol{\Sigma}_v$ and $\boldsymbol{\Sigma}_e$. The approach to incorporate knowledge about the approximation error by means of a Gaussian distribution $\boldsymbol{e} \propto \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{e}}, \boldsymbol{\Sigma}_{\boldsymbol{e}})$ is often referred to as enhanced error model (EEM) [1]. In its original form $\boldsymbol{\mu}_{\boldsymbol{e}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{e}}$ are obtained by samples over the prior $\pi(\boldsymbol{x})$. In this sense, the distribution $\pi(\boldsymbol{e})$ has a global character. In contrast to the fundamental idea of the EEM, the approach given by (2) to (4) builds and adaptive version for $\pi(\boldsymbol{e}_n)$ of form $\boldsymbol{e}_n = \mathcal{N}(\boldsymbol{\mu}_{\boldsymbol{e}_n}, \boldsymbol{\Sigma}_{\boldsymbol{e}_n})$ over the posterior distribution. Thus, $\pi(\boldsymbol{e}_n)$ is adopted with respect to the local behavior of the approximation error.

III. The Inverse Heat Transfer Problem

As a practical application of the proposed approach, we consider a 1D inverse heat transfer problem in a domain $\Omega : 0 \leq x \leq L$. The physical problem consists of a slab initially at the uniform temperature ϑ_0. The slab is heated by a uniform heat flux J at the boundary $x = 0$, while at the boundary $x = L$ it exchanges heat by convection with the surrounding media at the temperature ϑ_0, with a heat transfer coefficient α in $\mathrm{Wm^{-2}K^{-1}}$. There are no heat sources within the medium and the thermophysical properties are supposed constant. The mathematical formulation for this heat conduction problem is given by:

$$\frac{1}{k}\frac{d\vartheta}{dt} = \frac{\partial^2\vartheta}{\partial x^2} \qquad \text{in } 0 < x < L, \text{ for } t > 0 \quad (5)$$

$$-\lambda\frac{\partial\vartheta}{\partial x} = J \qquad \text{at } x = 0, \text{ for } t > 0 \quad (6)$$

$$\lambda\frac{\partial\vartheta}{\partial x} + \alpha\vartheta = \alpha\vartheta_0 \qquad \text{at } x = L, \text{ for } t > 0 \quad (7)$$

$$\vartheta = \vartheta_0 \qquad \text{for } t = 0, \text{ in } 0 < x < L \quad (8)$$

where k is the thermal diffusivity in ms^{-2} and λ is the thermal conductivity in $\mathrm{Wm}^{-1}\mathrm{K}^{-1}$. $\vartheta = \vartheta(x,t)$ is the temperature in K. k is given by $k = \frac{\lambda}{c\rho}$ where c is the specific heat capacity in $\mathrm{Jkg}^{-1}\mathrm{K}^{-1}$ and ρ is the density in kgm^{-3}. Equation (6) presents a Neumann type BC with the applied heat flux modeled by $J = k_1 I^2$, where I is the known heater current in A and k_1 (in $\Omega\mathrm{m}^{-2}$) represents an unknown factor describing the power input. For the numerical solution we use the Crank-Nicolson method in combination with a finite difference scheme. Note that the Courant-Friedrichs-Lewy (CFL) number $\frac{k\Delta t}{\Delta^2 x}$ (Δt and Δx present the time and space discretization, respectively) should be small for stability reasons.

α describes the heat exchange due to ventilation at the right side of the plate. A low value of α (i.e. $\alpha = 1$) represents steady conditions on the right side whereas a high value of α (i.e. $\alpha = 10$) represents a forced ventilation. The aim of the inverse problem is now to estimate λ, α, and k_1 from the trend $\vartheta(L,t)$ given ϑ_0 and I. Although this example seems generic due to the reductions the presented case appears in numerous practical applications like electrical machines, i.e. from the estimated parameters a thermal equivalent circuit model can be derived, which presents the thermal heat exchange.

A. Approximation of the Forward Map

In this subsection we use three approximation approaches to replace the accurate forward map F by a fast approximation or surrogate F^*. These approximations ideally have a considerable lower runtime with respect to the accurate model F but they introduce a deterministic model error $\boldsymbol{e} = F - F^*$. The approaches used in this work are:

- Reduced model F_1^*: use of the forward map with a lower degree of discretization (time and space).
- Surrogate approximation F_2^*: use of an artificial approximation.
- Surrogate approximation F_3^*: use of an adaptive artificial approximation.

F_1^* presents a trivial approach to use the forward map but with a lower degree of discretization. Hence, the approximation error is purely originated by the discretization error. One advantage of this approximation is given by the fact that it provides the same steady state solution as the exact model. In addition for a larger Δx the CFL number becomes automatically smaller, which ensures the numerical stability.

F_2^* and F_3^* are artificial approaches using surrogate models to replace F. Surrogate models rely on methods out of approximation theory. The aim is to approximate the behavior of some function or model F by methods like look up tables (LUT), neuronal networks, Gaussian regression models, Kriging techniques, radial basis function (RBF's), etc. [3]. Because of the artificial nature of $F_{2,3}^*$, surrogate models are generally local models whereas reduced models work over the full parameter space. This means, that the quality of surrogates typically decreases if they are applied over the whole input space, whereas they can be made arbitrarily accurate for a small range. However, because of the nature of inverse problems a surrogate has to work for the whole input parameter space, which makes its quality strongly dependent on the nonlinear behavior of F.

In this work we use a polynomial approximation of form $\boldsymbol{y}_i^* = \boldsymbol{p}_i^T \boldsymbol{x}_{a,i}$ for F_2^*, where $\boldsymbol{x}_{\mathrm{a},i} = \begin{bmatrix} 1 & \boldsymbol{x}^T & f(t_i) \end{bmatrix}^T$ is the augmented input vector. For the function f we recommend the use of some kind of bounded function to take care to saturation effects for larger values of t. In this work we used a sigmoid function. The entries of $\boldsymbol{p}_i$ can be obtained by a least squares fit from sampled data over the prior $\pi(\boldsymbol{x})$. For this samples $\boldsymbol{x}_i$ are generated and y_i is computed using F. Then the overdetermined equation system $\boldsymbol{\mathcal{X}}_a \boldsymbol{p}_i = \boldsymbol{y}_i$ can be formed, where the rows of the matrix $\boldsymbol{\mathcal{X}}_a$ are given by $\boldsymbol{x}_{\mathrm{a},i}^T$ for the samples and $\boldsymbol{y}_i$ contains the exact solutions. By this, $\boldsymbol{p}_i$ can be found using the pseudo inverse. With the same procedure also an initial EEM can be obtained for the DAMH. As the surrogate F_2^* is designed over the prior it has to cover a large range due to the lack of information. Thus, the design of F_2^* has to be done with care to keep the approximation error small over the full range.

The approximation F_3^* is an adaptive approximation, which takes use of the knowledge of the approximation error $\boldsymbol{e}$ to adaptively correct the approximation. The update strategy for F_3^* strongly depends on the general form of F_3^* and on the way how its tuning parameters can be manipulated to incorporate knowledge about the actual error. A considerable simple update strategy can be applied, if F_3^* is based on the polynomial approximation F_2^*. Because of the use of the augmented input vector $\boldsymbol{x}_{\mathrm{a},i}$, the approximation F_2^* is nonlinear with respect to $\boldsymbol{x}$ but it is linear with respect to the coefficients of the vectors $\boldsymbol{p}_i$. An efficient update strategy for $\boldsymbol{p}_i$ is given by the LMS algorithm [9], where the temporal update is given by $\boldsymbol{p}_{i,n+1} = \boldsymbol{p}_{i,n} + \mu e_{i,n} \boldsymbol{x}_{\mathrm{a},i,n}$. The parameter μ is a step size parameter of the LMS. The choice of μ itself is critical, as a too large value of μ can lead to numerical instability of the update procedure, whereas a too small value of μ results in a slow adaptation process of F_3^*. However, due to the usually large numbers of MCMC iterations a slow adaptation is less critical as even a slow adaptation provides an improvement. It is notable to mention, that the error $\boldsymbol{e} = F - F_3^*$ has to be reevaluated for the corrected model $F_{3,n+1}^*$ to update the enhanced error model. Given the update strategy, the surrogate F_3^* can adopt to local specifics about F over the posterior. Thus, the surrogate F_3^* is able to locally improve its quality.

IV. Experiments and Results

A. Simulation Setup

In this section we will present the results of numerical experiments. We assumed an arrangement as described in III. For the true parameters we used: $L = 0.05\,\mathrm{cm}$, $\rho = 8700\,\mathrm{kgm}^{-3}$, $c = 502\,\mathrm{Jkg}^{-1}\mathrm{K}^{-1}$, $\lambda = 430\,\mathrm{Wm}^{-1}\mathrm{K}^{-1}$, $\alpha = 3.5\,\mathrm{Wm}^{-2}\mathrm{K}^{-1}$ and $k_1 = 0.005\,\Omega\mathrm{m}^{-2}$. The ambient

TABLE I
SUMMARY OF THE RESULTS.

Nr.	Experiment	σ_v K	μ_λ $\mathrm{Wm^{-1}K^{-1}}$	σ_λ $\mathrm{Wm^{-1}K^{-1}}$	μ_α $\mathrm{Wm^{-2}K^{-1}}$	σ_α $\mathrm{Wm^{-2}K^{-1}}$	μ_{k_1} $\Omega\mathrm{m}^{-2}$	σ_{k_1} $\Omega\mathrm{m}^{-2}$	T_{sim} s
	true		430.0		3.5		5×10^{-3}		
1	F	0.1	431.9	2.54	2.3	14×10^{-3}	3.3×10^{-3}	21.6×10^{-6}	1255
2	F_1^*	0.1	436.3	2.20	1.8	6×10^{-3}	2.7×10^{-3}	9.1×10^{-6}	353
3	F_2^*	0.1	433.5	2.52	2.1	10.7×10^{-3}	3.0×10^{-3}	13.4×10^{-6}	342
4	F_3^*	0.1	429.9	2.52	2.5	6.8×10^{-3}	3.5×10^{-3}	6.8×10^{-6}	288
5	F	0.5	427.2	12.1	2.3	0.18	3.3×10^{-3}	0.26×10^{-3}	1252
6	F_1^*	0.5	434.2	10.4	2.4	0.12	3.5×10^{-3}	0.18×10^{-3}	537
7	F_2^*	0.5	415.7	11.0	2.0	0.04	2.9×10^{-3}	52.5×10^{-6}	558
8	F_3^*	0.5	431.2	12.0	1.5	0.08	2.1×10^{-3}	0.1×10^{-3}	538

temperature ϑ_0 was set to $\vartheta_0 = 20\,^\circ\mathrm{C}$, the current I was set to $I = 100\,\mathrm{A}$. For the accurate forward map F we used $N = 30$ equidistant intervals, the time step width was set to $\Delta t = 0.05\,\mathrm{s}$. For F_1^* we used $N = 3$ and $\Delta t = 0.2\,\mathrm{s}$. We assumed that $\vartheta(L,t)$ is measured every second for $50\,\mathrm{s}$. The standard deviation of the additive measurement noise was set to $0.1\,^\circ\mathrm{C}$ and $0.5\,^\circ\mathrm{C}$. For the prior knowledge about $\boldsymbol{x} = \begin{bmatrix} \lambda & \alpha & k_1 \end{bmatrix}^T$ we assumed uniform distributions with the boundaries $380\,\mathrm{Wm^{-1}K^{-1}} \leq \lambda \leq 580\,\mathrm{Wm^{-1}K^{-1}}$, $1\,\mathrm{Wm^{-2}K^{-1}} \leq \alpha \leq 10\,\mathrm{Wm^{-2}K^{-1}}$ and $0.001\,\Omega\mathrm{m}^{-2} \leq k_1 \leq 0.01\,\Omega\mathrm{m}^{-2}$. To set up the model F_2^* we draw random samples over the prior distribution of $\boldsymbol{x}$ and performed the least squares fit of all $\boldsymbol{p}_i$ using the pseudo inverse. For the initial surrogate F_3^* we used F_2^*. The LMS step with was set to $\mu = 10^{-10}$. The proposal generation is done by means of a Gaussian distribution. One randomly selected component of the state vector $\boldsymbol{x}$ is manipulated by an additive Gaussian distributed random variable with a standard deviation being $4\,\%$ of the range given by the prior.

Figure 1 depicts the trend of $\vartheta(L,t)$ as well as the trends of F_1^* and F_2^* for the described simulation setup. One can see, that F_1^* provides a too low temperature, whereas the approximated trend by F_2^* is too high. These approximation errors will be compensated by the EEM within the DAMH. The evaluation time of the exact mode F is $24.3\,\mathrm{ms}$. In contrast the evaluation of F_1^* and F_2^* last $4.9\,\mathrm{ms}$ and $49.8\,\mu\mathrm{s}$, respectively. These times were evaluated on an Intel Core II Duo CPU with $3\,\mathrm{GHz}$ clock frequency.

B. Simulation Experiments and Results

To test the approaches the following experiments are performed: i) Standard MH, ii) DAMH with F_1^*, iii) DAMH with F_2^*, iv) DAMH with F_3^*. For all three types of setups 50000 MCMC iterations were performed. The first 5000 samples were considered as burn in phase and rejected for the evaluation. Table I summarizes the results. All three algorithms show similar behavior for the two different noise conditions. The results for α and k_1 are yet too small with respect to the true values. This is an effect of the counteractive behavior of α and k_1. Thus, more (spatially distributed) measurements would be required to determine this variables. As can be seen, the algorithms provide almost

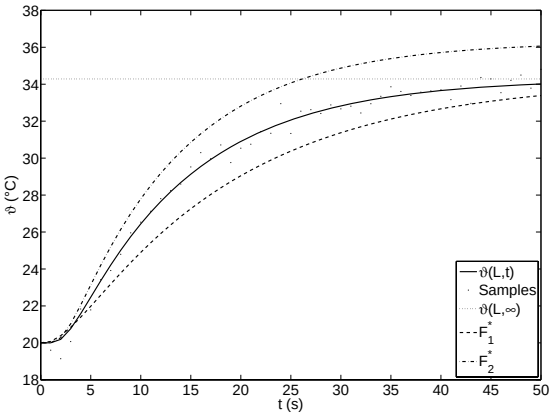


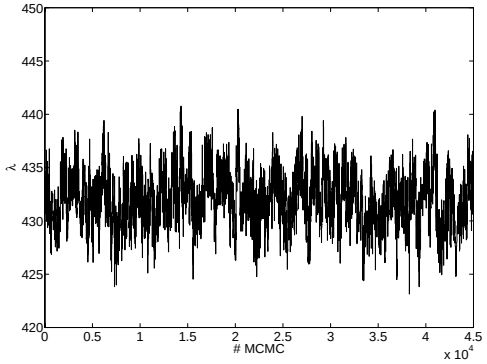
Fig. 1. Trends of $\vartheta(L,t)$ for the described settings and output of the approximations F_1^* and F_2^*.

the same mean estimate for both noise levels.The increased standard deviations for the higher noise level express the increased uncertainty of the result due to the noise. This is the benefit of the Bayesian inversion, as the Bayesian framework provides the possibility to quantify the quality of the result. Figure 2 presents samples for λ found by method i. From these plots one can gain knowledge about the distribution of λ.

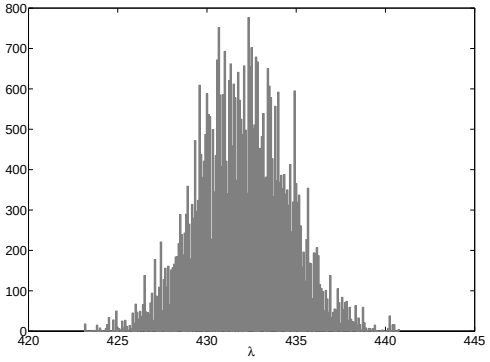
From the computation times in table I one can see the gain due to the use of the approximations. However, this speed up appears quite small with respect to the stated evaluation times of the forward map. Table II provides an analysis of the behavior of the Markov chains for the different algorithms.

The ratio Ac_α states the percentage of accepted proposal candidates. One can see from the MH algorithms (line 1 and 5), that this ratio is considerable low, which results in a huge number of unnecessary evaluations of F. For the DAMH algorithm, Ac_α rates an acceptance in the first step and Ac_β an acceptance in the second step. As can be seen, Ac_α is larger with respect to the MH. This effect comes due to the enhanced error model which adapts the likelihood for the evaluation of $\pi^*(\boldsymbol{x}'|\tilde{\boldsymbol{d}})$. The ratio $Ac_{\beta|\alpha}$ states that a proposal is accepted by both of the nested MH algorithms. Hence, $Ac_{\beta|\alpha}$ is a measure for the quality of the used approximation. As can be seen, the quality of the approximations is considerable low. The low gain in speed can now be explained by the low

quality of the approximations and the increased acceptance ratio Ac_α. These two factors result in a still large amount of unprofitable evaluations of F which diminish the speed up of the DAMH. Thus, the low speed up can be explained by the too large approximation errors of F_1^* and F_2^*. This appears worse but is yet not a big drawback. With respect to the design of the approximations it can be stated, that F_1^* and F_2^* are designed to be at the lower end in terms of quality. Nevertheless all DAMH variants produced the same results as the standard MH and by a more sophisticated setup of the approximations a further speed up can be provided. This can be done by the use of a better discretization for F_1^* or by more coefficients for F_2^*, respectively. The positive effect of the adaptive approximation can be seen in line 4 versus line 3 of table II. The adaption improves the quality of the approximation, as can be seen by the values of $Ac_{\beta|\alpha}$. The lower value of Ac_α results in the fastest solution time as can be seen in line 4 of table I. For the increased variance (line 7 and 8) this effect cannot be observed. The origin of this effect can be explained by the increased variance of the posterior. Thus, F_3^* is trained over a larger range resulting in surrogate with a lower quality. The last column in table II contains the integrated autocorrelation times (IACT) τ_{IACT} of the Markov chains. The IACT's were computed by methods explained in [10] and present the statistical efficiency of the algorithm. Figure 3 presents an evaluation of the MH algorithm. From the exponential trends on can see that the chain behaves well.



(a) MCMC output for λ.

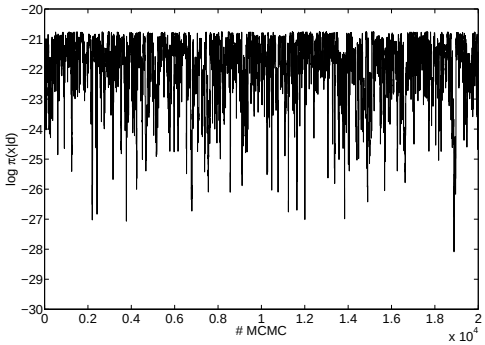


(b) Histogram plot for λ.

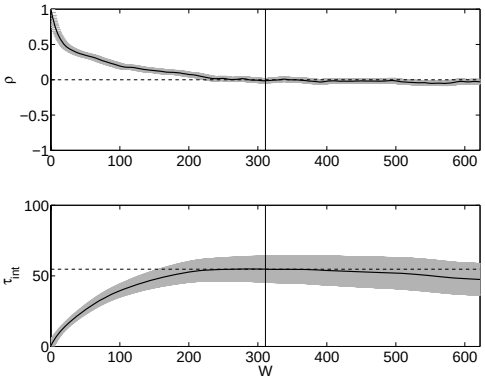
Fig. 2. Output analysis for λ for method i (standard MH MCMC).

TABLE II
BEHAVIOR OF THE CHAINS.

Nr.	Experiment	σ_v K	Ac_α %	$Ac_{\beta\|\alpha}$ %	Ac_β %	τ_{IACT}
1	F	0.1	9.0	X	X	54
2	F_1^*	0.1	10.4	56.4	5.9	190
3	F_2^*	0.1	27.4	27.4	7.5	140
4	F_3^*	0.1	23.2	34.9	8.1	136
5	F	0.5	27	X	X	22.8
6	F_1^*	0.5	28.6	80.8	23.13	50
7	F_2^*	0.5	45.2	50.1	22.6	50
8	F_3^*	0.5	54.4	45.0	24.5	45



(a) MCMC output.



(b) IACT

Fig. 3. Analysis of the behavior of the chain.

V. CONCLUSION

In this work an approach for accelerating Bayesian inference for an inverse heat transfer problem has been presented. The work demonstrates the successful application of approximations to speed up MCMC methods. The analysis of the chains brought the result that a further speed up can be achieved by better approximations. The presented framework can be applied to different problem types to perform Bayesian inference for the solution of indirect measurement problems.

REFERENCES

[1] J. Kaipio and E. Somersalo, *Statistical and Computational Inverse Problems*, ser. Applied Mathematical Sciences. Springer, 2005, vol. 160.

[2] J. M. Bernardo and A. F. M. Smith, *Bayesian Theory*. New York: John Wiley & Sons, 1994 (ISBN: 0-471-92416-4).

[3] A. Forrester, A. Sobester, and A. Keane, *Engineering Design Via Surrogate Modelling: A Practical Guide*. Wiley, 2008.
[4] H. R. B. Orlande, M. J. Colao, and G. S. Dulikravich, "Approximation of the likelihood function in the bayesian technique for the solution of inverse problems," *Inverse Problems in Science & Engineering*, vol. 16, pp. 677–692, 2008.
[5] L. Tierney, "Markov chains for exploring posterior distributions," *Annals of Statistics*, vol. 22, pp. 1701–1762, 1994.
[6] W. Hastings, "Monte Carlo sampling using Markov chains and their applications," *Biometrica*, vol. 57, no. 1, pp. pp. 97–109, 1970.
[7] J. A. Christen and C. Fox, "Markov chain Monte Carlo Using an Approximation," *Journal of Computational and Graphical Statistics*, vol. 14, no. 4, pp. 795–810, 2005. [Online]. Available: http://pubs.amstat.org/doi/abs/10.1198/106186005X76983
[8] T. Cui, "Bayesian Calibration of Geothermal Reservoir Models via Markov Chain Monte Carlo," Ph.D. dissertation, University of Auckland, 2010.
[9] S. Haykin, *Adaptive Filter Theory (4th Edition)*. Prentice Hall, Sept.
[10] U. Wolff, "Monte Carlo errors with less errors," *Computer Physics Communications*, vol. 156, no. 2, pp. 143 – 153, 2004. [Online]. Available: http://www.sciencedirect.com/science/article/B6TJ5-4B3NPMC-3/2/94bd1b60aba9b7a9ea69ac39d7372fc5