Weierstraß-Institut
für Angewandte Analysis und Stochastik
Leibniz-Institut im Forschungsverbund Berlin e. V.

Sampling-free Bayesian inversion with a hierarchical tensor representation

Martin Eigel[1], Manuel Marschall[1], Reinhold Schneider[2]

submitted: December 22, 2016 (revision: February 23, 2017)

[1] Weierstrass Institute Mohrenstr. 39 10117 Berlin Germany E-Mail: martin.eigel@wias-berlin.de manuel.marschall@wias-berlin.de	[2] TU Berlin Straße des 17. Juni 136 10623 Berlin Germany E-Mail: schneidr@math.tu-berlin.de

No. 2363

Berlin 2016

Sampling-free Bayesian inversion with a hierarchical tensor representation

Martin Eigel, Manuel Marschall, Reinhold Schneider

ABSTRACT. The statistical Bayesian approach is a natural setting to resolve the ill-posedness of inverse problems by assigning probability densities to the considered calibration parameters. Based on an affine-parametric deterministic representation of a linear forward model, a sampling-free approach to Bayesian inversion with an explicit representation of the parameter densities with respect to a uniform prior is developed. The approximation of the involved randomness inevitably leads to several high dimensional expressions, which are often tackled with classical sampling methods such as MCMC. To speed up these methods, the use of a surrogate model is beneficial since it allows for faster evaluation with respect to arbitrary calibration parameters. However, the inherently slow convergence can not be remedied by this. As an alternative, a complete functional treatment of the inverse problem is feasible as demonstrated in this work, with functional representations of the parametric forward solution as well as the probability densities of the calibration parameters, determined by Bayesian inversion.

The proposed sampling-free approach is discussed in the context of hierarchical tensor representations, which are employed for the adaptive evaluation of a random PDE (the forward problem) in generalized chaos polynomials and the subsequent high-dimensional quadrature of the log-likelihood. This modern compression technique alleviates the curse of dimensionality by hierarchical subspace approximations of the respective low-rank (solution) manifolds. All required computations can then be carried out efficiently in the low-rank format. An interesting aspect is the evaluation of the exponential of the Bayesian potential by means of an adaptive Runge-Kutta method with tensors. A priori convergence of the posterior is examined, considering all approximations that occur in the method. Numerical experiments demonstrate the performance and confirm the theoretical results.

1. INTRODUCTION

Mathematical models in engineering and science applications are typically characterized by calibration parameters, which are uncertain due to incomplete knowledge. Hence, it is a common task to identify these parameters based on noisy and incomplete measurement data related to the system response. For this, the response of the model is evaluated with different parameter realizations ("exitations"), a functional of which is used to adjust the guess for the parameters. This inverse problem of identification has been tackled with a variety of methods. In a deterministic setting, the problem is ill-posed in the sense of Hadamard and has to be regularized in some way in order to become solvable. As an alternative approach, we are concerned with the Bayesian setting in which the parameters are considered as random variables [35, 7]. The aim then is to determine the posterior density subject to measurements perturbed in accordance with some predetermined noise assumption. With this Bayesian notion of the task, the problem is regularized in the sense that a posterior probability distribution can be obtained.

Computational methods for the efficient evaluation of (the expectation of) the parameter densities have received considerable interest in recent years, in particular as part of the research efforts in the field of Uncertainty Quantification (UQ). The most widely used methods are based on statistical sampling from the posterior measure, namely Monte Carlo (MC) type algorithms such as the popular Markov-Chain Monte Carlo method (MCMC) [17] and variants [9, 8] and [29]. While these methods are well-understood analytically and are relatively simple to implement, a major drawback is the inherently slow convergence, which is limited by the convergence order $1/2$ of MC methods. Since for each sampling of the Markov chain a realization of the governing equation has to be computed, these methods can quickly become very costly computationally.

For the solution of random forward problems, significant progress could be witnessed over the last decade. In particular, methods which aim at the construction of an adequate surrogate model were shown to potentially converge at much higher rates than classical MC methods [32]. These findings are also supported by recent analytical results regarding the sparsity of the solution manifold. In addition to a priori results, adaptive algorithms can be derived which steer the problem-dependent

adjustment of the discretization parameters based on some a posteriori error indicator or even a reliable error estimator [12]. To improve the efficiency of the Bayesian inversion, such surrogate models, e.g. given as a functional representation in generalized chaos polynomials, can be used in combination with sampling methods.

However, also a complete functional representation of the posterior density is feasible, leading to a sampling-free method of the statistical inverse problem. For this, we employ a hierarchical tensor representation of the stochastic forward solution [14]. Hierarchical tensor formats have only recently been investigated more thoroughly in the community of numerical mathematics [18, 19], although these techniques have been used for a long time in physics and chemistry. With hierarchical tensor representations, the low-rank structure of the solution operator and the solution manifold can be fully exploited, which leads to very efficient methods for the evaluation of the system response [23]. We make use of our previous results on adaptive stochastic Galerkin methods in tensor representations.

With a functional representation in the tensor train (TT) format at hand, we derive a representation of the Bayesian potential. Subsequently, the entire Bayesian inversion can be carried out in the low-rank approximation. To describe the posterior density, we introduce an interpolation in parameter space. For the computation of the high-dimensional likelihood in tensor format, we suggest an adaptive Euler scheme.

All employed approximations, i.e. for the forward problem, the tensor exponential and the polynomial interpolation of the probability densities, can be estimated a priori. With this, we provide a convergence analysis of the posterior measure in the Hellinger distance [36].

The structure of this work is as follows: Section 2 reviews the Bayesian setting and introduces the used notation. Moreover, the parametric model problem is defined. Section 3 is concerned with hierarchical tensor formats, which form the basis for the derived method. As a special case, the popular TT format is introduced and functional representations with this format are discussed. Key to this work is the low-rank approximation of the Bayesian potential using an adaptively computed stochastic Galerkin solution of the parametric problem. In order to carry out the Bayesian inversion, the evaluation of the likelihood makes use of an adaptive Euler method for which numerical observations are presented. These preparations culminate in Section 4 where the Bayesian inversion in the hierarchical tensor representation is described. Section 5 is devoted to the derivation of an a posteriori convergence result for the Bayesian posterior, taking into account all occurring approximations. The concluding Section 6 demonstrates the performance of the proposed novel approach.

2. Bayesian Inversion of Operator Equations

The basis for the presented approach is the abstract Bayesian framework derived and analysed in [35, 7]. It is outlined in this section and employed henceforth. We also draw from other works in the same vain such as [32, 30, 22].

2.1. **Bayesian setup.** We consider a class of operator equations depending on some uncertain datum u taking values in a separable Banach space $\tilde{X}$. The datum u, which for instance is a random coefficient field in a PDE, is determined by a countable infinite set of parameters $y = (y_j)_{j\in\mathbb{N}}$. By the observation of the system response δ in the separable Banach space Y, the goal is to gain knowledge about the unknown u by means of Bayesian estimation with respect to some prior measure π_0 on a full measure subset $X \subset \tilde{X}$. The measurement data is perturbed by Gaussian noise $\eta \sim N(0, \Gamma)$ on Y. We assume that Y is finite-dimensional and the covariance operator Γ is non-degenerate, i.e. $Y \subseteq \mathbb{R}^K$ for $K < \infty$ noisy measurements.

Assume a "forward" response operator $G : X \to \mathcal{X}$ mapping from the separable Banach space X of uncertain distributed parameters u into the reflexive Banach space of responses $\mathcal{X}$. The operator $\mathcal{O} = (o_1, \ldots, o_K)^T \in (\mathcal{X}')^K$ models the observation at K sensors including estimated noise from

model and measurement errors such that

$$\delta = (\mathcal{O} \circ G)(u) + \eta : X \to L^2_\Gamma(\mathbb{R}^K). \tag{2.1}$$

The spaces are equipped with the norms $\|\cdot\|_X$ and $\|\cdot\|_{\mathcal{X}}$, respectively.

As model response, we consider solutions of the linear parametric operator equation

$$\text{Given } u \in X, \ f \in \mathcal{Y}' \quad \text{find} \quad q \in \mathcal{X} \quad \text{s.t.} \quad A(u)q = f, \tag{2.2}$$

with a uniformly boundedly invertible random linear operator $A(u) \in \mathcal{L}(\mathcal{X}, \mathcal{X}')$ depending on the random data $u \in X$. With known forcing $f \in \mathcal{X}'$, the response of (2.2) is given by

$$X \ni u \mapsto q(u) := G(u, f) =: G(u) = (A(u))^{-1} f \in \mathcal{X}. \tag{2.3}$$

Equation (2.2) is referred to as the *forward* problem throughout this article. The main goal is to solve the *backward* or *inverse* problem:

$$\text{Given } \delta = (\mathcal{O} \circ G)(u) + \eta \colon X \to L^2_\Gamma(\mathbb{R}^K), \quad \text{find} \quad u \in X. \tag{2.4}$$

Here, $L^2_\Gamma(\mathbb{R}^K)$ denotes the *weighted* space of square integrable functions over $\mathbb{R}^K$ equipped with the norm

$$\|v\|^2_\Gamma := \langle v, v \rangle_\Gamma = \langle v, \Gamma^{-1} v \rangle \tag{2.5}$$

where $\langle \cdot, \cdot \rangle$ is the Euclidean inner product in $\mathbb{R}^K$ and Γ is the symmetric positive definite covariance matrix of the noise η.

In case of an deterministic error η, problem (2.4) is usually ill-posed. However, interpreting the involved objects as random variables over some probability space, the existences of a desired probability distribution can be deduced in the sense of Bayes' theorem, which we recall in the following.

As a measure to quantify the probability of u given δ, we introduce a likelihood model by

$$L \colon \mathcal{B}(\mathbb{R}^K) \times X \to [0, 1]. \tag{2.6}$$

The likelihood explains how well the uncertain data u fits the measurements δ. For any set $E \in \mathcal{B}(\mathbb{R}^K)$ and with the Lebesgue density of η denoted by ϱ, a sensible choice is

$$L(E|u) = \mathbb{P}(\delta \in E|u) = \int_E (\delta - (\mathcal{O} \circ G)(u)) \varrho \, \mathrm{d}\delta. \tag{2.7}$$

The joint random variable (u, δ) is distributed according to the joint (possibly non product) measure μ, which, for $\tilde{E} \in \mathcal{B}(X \times \mathbb{R}^K)$, is given by

$$\mu(\tilde{E}) := \int_X \int_{\mathbb{R}^K} \mathbb{1}_{\tilde{E}}(u, \delta) L(\mathrm{d}\delta|u) \pi_0(\mathrm{d}u). \tag{2.8}$$

Here, π_0 is the *prior* measure on the uncertain data u, containing à priori information about the unknown with $\pi_0(X) = 1$. The suitable choice of the prior is a delicate task and depends on the specific problem.

The sought posterior measure π_δ describes the distribution of u given δ. It results from conditioning the joint measure to the $\mathbb{R}^K$-fibre represented by the measurement vector.

With the *Bayesian potential*[1] defined by

$$\Phi(u; \delta) := -\log \varrho(\delta - (\mathcal{O} \circ G)(u)), \tag{2.9}$$

the theorem of Bayes holds, see [35].

Theorem 2.1 (Bayes' theorem)**.** *Assume that the normalization factor Z is non-negative, i.e.,*

$$Z := \int_X \varrho(\delta - (\mathcal{O} \circ G)(u)) \pi_0(\mathrm{d}u) > 0. \tag{2.10}$$

Then, the posterior measure π_δ of u given δ is absolutely continuous and has a Radon-Nikodym derivative with respect to the prior measure π_0 given by

$$\frac{\mathrm{d}\pi_\delta}{\mathrm{d}\pi_0}(u) = \frac{1}{Z} \varrho\left(\delta - (\mathcal{O} \circ G)(u)\right) = \frac{1}{Z} \exp(-\Phi(u; \delta)). \tag{2.11}$$

---

[1]also called *misfit* or *negative log likelihood*

Since we have chosen η to be a centered Gaussian noise, there exists an explicit expression of the Bayesian posterior in (2.11). Using that the translation of a $\mathcal{N}(0, \Gamma)$ random variable by $(\mathcal{O} \circ G)(u)$ is distributed according to $\mathcal{N}((\mathcal{O} \circ G)(u), \Gamma)$, we obtain

$$\Phi(u; \delta) = \|\delta - (\mathcal{O} \circ G)(u)\|_\Gamma^2 = \frac{1}{2}\langle \delta - (\mathcal{O} \circ G)(u), \Gamma^{-1}(\delta - (\mathcal{O} \circ G)(u))\rangle. \qquad (2.12)$$

Remark 2.2. In general, Bayes' theorem requires the measurability and boundedness of the potential Φ. In our setting, this follows immediately by using the local Lipschitz property of the forward operator and the continuity of the observation operator, see [32].

Remark 2.3. To show the positivity of the normalization constant Z in (2.11), we need the boundedness of $\Phi(u; y)$, which can directly be deduced from the boundedness of $(\mathcal{O} \circ G)(u)$ in $\mathbb{R}^K$ and the η-almost sure finiteness of δ. Hence, the potential is bounded in X by some constant $C(\delta) = C < \infty$ and it follows that

$$Z = Z(y) = \int_X \exp(-\Phi(u; \delta))\pi_0(\,\mathrm{d}u) \geq \int_X \exp(-C)\pi_0(\,\mathrm{d}u) = \exp(-C) > 0. \qquad (2.13)$$

2.2. **Parametric uncertainty model.**

In the following, we introduce a dependency of the uncertain system input u on a set of parameters $y = (y_1, y_2, \ldots) \in \Xi$ in a parameter space Ξ. Hence, the same holds for the solution $G(u(y)) \in \mathcal{X}$ and consequently also the Bayesian formulation in (2.11). For numerical computations to become feasible, several approximations have to be applied.

We make the representation of u more specific and assume a Schauder basis $(\psi_j)_{j=1}^\infty$ of X. Moreover, suppose $y = (y_j)_{j=1}^\infty$ is a set of independent identically distributed random variables with $y \sim \pi_0 = \bigotimes_{j\geq1} \pi_0^j$. With this, we assume an unconditionally L^2-convergent expansion of the form

$$u = u(x, y) := \langle u\rangle(x) + \sum_{j\geq1} \psi_j(x)y_j, \qquad (2.14)$$

where $\langle u\rangle$ is a deterministic nominal value of u. Examples for such decompositions are the Karhunen-Loève expansion, cf. [6, Prop. 2.1.6] and [16, 25, 24], or the principal component analysis [15]. Throughout this article we assume $y \in \Xi = [-1, 1]^\infty$ equipped with the uniform distribution $\pi_0^{(1)} := \pi_0^j = \mathcal{U}(-1, 1)$ on $[-1, 1]$ for all j and suppose $(\psi_j)_{j=1}^\infty$ belongs to $L^\infty(D)$ for some Lipschitz domain $D \subset \mathbb{R}^d$ ($d = 1, 2, 3$). Moreover, we require uniform ellipticity of $u(x, y)$, i.e., for some $\check{u}, \hat{u} > 0$ it holds

$$0 < \check{u} \leq u(x, y) \leq \hat{u} < \infty \qquad \text{for all } x \in D \text{ and } y \in \Xi. \qquad (2.15)$$

As a common model problem, we introduce the stationary diffusion problem on the square domain $D = (0, 1)^2$ with homogeneous Dirichlet boundary conditions, given by

$$\begin{aligned} \operatorname{div}(u(x, y)\nabla q(x, y)) &= f(x) & \text{in } D \times \Xi, \\ q(x, y) &= 0 & \text{on } \partial D \times \Xi. \end{aligned} \qquad (2.16)$$

Due to the uniform ellipticity of the coefficient (2.15), the paramteric operator equation

$$A(y)q = f \qquad (2.17)$$

admits a unique solution $G(y) := G(u(y)) = A^{-1}(y)f$, see e.g. [33] and the references therein for details. In this setting, the Bayesian posterior (2.11) following from Theorem 2.1 and (2.14) is well-defined.

Proposition 2.4 ([30, Prop. 2.3]). *The Bayesian posterior π_δ of $u \in X$ given data $\delta \in \mathbb{R}^K$ is absolutely continuous with respect to the prior π_0 and it holds*

$$\frac{\mathrm{d}\pi_\delta}{\mathrm{d}\pi_0}(y) = \frac{1}{Z} \exp(-\Phi(u(y); \delta))|_{u=\langle u\rangle + \sum_{j\geq1} \psi_j y_j}. \qquad (2.18)$$

One is often interested in functionals of the solution depending on the uncertain data, referred to as quantities of interest (QoI) and denoted by $\varphi \colon X \to \mathbb{R}$. Common QoIs are for instance moments of u, i.e. $\varphi_m(u) = \int_X u^m \,\mathrm{d}\pi_\delta$.

Proposition 2.5. *Given a quantity of interest $\varphi\colon X \to \mathbb{R}$ and noisy data $\delta \in \mathbb{R}^K$, the Bayesian estimate, i.e. the expectation of the QoI φ with respect to the posterior π_δ takes the form*

$$\mathbb{E}_{\pi_\delta}[\varphi] = \frac{1}{Z}\mathbb{E}_{\pi_0}\Big[\exp(-\Phi(u;\delta))\varphi(u)|_{u=\langle u \rangle + \sum_{j\geq 1}\psi_j y_j}\Big]. \tag{2.19}$$

2.3. **Dimension truncation and forward operator approximation.** For practical computations, the infinite expansion (2.14) has to be truncated to a finite number $M < \infty$ of terms. We assume a generalized polynomial chaos (gPC) expansion of u and estimate the error as follows, see [4, Theorem 1] and also [34, 5] for further results on gPC approximations. For more details also see section 3.2.

Theorem 2.6. *Under the uniform ellipticity assumption (2.15) and if $\left(\|\psi_j\|_{L^\infty(D)}\right) \in \ell^p(\mathbb{N})$ for some $0 < p < 1$, there exists a constant $0 < C < \infty$ such that the equation (2.17) admits a unique uniformly bounded solution represented by a gPC expansion*

$$\sup_{y\in\Xi}\|G(y)\|_{\mathcal{X}} \leq C \quad and \quad G(y) = \sum_{\nu\in\mathcal{F}} g_\nu P_\nu(y) \tag{2.20}$$

where $P_\nu(y) := \prod_{i=1}^M P_{\nu_j}(y_j)$, with coefficients $g_\nu \in \mathcal{X}$, P_n denoting the univariate Legendre polynomial of degree n for the interval $[-1,1]$ nomalized such that $\|P_n\|_{L^\infty([-1,1])} = 1$ and $\mathcal{F}$ denotes the countable set of all finitely supported sequences $\nu \in \mathbb{N}_0^M$. Moreover, there exists a downward closed index set $\Lambda_N \subset \mathcal{F}$ with at most N indices such that the dimension independent convergence rate hold

$$\sup_{y\in\Xi}\|G(y) - \sum_{\nu\in\Lambda_N} g_\nu P_\nu(y)\|_{\mathcal{X}} \leq C_{Trun}N^{-\left(\frac{1}{p}-1\right)}. \tag{2.21}$$

Here, the constant C_{Trun} does not depend on N nor on the set of active indices, i.e. $\max\{\#\{j \in \mathbb{N} : \nu_j \neq 0\} : \nu \in \Lambda_N\}$.

The discretization in gPC in the stochastic variables y is denoted by

$$G^N(y) := \sum_{\nu\in\Lambda_N} g_\nu P_\nu(y), \quad y \in \Xi. \tag{2.22}$$

Further, we assume that the forward operator is approximated by a Galerkin finite element solution on some simplicial mesh with maximal element diameter h, which admits a well-known quasi-optimality result.

Proposition 2.7 ([8] Prop 2.3). *Let $(\mathcal{X}^h)_{h>0} \subset \mathcal{X}$, be a sequence of finite dimensional subspaces. Then, given unique solvability of the forward problem corresponding to G, for every $y \in \Xi$ there exists the Galerkin approximation $G^h(y) \in \mathcal{X}^h$ to $G(y) \in \mathcal{X}$. Furthermore, we obtain quasi-optimal convergence, i.e., there exists a constant $C_{Gal} > 0$ independent of h and y such that*

$$\|G(y) - G^h(y)\|_{\mathcal{X}} \leq C_{Gal}\inf_{0\neq v\in\mathcal{X}^h}\|G(y) - v\|_{\mathcal{X}}. \tag{2.23}$$

Moreover, assume a sequence of subspaces $(\mathcal{X}_t)_{t>0} \subset \mathcal{X}$ with $\mathcal{X}_{t_2} \subset \mathcal{X}_{t_1}$ for $t_1 < t_2$ dense in $\mathcal{X}$ and for every $v \in \mathcal{X}_t$, it holds

$$\inf_{v^h\in\mathcal{X}^h}\|v - v^h\| \leq C_{Gal}h^t\|v\|_{\mathcal{X}_t}. \tag{2.24}$$

Then, for $t > 0$,

$$\|G(y) - G^h(y)\|_{\mathcal{X}} \leq C_{Gal}h^t\sup_{y\in\Xi}\|G(y)\|_{\mathcal{X}_t}. \tag{2.25}$$

Remark 2.8. From Theorems 2.6 and 2.7 we can deduce a combined error bound for the forward operator, namely

$$\|G(y) - G^{h,N}(y)\|_{\mathcal{X}} \leq C_{\text{Trun,Gal}}(h^t + N^{-\frac{1}{p}+1}), \quad y \in \Xi. \tag{2.26}$$

3. Hierarchical Tensor Formats

The presented method relies on an adequate functional representation, which allows for the efficient evaluation of high-dimensional integrals as they occur in the Bayesian inversion with parametric PDEs. Recently, hierarchical tensor methods were shown to be well-suited for such high-dimensional problems, in particular since they allow to alleviate the curse of dimensionality in case the considered functions are on low-rank manifolds. This section provides a general introduction to hierarchical tensor formats and in particular scrutinizes the tensor train (TT) format as a convenient specialization employed henceforth. The representation of multivariate functions in a geometric tensor setting was e.g. introduced in [20]. We also refer to [31] for an overview in the context of parametric PDEs. In [14], an adaptive tensor approach was described which is used as the solver of the forward model. Section 3.2 gives a brief overview of the stochastic Galerkin method.

Hierarchical tensor product spaces are defined based on the notion of *dimension partition trees*. Therefore, let $\widetilde{\mathcal{V}}_1, \ldots, \widetilde{\mathcal{V}}_M$ be real Hilbert spaces of functions depending on the variables x_i, $i = 1, \ldots, M$, and consider the topological tensor product

$$\widetilde{\mathcal{V}} := \bigotimes_{m=1}^{M} \widetilde{\mathcal{V}}_m. \tag{3.1}$$

Define $\mathbb{T} \subset \mathcal{F}(\{1, \ldots, M\})$, i.e. a subset of the power set of the index set, as *dimension partition tree* with the following properties:

  1 The root $\alpha^* := \{1, \ldots, M\}$ corresponds to the full index set.
  2 Every *node* $\alpha \in \mathbb{T}$ is either a *leaf*, i.e. $|\alpha| = 1$, or there exists $\alpha_1, \alpha_2 \in \mathbb{T}$ such that $\alpha = \alpha_1 \cup \alpha_2$ and $\alpha_1 \cap \alpha_2 = \emptyset$.

This construction is used for binary trees and gives direct access to the length of the tree and allows for simple traversal algorithms. For a node $\alpha \in \mathbb{T}\backslash\{\alpha^*\}$ with sons α_1 and α_2 assume an associated subspace $\mathcal{U}_\alpha \subset \bigotimes_{j \in \alpha} \widetilde{\mathcal{V}}_j$ of dimension r_α spanned by the basis $\{\times_{\mu \in \alpha}\{x_\mu\} =: x_\alpha \mapsto U^\alpha[x_\alpha, k_\alpha]\}_{k_\alpha=1}^{r_\alpha}$ of $\mathcal{U}_\alpha$ and it holds the nestedness property

$$\mathcal{U}_\alpha \subset \mathcal{U}_{\alpha_1} \otimes \mathcal{U}_{\alpha_2}, \quad \alpha \in \mathbb{T}\backslash\{\widehat{\alpha}\colon \widehat{\alpha} \text{ is a leaf or the root}\} =: \widehat{\mathbb{T}}. \tag{3.2}$$

The basis can be constructed recursively by

$$U^\alpha[x_\alpha, k_\alpha] = \sum_{k_1=1}^{r_{\alpha_1}} \sum_{k_2=1}^{r_{\alpha_2}} B^\alpha[k_1, k_2, k_\alpha] U^{\alpha_1}[x_{\alpha_1}, k_1] U^{\alpha_2}[x_{\alpha_2}, k_2], \quad \alpha \in \widehat{\mathbb{T}}, k_\alpha = 1, \ldots, r_\alpha \tag{3.3}$$

where B^α is an order three coefficient tensor and x_α denotes the tuple of variables represented by α.

For any element $\mathbf{V}$ of $\mathcal{U}_{\alpha^*} = \mathcal{U}_{\alpha_1^*} \otimes \mathcal{U}_{\alpha_2^*}$, there exists a basis representation with respect to the associated subspaces such that, for multi-index $\mathbf{x}_{\alpha^*} = (x_{\alpha_1^*}, x_{\alpha_2^*})$,

$$\mathbf{V}[\mathbf{x}_{\alpha^*}] = \sum_{k_{\alpha_1^*}=1}^{r_{\alpha_1^*}} \sum_{k_{\alpha_2^*}=1}^{r_{\alpha_2^*}} B^{\alpha^*}[k_{\alpha_1^*}, k_{\alpha_2^*}] U^{\alpha_1^*}[x_{\alpha_1^*}, k_{\alpha_1^*}] U^{\alpha_2^*}[x_{\alpha_2^*}, k_{\alpha_2^*}]. \tag{3.4}$$

Expanding this structure recursively using (3.3), one obtains a tree network structure which represents a multilinear low-rank subspace approximation, see [1] for more details. The essential point is that only the order three component tensors $B^\alpha[k_{\alpha_1}, k_{\alpha_2}, k_\alpha]$, $\alpha \in \mathbb{T}$, (or order two for the root α^*) need to be determined in order to obtain the subspace representation. The overall storage complexity is determined by the component tensor tuple $(B^\alpha)_{\alpha \in \mathbb{T}}$ which, in contrast to e.g. the Tucker format, scales only polynomially in the dimension while linear algebra operations can be carried out with a similarly low complexity as with other tensor formats.

3.1. Tensor Train format.
A popular subclass of the hierarchical tensors introduced above is the tensor train (TT) format. It corresponds to a linearized unsymmetric tree, where for every index set $\alpha \in \widehat{\mathbb{T}}$ it holds $\alpha := \{1, \ldots, j\} = \alpha_1 \cup \alpha_2$ for $\alpha_1 := \{1, \ldots, j-1\}$, $\alpha_2 := \{j\}$ denoting nodes

in $\mathbb{T}$. Its parametrized form[2] renders it simpler to handle while maintaining the main features of more general hierarchies, see [1] and [18, Sec. 11]. Historically, this format has been known for a long time in quantum chemistry and physics under the name of *matrix product states* (MPS). The TT format became known to a broader community in applied mathematics by recent publications such as [28, 19] and [23] and we provide a brief overview in this section. Subsequently, the nodes $\alpha_j = \{1, \ldots, j\} \in \widehat{\mathbb{T}}$ are denoted by j for the sake of simplicity.

Consider an element $\mathbf{V} \in \bigotimes_{m=1}^{M} \widetilde{\mathcal{V}}_m$ in the topological tensor product spaces (3.1). We say $\mathbf{V}$ is in TT format if there exists a rank vector $\mathbf{r} = (r_1, \ldots, r_M) \in \mathbb{N}^M$ such that $\mathbf{V}$ admits the representation

$$\mathbf{V}[x_1, \ldots, x_M] = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} V_1[k_1, x_1] V_2[k_1, x_2, k_2] \cdots V_M[k_{M-1}, x_M] \qquad (3.5)$$

with *core matrices* $V_j[x_j] := (V_j[k_{j-1}, x_j, k_j])_{k_{j-1}, k_j} \in \mathbb{R}^{r_{j-1}, r_j}$ and $r_0 = r_M = 1$ by definition. Hence, every entry of $\mathbf{V}$ is determined by the matrix product

$$\mathbf{V}[x_1, \ldots, x_M] = V_1[x_1] \cdots V_M[x_M]. \qquad (3.6)$$

The key element to the construction of such a TT representation is the *higher order singular value decomposition* (HOSVD) [28, Thm 2.1]. We discuss the respective algorithm briefly and also point out a practical complexity reduction via truncation. The procedure is described for the finite-dimensional case where we use discrete variables μ instead of continuous variables x: Given a tensor $\mathbf{U} \in \bigotimes_{m=1}^{M} \widetilde{\mathcal{V}}_m$ with $\widetilde{\mathcal{V}}_m := \mathbb{R}^{n_m}$, $n_m \in \mathbb{N}$ for $m = 1, \ldots, M$ and a multirank $\mathbf{r} = (r_1, \ldots, r_M)$, consider, for indices $\mu = (\mu_1, \ldots, \mu_M) \in \bigtimes_{m=1}^{M} \{1, \ldots, n_m\}$, the *unfolding matrix*

$$\mu \mapsto A_k[\mu_1, \ldots \mu_k; \mu_{k+1}, \ldots, \mu_M] := \mu \mapsto \mathbf{U}[\mu], \qquad (3.7)$$

corresponding to a reshaping of the tensor. Due to the assumption that $\mathbf{U}$ admits a TT representation, the unfolding A_1 of $\mathbf{U}$ has rank r_1. Hence, there exists the QR decomposition $A_1 = QR^T$ which yields

$$A_1[\mu_1; \mu_2, \ldots, \mu_M] = Q[\mu_1] R[\mu_2, \ldots, \mu_M] = \sum_{\alpha_1=1}^{r_1} Q[\mu_1, \alpha_1] R[\alpha_1, \mu_2, \ldots, \mu_M]. \qquad (3.8)$$

Due to orthogonality, $R = A_1^T Q (Q^T Q)^{-1} =: A_1^T G_1$ and it follows

$$R[\mu_1, \mu_2, \ldots, \mu_M] = \sum_{\alpha_1=1}^{r_1} \mathbf{U}[\alpha_1, \mu_2, \ldots, \mu_M] G_1[\mu_1, \alpha_1]. \qquad (3.9)$$

By construction of the unfolding matrices of $\mathbf{U}$, one can deduce that rank $R \leq r_1$. Thus, the process can be repeated inductively for the next index tuple (α_1, μ_2) to obtain the remaining cores $G_k(\alpha_{k-1}, \mu_k, \alpha_k)$ for $k = 2, \ldots, M$, leading to the TT representation

$$\mathbf{U}[\mu_1, \ldots, \mu_M] = \sum_{\alpha_1, \ldots, \alpha_{M-1}}^{r_1, \ldots, r_{M-1}} G_1[\mu_1, \alpha_1] G_2[\alpha_1, \mu_2, \alpha_2] \ldots G_M[\alpha_{M-1}, \mu_M]. \qquad (3.10)$$

By replacing the QR decomposition in the construction with a singular value decomposition (SVD) $A_1 = U \Sigma V^T$ with $\Sigma = \operatorname{diag}(\sigma_1, \ldots, \sigma_{r_1})$ containing the singular values of A_1, we can replace Σ by the truncated $\Sigma_s = \operatorname{diag}(\sigma_1, \ldots, \sigma_s)$ for $s \leq r_1$. This fixed s-term truncation (hard thresholding) yields an approximation to A_1 which is optimal in the Frobenius norm of matrices. It can be used to *round* a tensor to some prescribed rank which might make numerical computations feasible again when too large ranks otherwise would render it impossible. Nevertheless, this process is an approximation and the error has to be controlled. For a more rigorous treatment of these aspects, we refer to [1, Sec. 3.7]. An error bound for the approximation with a rank-$\mathbf{r}$ tensor is given by the next lemma.

---

[2]The tensor train format parametrizes every element of a fixed rank mannifold by the cores in the representation (3.5).

Lemma 3.1 ([28] Theorem 2.2). *Given a tensor $\mathbf{U} \in \bigotimes_{m=1}^{M} \mathbb{R}^{n_m}$ and a rank vector $\mathbf{r} = (r_1, \ldots, r_M)$, the* higher order SVD *with hard thresholding to* $\mathbf{r}$ *yields an approximation* $\mathbf{U}_r := HOSVD(\mathbf{U})$ *with*

$$\|\mathbf{U} - \mathbf{U}_r\|_F^2 \leq \sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \sigma_k[m]^2, \tag{3.11}$$

where σ_k is the vector of singular values of the k-th unfolding matrix (3.7).

Remark 3.2. The storage complexity for a tensor in the TT format can be estimated by

$$\mathcal{O}(nr^2 M), \tag{3.12}$$

where $n = \max\{n_i : i \in \{1, \ldots, M\}\}$, $r = \max\{r_i : i \in \{1, \ldots, M-1\}\}$. Hence, it is apparent that the otherwise exponential complexity growth with the number of dimensions (*curse of dimensionality*) can be reduced to a polynomial complexity.

Remark 3.3. The *Hadamard* product of two TT-tensors $\mathbf{U}$ and $\mathbf{V}$ is the element-wise multiplication in the full tensor representation. It can be extended easily to the infinite-dimensional setting of multivariate continuous functions in which case it is the point-wise multiplication defined by

$$\mathbf{D}[x_1, \ldots, x_M] = (\mathbf{U} \circ \mathbf{V})[x_1, \ldots, x_M]. \tag{3.13}$$

The Hadamard product is a binary operation between two tensors. When applying this operation in the TT-format,

$$\begin{aligned}
\mathbf{D}[x_1, \ldots, x_M] &= U_1[x_1] \cdots U_M[x_M] V_1[x_1] \cdots V_M[x_M] \\
&= (U_1[x_1] \otimes V_1[x_1]) \cdots (U_M[x_M] \otimes V_M[x_M]) \\
&=: D_1[x_1] \cdots D_M[x_M],
\end{aligned}$$

the resulting TT-tensor exhibits a rank which is smaller or equal the product of the individual representation ranks.

3.2. Stochastic Galerkin FEM in an extended TT format.
The availability of the forward map G is a prerequisite for the evaluation of the potential Φ. Similar to [14], we introduce a discretization with finite elements and gPC polynomials in the TT format. In case of uniform random variables, the orthonormal system consists of tensorized Legendre polynomials.

Let

$$\mathcal{V}_M := \mathcal{X} \otimes \mathcal{Y}_M := H_0^1(D) \otimes \left(\bigotimes_{m=1}^{M} L_{\pi_m}^2([-1,1]) \right), \qquad M \geq 1, \tag{3.14}$$

and set $\mathcal{V} := \mathcal{V}_\infty$. Moreover, define the set of finitely supported multi-indices by

$$\mathcal{F}_M := \{\mu \in \mathbb{N}_0^M : |\operatorname{supp} \mu| < \infty\}, \qquad M \geq 1, \tag{3.15}$$

and set $\mathcal{F} := \mathcal{F}_\infty$. Assume univariate Legendre polynomials P_j of degree j, orthonormal in $L_{\pi_0^1}^2([-1,1])$. Then, the tensorized polynomials $\{P_\mu(y) := \prod_{j \in \operatorname{supp} \mu} P_j(y_j)\}_{\mu \in \mathcal{F}_M}$ form an orthonormal basis of $\mathcal{V}_M$ with respect to the canonical inner product $\langle \cdot, \cdot \rangle_{\mathcal{V}_M}$.

We consider a TT representation of the solution q of model problem (2.2) in a discrete subspace of the solution space $\mathcal{V}$. For this, we define a multi-index set based on the dimension vector $(n_m)_{m=1,\ldots,M} \in \mathbb{N}_0^M$ by

$$\Lambda_M := \{(\mu_1, \ldots, \mu_M) \in \mathcal{F} : \mu_m = 0, \ldots, n_m - 1; \ m = 1, \ldots, M\}, \tag{3.16}$$

and obtain the semi-discrete space

$$\mathcal{V}(\Lambda_M) := \left\{ v(x,y) = \sum_{\mu \in \Lambda_M} v_\mu(x) P_\mu(y) : v_\mu \in \mathcal{X} \right\} \subset \mathcal{V}. \tag{3.17}$$

For the discretization of the physical space $\mathcal{X}$, we introduce a conforming finite element space with piecewise polynomials of order p on some simplicial regular triangulation $\mathcal{T}$ of the domain. The

resulting space $\mathcal{X}_p(\mathcal{T}) \subset \mathcal{X}$ is spanned by FE basis functions $\{\varphi_i\}_{i=0,\dots,N_\mathcal{X}-1}$ and $N_\mathcal{X} = \dim \mathcal{X}_p$. For the sake of simplicity, we assume $\mathcal{T}$ to exactly represent D. Hence, the fully discrete space is given by

$$\mathcal{V}_p(\Lambda_M, \mathcal{T}) := \left\{ v(x,y) = \sum_{\mu \in \Lambda_M} v_\mu(x) P_\mu(y) \ : \ v_\mu \in \mathcal{X}_p(\mathcal{T}) \right\} \subset \mathcal{V}(\Lambda_M). \tag{3.18}$$

The Galerkin projection of solution q of (2.2) onto $\mathcal{V}_p(\Lambda_M, \mathcal{T})$ is then obtained by solving a tensor system for coefficient tensor $U \in \mathbb{R}^{N_\mathcal{X} \times n_1 \times \cdots \times n_M}$,

$$A(U) = F. \tag{3.19}$$

Further details on the problem structure can be found in [14]. Upon solving (3.19) and performing a TT compression, e.g. by the HOSVD, we obtain the solution map representation

$$G^{h,N}(x,y) = \sum_{k=1}^{N} \sum_{\mu \in \Lambda_M} U[k,\mu] \varphi_k(x) P_\mu(y) \tag{3.20}$$

$$= \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} \left(\sum_{\mu_0=0}^{N_\mathcal{X}} U_0[\mu_0, k_1] \varphi_{\mu_0}(x) \right) \times$$

$$\times \left(\sum_{\mu_1=0}^{n_1} U_1[k_1, \mu_1] P_{\mu_1}(y_1) \right) \cdots \left(\sum_{\mu_M=0}^{n_M} U_M[k_{M-1}, \mu_M] P_{\mu_M}(y_M) \right),$$

and U exhibits the TT structure

$$\mathbf{U}[x, y_1, \dots, y_M] = U_0[x] U_1[y_1] U_2[y_2] \cdots U_M[y_M] \quad \text{for } x \in D, y \in \Xi_M. \tag{3.21}$$

Remark 3.4. The approximation parameters M, Λ and $\mathcal{T}$ can be determined adaptively based on reliable a posteriori error estimators for the stochastic and physical discretizations. This was introduced in [11, 12], transferred to the TT setting in [14], and is used in the numerical experiments of Section 6. A hierarchical a posteriori approach was considered in [2].

3.3. **Interpolation in the TT parameter space.**
The representation (3.20) has the property that the basis elements of (3.3) are given by $U^j[x_j, k_j] = P_{k_j}(x_j)$, i.e., the leafs of the dimension tree are evaluations of the gPC bases P_μ instead of a general Tucker basis. We call this the *extended TT format*. For an element $\mathbf{V} \in \mathcal{Y}_M$, the corresponding core tensor is denoted by $\tilde{\mathbf{V}}$ and it holds

$$\mathbf{V}[\mathbf{y}] = \sum_{\mu_1=1}^{\infty} \cdots \sum_{\mu_M=1}^{\infty} \tilde{\mathbf{V}}[\mu_1, \dots, \mu_M] \prod_{i=1}^{M} P_{\mu_i}(y_i), \quad \text{for } \mathbf{y} \in \Xi_M := [-1,1]^M. \tag{3.22}$$

We recall Remark 3.3 regarding the Hadamard product of tensor trains as a pointwise multiplication. In the special case of the extended TT format (3.22) we employ an interpolation approach to solve the computational tasks efficiently. In fact, with the multi-index set Λ_M of the representation at hand, the interpolation can be made exact. Moreover, when using a non-optimal interpolation order for the tensor product, the interpolation errors can be separated from the tensor approximation by e.g. a HOSVD.

Given some function $f : \Xi_M \to \mathbb{R}$, we define the N-th order *tensor product interpolation operator* $\mathcal{I}_N$ by an univariate interpolation basis $\mathbf{L} = (L_m)_{m=1}^{N}$ and interpolation nodes $\hat{\mathbf{y}}_\mu \in \Xi_M$, $\mu \in \Lambda_M$, such that for all $(y_1, \dots, y_M) \in \Xi_M$,

$$\mathcal{I}_N f(y_1, \dots, y_M) := \sum_{\mu \in \Lambda_M} f(\hat{y}_{\mu_1}, \dots, \hat{y}_{\mu_M}) L_{\mu_1}(x_1) \cdots L_{\mu_M}(x_M). \tag{3.23}$$

Here, the multi-index set Λ_M specifies the polynomial degrees used in the representation (3.23). It can be given by (3.16) or any set of multi-dimensional indices of length M and maximal index N. We define the tensor

$$\mathbf{F}[\mu_1, \dots, \mu_d] := f(\hat{y}_{\mu_1}, \dots, \hat{y}_{\mu_M}), \quad \mu \in \Lambda_M. \tag{3.24}$$

For s large enough[3] and $f \in H^s(\Xi_M)$, the interpolation error can be estimated by

$$\|f - \mathcal{I}_N f\|_{\mathcal{Y}} \leq C N^{-s} \|f\|_{H^s(\Xi_M)}. \tag{3.25}$$

The error of (3.25), using the quasi best rank-$\mathbf{r}$ TT representation of $\mathbf{F}$, denoted by $\mathbf{F_r} := \mathrm{HOSVD}(\mathbf{F})$, can be bounded with the previous results (3.11) and (3.23),

$$\|f - \sum_{\mu \in \Lambda_M} \mathbf{F_r}[\mu]\mathbf{L}_\mu\|_{\mathcal{Y}} \leq \|f - \mathcal{I}_N f\|_{\mathcal{Y}} + \|\sum_{\mu \in \Lambda_M} (\mathbf{F}[\mu] - \mathbf{F_r}[\mu])\mathbf{L}_\mu\|_{\mathcal{Y}}.$$

This yields the bound

$$\|f - \sum_{\mu \in \Lambda} \mathbf{F}^*[\mu]\mathbf{L}_\mu\|_{\mathcal{Y}} \leq c\left(N^{-s}\|f\|_{H^s(\Xi_M)} + c_2 \sqrt{\sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \sigma_k[m]^2}\right). \tag{3.26}$$

The constant c_2 can be improved by expanding $\mathcal{I}_N f$ into an orthonormal Legendre basis instead of the interpolation basis with an ill-conditioned Gram matrix. This can be achieved by a rank-one transformation $\otimes_{i=1}^M T_i : \mathbf{F} \to \hat{\mathbf{F}}$. For details see Section 4.1.

Remark 3.5. If we apply the Hadamard product to two tensors $\mathbf{U}$ and $\mathbf{V}$ in the extended TT format, the resulting polynomial degree is twice as large as in the initial tensors. Hence, this is not equivalent to the Hadamard poduct of $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$. However, since we know the degrees of the polynomial bases of $\mathbf{U}$ and $\mathbf{V}$ in each dimension, we can appropriately choose the number of interpolation nodes to eliminate the pure interpolation error in (3.26). Therefore, we approximate the Hadamard poduct of $\mathbf{U}$ and $\mathbf{V}$ by interpolation at the nodes, i.e.,

$$\mathbf{T}[\mu_1, \ldots, \mu_d] := \mathbf{U}[\hat{y}_{\mu_1}, \ldots, \hat{y}_{\mu_d}]\mathbf{V}[\hat{y}_{\mu_1}, \ldots, \hat{y}_{\mu_d}].$$

An error bound for the coefficient tensor Hadamard product can be obtained from (3.26) with $\mathbf{F} = \mathbf{U} \circ \mathbf{V}$ and $\mathbf{F_r} = \mathbf{T_r}$.

4. Bayesian Inversion Using Low-Rank Tensor Approximation

In this Section, we introduce a low-rank approximation of the Bayesian potential (2.12) based on the solution described in Section 3.2. With this, we explain an ODE-based numerical approach to determine the exponential of the TT tensor, required for the evaluation of the Bayesian posterior.

4.1. **Low-rank approximation of the Bayesian potential.** We recall the Radon-Nikodym derivative (2.18), which we aim to evaluate,

$$\frac{\mathrm{d}\pi_\delta}{\mathrm{d}\pi_0}(u) = \frac{1}{Z} \exp\left(-\frac{1}{2}\langle \delta - (\mathcal{O} \circ G)(u), \Gamma^{-1}(\delta - (\mathcal{O} \circ G)(u))\rangle\right). \tag{4.1}$$

In actual computations, we use the approximate solution $G^{h,N}$ of the forward problem given as a multivariate polynomial representation in TT format (3.20), where the discretization parameters h and N determine the maximal element diameter of some simplical FE mesh and the number of indices in the gPC expansion of the random data, respectively. Furthermore, $\hat{x} = (x_1, \ldots, x_K)$ are the nodes in the physical domain D chosen for the measurement operator $(\mathcal{O} \circ G)(y) = (G(y)(x_1), \ldots, G(y)(x_K))$. The observed approximated system response $\mathcal{O} \circ G^{h,N}$ again is a TT tensor whose first core is indexed by $k \in \{1, \ldots, K\}$, enumerating the physical measurements x_k,

$$(\mathcal{O} \circ G^{h,N})(k, y) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} (U_0[k, k_1]) \cdots \left(\sum_{\mu_M=0}^{d_M} U_M[k_{M-1}, \mu_M]P_{\mu_M}(y_M)\right). \tag{4.2}$$

Note that this tensor object represents the stochastic solution at certain measurement points $\hat{x}$ parametrized with a polynomial basis in y. This has to be considered when evaluating the inner product (4.1) in $\mathbb{R}^K$. In order to make the computation feasible and to avoid handling the combinatorially large products of polynomials which would normally arise, we introduce an interpolation

---

[3]note that in our setting f is analytic

in the parametric dimensions. For this, we employ univariate *Chebyshev* nodes of the same order L in every dimension to obtain a full tensor grid in Ξ_M. To make the construction precise, choose $\hat{y}_\nu = (\hat{y}_{\nu_1}, \ldots, \hat{y}_{\nu_M})$ with $\hat{y}_{\nu_j} = \cos\left(\frac{2\nu_j - 1}{2L}\pi\right)$ and $1 \le \nu_j \le L$ for $j = 1, \ldots M$. Then, we evaluate the sum over every tensor core at any combination of nodes and store the results in the corresponding tensor cores $\tilde{U}_0, \ldots, \tilde{U}_M$,

$$
\begin{aligned}
(\mathcal{O} \circ G^{h,N})(k, \hat{y}_\nu) &= \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} (U_0[k, k_1]) \left(\sum_{\mu_1=0}^{d_1} U_1[k_1, \mu_1, k_2] P_{\mu_1}(\hat{y}_{\nu_1}) \right) \times \\
&\quad \times \cdots \times \left(\sum_{\mu_M=0}^{d_M} U_M[k_{M-1}, \mu_M] P_{\mu_M}(\hat{y}_{\nu_M}) \right) \\
&= \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} \tilde{U}_0[k, k_1] \tilde{U}_1[k_1, \nu_1, k_2] \cdots \tilde{U}_M[k_{M-1}, \nu_M] \\
&= \tilde{U}_0[\ell] \tilde{U}_1[\nu_1] \cdots \tilde{U}_M[\nu_M] =: \tilde{U}[k, \nu_1, \ldots, \nu_M],
\end{aligned}
\tag{4.3}
$$

for $k = 1, \ldots, K$. The resulting tensor of the Bayesian potential can then be evaluated pointwise at L^M interpolation nodes in Ξ_M. For $\nu = (\nu_1, \ldots, \nu_M) \in \{1, \ldots, L\}^M =: \Lambda_L$ we define the discrete coefficient tensor

$$
\hat{U}_L^{h,N}[\nu] := \frac{1}{2}(\delta - \tilde{U}[\cdot, \nu])^T \Gamma^{-1}(\delta - \tilde{U}[\cdot, \nu]),
\tag{4.4}
$$

using tensor train operations like summation and Hadamard multiplication.

The interpolation in Ξ_M follows from tensorization of the univariate Lagrange polynomials $L_\mu(y) := \prod_{m=1}^M L_{\mu_m}(y_m)$ in $y \in \Xi_M$ such that $L_\mu(\hat{y}_\nu) = \delta_{\mu\nu}$ for $\mu, \nu \in \Lambda_L$. An approximation of the Bayesian misfit function (2.12) is then given by

$$
\Phi_L^{h,N}(y; \delta) \colon \Xi \to \mathbb{R}, \quad (y_1, y_2, \ldots) \mapsto \sum_{\mu \in \Lambda_L} \hat{U}_L^{h,N}[\mu] L_\mu(y).
\tag{4.5}
$$

This representation does not depend on the physical space anymore and satisfies the following property.

Lemma 4.1. *Let $\Phi^{h,N}(\cdot; \delta) \in H^s(\Xi)$, $s > 0$, be an approximation of the Bayesian potential which is Lipschitz in the first argument. Given a prescribed maximal tensor rank $\mathbf{r} \in \mathbb{N}^M$, there holds*

$$
\Phi_L^{h,N}(\cdot; \delta) \colon \Xi \to \mathbb{R}, \quad (y_1, y_2, \ldots) \mapsto \sum_{\mu \in \Lambda_L} U_L^{h,N}[\mu] P_\mu(y),
\tag{4.6}
$$

where $(P_\mu)_{\mu \in \Lambda_L} = \left(\prod_{m=1}^L P_{\mu_m} \right)_{\mu \in \Lambda_L}$ are the orthonormal Legendre polynomials as in Theorem 2.6 and $U_L^{h,N}$ is the coefficient tensor given by (4.4) with a one-dimensional basis change in every tensor core by multiplication with $T = (t_{i,j})_{i,j=1,\ldots,L}$, where

$$
t_{i,j} := \langle L_i, P_j \rangle = \int_{-1}^1 L_\mu(x) P_\nu(x) \, \mathrm{d}x.
\tag{4.7}
$$

Furthermore, there exists a constant $C > 0$ such that

$$
\|\Phi^{h,N}(\cdot; \delta) - \Phi_L^{h,N}(\cdot; \delta)\|_{L_{\pi_0}^2(\Xi)} \le C \left(\sqrt{\sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \sigma_k[m]^2} + L^{-s} \|\Phi(\cdot; \delta)\|_{H^s(\Xi)} \right),
\tag{4.8}
$$

where σ_k are the singular values of the unfolding matrices of the coefficient tensor $U_L^{h,N}$.

Proof. The proof is an immediate consequence of (3.26) applied to the construction of the Bayesian potential approximation. $\qquad\square$

Remark 4.2. The Euclidean inner product of the approximate Bayesian potential (4.4) (expanded in (3.20)) can be calculated more efficiently by considering the sum of the terms $\hat{U}_L^{h,M}[\nu] =: A + B + C$,

$$A := \delta^T \Gamma^{-1} \delta, \tag{4.9}$$

$$B := -2\delta^T \Gamma^{-1} \tilde{U}[\cdot, \nu], \tag{4.10}$$

$$C := \tilde{U}[\cdot, \nu]^T \Gamma^{-1} \tilde{U}[\cdot, \nu]. \tag{4.11}$$

The first term A is a real number which has to be represented as an M-dimensional tensor product, constant in all dimensions. This can be carried out by setting $\tilde{A} = A \otimes e \otimes \cdots \otimes e$, where e is the vector of all ones. The matrix tensor product in the mixed term B acts only on the first core of the interpolated solution. In order to avoid rank-increasing operations, we do not implement the summation of the remaining vector product but instead treat the physical and stochastic dimensions separately. For this, we create a copy of the TT tensor $\tilde{U}[\cdot, \nu]$, set all entries in the first core to zero, multiply it by $K - 1$ and add the result to the already calculated $\Gamma^{-1}\tilde{U}[\cdot, \nu]$. Computationally more involved is the quadratic term C. Usually the TT-ranks add up with every summation and multiply with every multiplication. Hence, while increasing the number of measurements K clearly adds information to the problem, this may also lead to a substantial growth in the tensor ranks which has to be compensated by a compression of the low-rank representation. Nevertheless, the proposed splitting of the inner product allows for a parallel computation and reduces the rank increase when compared to a straightforward approach.

Remark 4.3. Note that the proposed interpolation error in (4.8) can be reduced for $L \in \mathbb{N}$ sufficiently large, namely twice the maximal polynomial degree of the solution in (3.20). The Bayesian potential calculated from the tensor representation,

$$\Phi^{h,N}(y; \delta) = \|\delta - (\mathcal{O} \circ G^{h,N})(y)\|_\Gamma, \tag{4.12}$$

is a polynomial depending on the active multiindex set $\Lambda_M \subset \mathcal{F}$. With this adaptively constructed set at hand, we can easily choose an interpolation degree L sufficiently large to yield an exact representation. To make this more efficient, we can choose the interpolation degree separately for each parameter dimension according to the anisotropic active set Λ_M resulting from an adaptive SFEM as described in [14] which consequently leads to fewer interpolation nodes in Ξ_M.

4.2. **Exponential of a TT-Tensor.**

The evaluation of the exponential of the low-rank Bayesian potential $\Phi^{h,N}$ of Section 4.1 is the next important step to determine the posterior in (4.1). Since $\Phi^{h,N}$ is a high-dimensional object, the computation is not straigtforward. To obtaining a closed form representation, we derive a TT representation from (4.5), analog to (4.3),

$$\widehat{\Phi}(\hat{y}; \delta) := \sum_{\mu \in \Lambda_L} U_L^{h,N}[\mu] P_\mu(\hat{y}), \quad \hat{y} \in \Xi_M. \tag{4.13}$$

Here, the exponential function needs to be evaluated pointwise for every interpolation node $\hat{y}$ without leaving the tensor format to stay efficient. In [27], several possibilities to calculate the matrix exponential are described, which in principle could also be applied in the tensor framework. We make use of the well-known fact that the exponential is the solution of the basic ordinary differential equation (ODE)

$$\frac{\mathrm{d}}{\mathrm{d}t} W(t, \hat{y}, \delta) = -W(t, \hat{y}; \delta) \circ \widehat{\Phi}(\hat{y}, \delta), \quad \text{with } W(0, \hat{y}, \delta) = 1. \tag{4.14}$$

The solution to this initial value problem is given by

$$W(t, \hat{y}, \delta) = \exp(-t\widehat{\Phi}(\hat{y}, \delta)), \tag{4.15}$$

and $W(1, \hat{y}, \delta)$ is equivalent to the exponential of the negative Bayesian potential. Assuming the terms of the initial value problem (4.14) are tensor trains, the problem translates to a low-rank approximation of the original system interpreting the point-wise multiplication as Hadamard product. It remains to solve this ODE in the TT format, which is discussed subsequently.

Remark 4.4. In [26], the problem of quasi-optimal approximation on larger time-scales is explained. It gives rise to dynamical algorithms which stay in the desired rank manifold without the need of hard-thresholding.

4.2.1. *Runge-Kutta methods.* Numerical methods for the solution of ODEs are a classical topic [21, 3]. For our purposes, we adapt a well-known class of explicit s-stage Runge-Kutta methods to the TT-tensor framework. Runge-Kutta schemes are usually described by Butcher-Tableaus of the form

$$\begin{array}{c|c} \mathbf{c} & A \\ \hline & \mathbf{b^T} \end{array} \tag{4.16}$$

where, in explicit approaches, $A = [a_{i,j}]$ is a strict lower-triangular $s \times s$ matrix. We cite a result for the general scheme.

Proposition 4.5. *Let* $Y(t)$ *be the unique solution to the initial value problem*

$$\frac{\mathrm{d}}{\mathrm{d}t}Y(t) = f(t, Y(t)), \qquad Y(0) = Y_0. \tag{4.17}$$

Then, the approximation method given by a temporal grid $(t_l)_{l \geq 0} = (l\tau)_{l \geq 0})$ *with step width* $\tau > 0$, *the Butcher-tableau* (4.16) *and the iterative procedure*

$$Y_{n+1} = Y_n + \tau \sum_{i=1}^{s} b_i k_i \quad \text{with} \quad k_i = f\left(t_n + c_i\tau, Y_n + \tau \sum_{j=1}^{i-1} a_{i,j}k_j\right), \tag{4.18}$$

is consistent if and only if $\sum_{i=1}^{s} b_i = 1$.

Assuming an admissible right-hand side f (continuous in the first argument and Lipschitz in the second argument), we obtain convergence rates equal to the resulting consistency rate, which depends on the choice of A, b and c. Examples are the explicit Euler scheme with convergence order 1,

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}, \tag{4.19}$$

the Heun method with convergence order 2,

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}, \tag{4.20}$$

and the classical Runge-Kutta method which is convergent of order 4,

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array}. \tag{4.21}$$

Remark 4.6. In our context, we only need to consider a stationary right-hand side f. Hence, the calculations are independent of the nodes vector c. However, the stability and convergence of the considered method still depends on c.

For the evaluation of the Bayes potential in TT format, we have to consider (4.14) as tensor ODE and extend the iteration scheme (4.18) to TT tensors. This is formally described in Algorithm 1. Figure 1 pictures the convergence of $V_{\exp}$ for different numerical schemes. The computation was executed for a randomly created order five TT tensor with dimensions up to 10 and ranks up to 20. The error for this plot is determined approximately by Monte-Carlo sampling of the tensor and comparison with the exact pointwise exponential. One can observe the expected convergence rates for the different numerical schemes. The implicit and explicit Euler schemes exhibit the same convergence but the implicit method yields a good approximation already for less than 10 iteration steps.

Algorithm 1: explicit Runge-Kutta method for TT tensors

Require: TT tensor V, number of iterations N, maximal rank r, rounding precision ϵ,
    Butcher-tableau (A, b)
1: $V_{exp} = (1, \ldots, 1) \otimes \cdots \otimes (1, \ldots, 1)$ according to dimensions and ranks of V.
2: **for** $\ell = 1 : N$ **do**
3:    **if** $\max(\text{TT-ranks of } V_{exp}) > r$ **then**
4:       $V_{exp} \leftarrow$ hard-thresholding to maximal rank r and precision ϵ.
5:    **end if**
6:    $\text{summ} = 0$
7:    **for** $i = 1 : s$ **do**
8:       **for** $j = 1 : i - 1$ **do**
9:          $\text{summ} = \text{summ} + a_{i,j} k_j$
10:      **end for**
11:     $k_j = \tau f(V_{exp} + \text{summ}) = \tau V \circ V_{exp}$
12:    **end for**
13:    $\text{summ} = 0$
14:    **for** $i = 1 : s$ **do**
15:       $\text{summ} = \text{summ} + b_i k_i$
16:    **end for**
17:    $V_{exp} = V_{exp} + \text{summ}$
18: **end for**
19: **return** V_{exp}
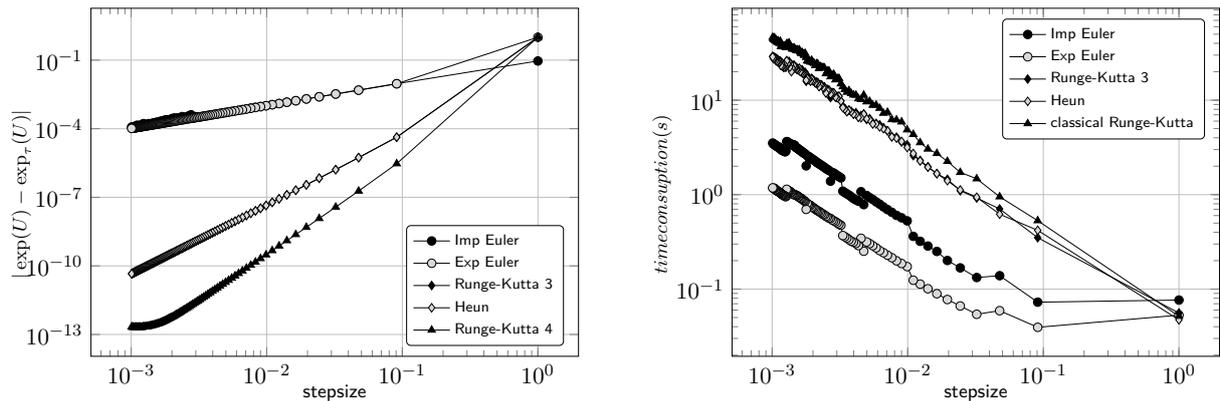
FIGURE 1. Convergence rates of error (left) and measured time (right) of different Runge-Kutta schemes for the evaluation of the tensor exponential for decreasing step sizes. The error is determined by Monte-Carlo sampling with respect to the point-wise exact exponential.

Remark 4.7. Note that the tensor multiplication in line 11 has to be carried out element-wise, i.e. in the Hadamard sense. It is apparent that with regard to complexity, the iteration process is dominated by these tensor multiplications. As mentioned above, this leads to a strong increase of tensor ranks due to the multiplication of the respective ranks of the involved tensors.

Since explicit schemes are not unconditionally stable and hence may require a very small step size and many iterations, the inevitable increase of tensor ranks in each step becomes a pressing issue. To keep the scheme computationally feasible, the tensor has to be recompressed to a prescribed tolerance in each iteration.

4.2.2. *Adaptive step size Runge-Kutta algorithm.* Adaptivity can often greatly reduce the computational complexity while maintaining a high accuracy of the solution. We hence discuss an embedded

step size control of the order q Runge-Kutta methods as described in Section 4.2.1. The main idea is to take a second Runge-Kutta approximation of higher order $q' > q$ which reuses the already calculated increments k_i of the order q scheme, i.e.,

$$Y_{n+1}^* = Y_n^* + \tau_n \sum_{i=1}^{s} b_i^* k_i.$$

The resulting error

$$e_{n+1} = \|Y_{n+1} - Y_{n+1}^*\|_F = \tau_n \|\sum_{i=1}^{s} (b_i - b_i^*) k_i\|_F \tag{4.22}$$

is of order q and it can be used as an error indicator to adaptively adjust the step size. Given an initial step size, a desired approximation tolerance $\mathrm{tol} > 0$ and a delay parameter $0 < \beta < 1$, we obtain an optimal step size for the next iteration step by

$$\tau_{n+1} = \begin{cases} \beta \tau_n \left(\frac{\mathrm{tol}}{e_{n+1}} \right)^{\frac{1}{q}} & \text{for } e_{n+1} >= \mathrm{tol} \\ \beta \tau_n \left(\frac{\mathrm{tol}}{e_{n+1}} \right)^{\frac{1}{q}+1} & \text{else.} \end{cases} \tag{4.23}$$

Proposition 4.8. *Assume $0 < q < q' < \infty$ and two Runge-Kutta methods with convergence of order q and q'. Then, the embedded Runge-Kutta method for (4.17) converges with order q.*

Proof. We have

$$Y_1 = Y(\tau) + \mathcal{O}(h^{q+1}) \quad \text{and} \quad Y_1^* = Y(\tau) + \mathcal{O}(h^{q'+1}). \tag{4.24}$$

Hence,

$$\|e_1\| = \|\mathcal{O}((h^{q+1}) + \mathcal{O}(h^{q'+1})\|. \tag{4.25}$$

This means, we obtain only an estimate of the lower order method. □

Remark 4.9. In fact, the error bound in (4.28) consists of multiple instances of the tensor rounding error, accumulated due to the iterative thresholding. Nevertheless, it is possible to calibrate the magnitudes of the rounding error $\sqrt{\sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \tilde{\sigma}_k[m]^2}$ and the general Runge-Kutta error of order τ^q by coupling the step size parameter to the rounding rank and using the condition

$$e_{n+1} = \tau_n \|\sum_{i=1}^{s} (b_i - b_i^*) k_i\|_F = \beta \tau^{q+1} + \sqrt{\sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \tilde{\sigma}_k[m]^2} = \mathrm{tol} > 0. \tag{4.26}$$

Possible rank adaptive extensions to (4.23) can be achieved by

1 prescribing a tolerance for the truncation of the singular values which can be modified if the new step size leaves a certain domain.
2 prescribing a rounding rank and use a hard thresholding or e.g. a rank preserving optimization such as the *Alternating Least Square* (ALS). Then, the rounding rank is increased if the step size is to small or decreased otherwise.

4.2.3. *Implicit Method.* Usually, the previously defined methods are convergent only in specific stability regions. In order to alleviate such limitations, we can take advantage of implicit methods which are unconditionally stable. The simplest implicit method is the first order convergent backward Euler scheme which can be written as

$$W_{k+1}(\hat{y}, \delta) = W_k(\hat{y}, \delta) + \tau \widehat{\Phi}(\hat{y}, \delta) \circ W_{k+1}(\hat{y}, \delta)$$

In every iteration step of the implicit Euler scheme a system of linear equations has to be solved, which in our setting consists of TT tensors. Solving linear equation system involving TT tensors is e.g. examined in [23]. For our computations, we use the *Alternating Minimal Energy* (AMEN) [10] algorithm with a random starting tensor. In experiments, as a result of its unconditional stability, the implicit Euler scheme already converges for a small number of steps, i.e., a large step size. We now state the main result of this section.

Algorithm 2: Implicit Euler method for TT tensors

Require: TT tensor V, number of iterations N, maximal rank r, rounding precision ϵ

1: $O = (1, \ldots, 1) \otimes \cdots \otimes (1, \ldots, 1)$ according to dimensions and ranks of V

2: $V_{\exp} = O$

3: **for** $j = 1 : N$ **do**

4:    **if** $\max(\text{TT ranks of } V_{\exp}) > r$ **then**

5:       $V_{\exp} \leftarrow$ hard-thresholding to maximal rank r and precision ϵ

6:    **end if**

7:    $V_{exp} = \mathrm{diag}(O - \frac{1}{N}V, \ldots, O - \frac{1}{N}V)V_{\exp}$

8: **end for**

9: **return** $V_{\exp}$

Lemma 4.10. *Let $\Phi_L^{h,N}$ be as in* (4.6) *an approximation of the Bayesian potential which is Lipschitz in the first argument. Assume a stable one-step Runge-Kutta method of convergence order $q \geq 1$ with step size $\tau > 0$. Then, for a prescribed maximal tensor rank $\mathbf{r} \in \mathbb{N}^M$, the following mapping is well-defined,*

$$\exp_\tau\left(-\frac{1}{2}\Phi_L^{h,N}(\cdot;\delta)\right) : \Xi \to \mathbb{R}, \quad (y_1, y_2, \ldots) \mapsto \sum_{\mu \in \Lambda_L} U_{L,\tau}^{h,N}[\mu] P_\mu(y), \tag{4.27}$$

where $U_{L,\tau}^{h,N}$ is the coefficient tensor of $\Phi_L^{h,N}$ from (4.6) *applied to* (4.14)*. Furthermore, if $e^{-\frac{1}{2}\Phi_L^{h,N}} \in H^{\tilde{s}}(\Xi)$ for some $\tilde{s} > 0$ then there exists a constant $\widetilde{C} > 0$ such that*

$$\|\exp_\tau\left(-\frac{1}{2}\Phi_L^{h,N}\right) - \exp\left(-\frac{1}{2}\Phi_L^{h,N}\right)\|_{L^2_{\pi_0}(\Xi)} \leq \widetilde{C}\left(\sqrt{\sum_{k=1}^{M-1}\sum_{m=r_k+1}^{n_k}\tilde{\sigma}_k[m]^2}\right. \tag{4.28}$$

$$\left. + L^{-\tilde{s}}\|e^{-\frac{1}{2}\Phi_L^{h,N}}\|_{H^{\tilde{s}}(\Xi)} + \tau^q\right), \tag{4.29}$$

where $\tilde{\sigma}$ are the singular values of the unfolding matrices of the coefficient tensor $U_L^{h,N}$.

Proof. The result follows directly from the stability of the Runge-Kutta method and (3.26). □

Corollary 4.11. *Using Remark 4.9 we can define a combined error bound $\tilde{\tau}$, which contains approximations due to the iterative step size algorithm and the rounding procedure. Hence, inequality* (4.28) *becomes*

$$\|\exp_\tau\left(-\frac{1}{2}\Phi_L^{h,N}\right) - \exp\left(-\frac{1}{2}\Phi_L^{h,N}\right)\|_{L^2_{\pi_0}(\Xi)} \leq \widehat{C}\left(L^{-\tilde{s}}\|e^{-\frac{1}{2}\Phi_L^{h,N}}\|_{H^{\tilde{s}}(\Xi)} + \tilde{\tau}^q\right), \tag{4.30}$$

with some $\tilde{C} > 0$.

4.3. **Low-rank approximation of the Bayesian posterior.** This section is concerned with the derivation of an explicit representation of the Bayesian posterior density. This is based on the results of the preceding sections, namely the TT representation of the Bayesian potential in terms of multivariate polynomials (4.5) and the approximation of the point-wise exponential of a TT tensor by some higher-order Runge-Kutta method as discussed in Section 4.2. Combining both ideas with a suitable orthogonal polynomial basis leads to an explicit functional representation of the approximate Bayesian posterior density

$$\exp_\tau\left(-\frac{1}{2}\Phi_L^{h,N}(y;\delta)\right) = \sum_{\mu \in \Lambda_L} U_{L,\tau}^{h,N}[\mu] P_\mu(y), \quad \text{for } y \in \Xi. \tag{4.31}$$

Employing the TT tensor structure (3.5) and setting $V_k := (U_{L,\tau}^{h,N})_k$ for the tensor cores $k = 0, \ldots, M$, we get

$$\exp_\tau\left(-\frac{1}{2}\Phi_L^{h,N}(y;\delta)\right) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{M-1}=1}^{r_{M-1}} V_0[k_1]\left(\sum_{\mu_1=0}^{L} V_1[k_1, \mu_1, k_2]P_{\mu_1}(y_1)\right) \times \tag{4.32}$$

$$\times \cdots \times \left(\sum_{\mu_M=0}^{L} V_M[k_{M-1}, \mu_M]P_{\mu_M}(y_M)\right).$$

The benefits of the representation in (4.32) become apparent when estimating otherwise expensive quantities as e.g. the normalization factor Z from (2.18),

$$Z = \mathbb{E}_{\pi_\delta}[1] = \mathbb{E}_{\pi_0}\left[\exp\left(-\frac{1}{2}\Phi(\cdot,\delta)\right)\right]. \tag{4.33}$$

The evaluation of this high-dimensional integral is a challenging task and sampling methods such as Markov-Chain Monte Carlo (MCMC) are quite popular for instance because of their simple approach to approximate this constant. In contrast to sampling methods, the presented TT tensor setting based on a functional representation in orthogonal polynomials allows for a computation by

$$Z_{L,\tau}^{h,N} = \mathbb{E}_{\pi_0}[\exp_\tau(-\Phi_L^{h,N}(y,\delta))] = \int_{\Xi_M} \exp_\tau(-\Phi_L^{h,N}(y,\delta))\,\mathrm{d}\frac{1}{2}\lambda(y) \tag{4.34}$$

$$= 2^{-M}\sum_{\mu\in\Lambda_L} V[\mu]\int_{-1}^{1} P_{\mu_1}(y_1)\,\mathrm{d}\lambda(y_1)\cdots\int_{-1}^{1} P_{\mu_M}(y_M)\,\mathrm{d}\lambda(y_M) \tag{4.35}$$

$$= 2^{-M}V[0, 0, \ldots, 0]. \tag{4.36}$$

Hence, this single tensor evaluation, consisting of $M+1$ matrix vector multiplications of the cores, enables the evaluation of the approximated Bayesian posteriori density in (2.11) by

$$\frac{\mathrm{d}\pi_{\delta,L,\tau}^{h,N}}{\mathrm{d}\pi_0}(y) = \frac{1}{Z_{L,\tau}^{h,N}}\exp_\tau\left(-\frac{1}{2}\Phi_L^{h,N}(y,\delta)\right). \tag{4.37}$$

This joint density contains all information about the unknown parameter in the model parametrized by the expansion (2.14). For parameter (density) estimation, the tensor train decomposed Bayesian potential can also be used to determine the marginal densities. Setting $y_{-k} = (y_1, \ldots, y_{k-1}, y_{k+1}, \ldots, y_M)$, the k-th marginal density for the parameter y_k is given by

$$\left(\frac{\mathrm{d}\pi_{\delta,L,\tau}^{h,N}}{\mathrm{d}\pi_0}\right)_k(y_k) = \frac{1}{Z_{L,\tau}^{h,N}}\int_{-1}^{1}\exp_\tau\left(-\frac{1}{2}\Phi_L^{h,N}(y,\delta)\right)\,\mathrm{d}\pi_0(y_{-k}) \tag{4.38}$$

$$= 2^{-M+1}\sum_{j=1}^{L} V[0, 0, \ldots, 0, j, 0, \ldots, 0]P_j(y_k), \tag{4.39}$$

where the orthogonality of the Legendre polynomials is used.

Note that the tensor format can also be exploited to efficiently estimate quantities of interest (2.19), e.g. moments of the forward solution with respect to the posterior probability measure.

5. Error Analysis

The presented approach is amenable to a stringent a priori error analysis for the posterior probability measure, which is the topic of this section. In the analysis, we follow the derivations of [36] and also of [8]. In particular, we take into account all approximation errors of the chosen explicit representation. We point out that, in principle, the presented analysis directly allows for an a posteriori error control since computable estimators for all required approximations can be defined.

5.1. **Convergence of the posterior.** In order to quantify the distance of the approximate posterior $\pi_{\delta,L,\tau}^{h,M}$ to the true posterior π_y, the Hellinger distance can be employed. It is defined by

$$
d_{\text{Hell}}(\pi_\delta, \pi_{\delta,L,\tau}^{h,N}) = \left(\frac{1}{2} \int_{\Xi_M} \left(\sqrt{\frac{d\pi_y}{d\pi_0}} - \sqrt{\frac{d\pi_{\delta,L,\tau}^{h,N}}{d\pi_0}} \right)^2 d\pi_0 \right)^{1/2}.
\tag{5.1}
$$

This error measure includes all approximation errors of $\pi_{\delta,L,\tau}^{h,N}$, namely the forward operator approximation by an N-term gPC series and a finite element method with mesh parameter h as described in (3.20), the polynomial interpolation in parameter space of order L in (4.4) and the ODE based tensor exponential approximation with step size τ in (4.14). In [36], an alternative approximation of the forward solution by means of Gaussian processes is presented, which only yields a dependence on the $L_{\pi_0}^2(X)$-norm of the solution operator. In principle, the approach derived in this article is similar, as can be seen from the analysis. However, due to the representation in orthogonal polynomials and the compression of the tensor format, it allows for more efficient and adaptive approximations. With the same arguments as in [36, Lem. 4.1.], it can be verified that

$$
Z_{L,\tau}^{h,N} > 0.
\tag{5.2}
$$

This is a requirement for the following theorem.

Note that we tacitly assume that both, the number of stochastic dimensions M and the dimension N of the gPC basis, are directly determined by the (same) multi-index set Λ.

Theorem 5.1. *Assume that* $\sup_{y \in \Xi} |\mathcal{O}(G(y))|_{\mathbb{R}^K} < \infty$ *and* $\sup_{y \in \Xi} \|G(y) - G^{h,N}(y)\|_{\mathcal{X}} \to 0$ *as* $N \to \infty$, $h \to 0$. *Then, there exists a constant* $C > 0$ *independent of the approximation parameters* h, N, L *and* τ, *such that*

$$
d_{Hell}(\pi_\delta, \pi_{\delta,L,\tau}^{h,N}) \leq C \left(h^{2t} + N^{-2\left(\frac{1}{p}-1\right)} + \sqrt{\sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \sigma_k[m]^2} + L^{-s} + L^{-\tilde{s}} + \tilde{\tau}^q \right),
\tag{5.3}
$$

with σ_k *from Lemma 3.1.*

Proof. We consider the Bayesian posterior density (2.18) and the approximation

$$
\frac{d\pi_{\delta,L,\tau}^{h,N}}{d\pi_0}(\xi) = \frac{1}{Z_{L,\tau}^{h,N}} \exp_\tau(-\Phi_L^{h,N}(\xi;\delta)),
\tag{5.4}
$$

with $\Phi_L^{h,N}(\xi;\delta)$ from (4.5) and $\exp_\tau$ is the approximated exponential as solution of the ODE (4.14). Then, Hölders' inequality yields

$$
\begin{aligned}
2d_{\mathsf{Hell}}(\pi_\delta, \pi_{\delta,L,\tau}^{h,N})^2 &= \mathbb{E}_{\pi_0}\left[\left(\frac{e^{-\frac{1}{2}\Phi(\cdot,\delta)}}{\sqrt{Z}} - \frac{e_\tau^{-\frac{1}{2}\Phi_L^{h,N}(\cdot,\delta)}}{\sqrt{Z_{L,\tau}^{h,N}}}\right)^2\right] \\
&= \frac{2}{Z}\underbrace{\mathbb{E}_{\pi_0}\left[\left(e^{-\frac{1}{2}\Phi(\cdot;\delta)} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}(\cdot,\delta)}\right)^2\right]}_{:=I} - \frac{2}{Z}\underbrace{\mathbb{E}_{\pi_0}\left[e_\tau^{-\Phi_L^{h,N}(\cdot,\delta)}\right]}_{=Z_{L,\tau}^{h,N}} \\
&\quad + \left(\frac{4}{Z} - \frac{2}{\sqrt{ZZ_{L,\tau}^{h,N}}}\right)\mathbb{E}_{\pi_0}\left[e^{-\frac{1}{2}\Phi(\cdot,\delta)}e_\tau^{\frac{1}{2}\Phi_L^{h,N}(\cdot,\delta)}\right] \\
&\leq I - \frac{2Z_{L,\tau}^{h,N}}{Z} + \frac{4\sqrt{ZZ_{L,\tau}^{h,N}}}{Z} - 2 \\
&= I - \underbrace{2Z_{L,\tau}^{h,N}\left(\frac{1}{\sqrt{Z}} - \frac{1}{\sqrt{Z_{L,\tau}^{h,N}}}\right)^2}_{:=II} \leq I + II.
\end{aligned}
\tag{5.5}
$$

We split the difference in I into two distinct error terms by the triangle inequality,

$$
\left|e^{-\frac{1}{2}\Phi(\xi;\delta)} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}(\xi;\delta)}\right| \leq \left|e^{-\frac{1}{2}\Phi(\xi;\delta)} - e^{-\frac{1}{2}\Phi_L^{h,N}(\xi;\delta)}\right| + \left|e^{-\frac{1}{2}\Phi_L^{h,N}(\xi;\delta)} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}(\xi;\delta)}\right|.
\tag{5.6}
$$

Due to the continuity of the observation operator $\mathcal{O}$ and the convergence of the forward operator, the exponential function admits a locally Lipschitz property, which means there exists a $Q > 0$ such that

$$
|e^{-\frac{1}{2}\Phi(\xi;\delta)} - e^{-\frac{1}{2}\Phi_L^{h,N}(\xi;\delta)}| \leq \frac{1}{4}Q|\Phi(\xi;\delta) - \Phi_L^{h,N}(\xi;\delta)|.
\tag{5.7}
$$

Note that the second term of (5.6) quantifies the error of the exponential approximation for a chosen numerical ODE scheme. Another triangle inequality yields the splitting of the potential interpolation error

$$
|\Phi(\xi;\delta) - \Phi_L^{h,N}(\xi;\delta)| \leq |\Phi(\xi;\delta) - \Phi^{h,N}(\xi;\delta)| + |\Phi^{h,N}(\xi;\delta) - \Phi_L^{h,N}(\xi;\delta)|,
\tag{5.8}
$$

where the second term is the interpolation error of the potential.

To simplify the notation, we omit the arguments of $\Phi := \Phi(\cdot, \delta)$ (and of all approximations of Φ). Then, by Young's inequality,

$$\frac{Z}{2}I \leq \frac{1}{16}Q^2\mathbb{E}_{\pi_0}\left[\left(\left|\Phi - \Phi^{h,N}\right| + \left|\Phi^{h,N} - \Phi_L^{h,N}\right|\right)^2\right] \tag{5.9}$$

$$+ \frac{1}{2}Q\mathbb{E}_{\pi_0}\left[\left(\left|\Phi - \Phi^{h,N}\right| + \left|\Phi^{h,N} - \Phi_L^{h,N}\right|\right)\left|e^{-\frac{1}{2}\Phi_L^{h,N}} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}}\right|\right]$$

$$+ \mathbb{E}_{\pi_0}\left[\left|e^{-\frac{1}{2}\Phi_L^{h,N}} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}}\right|^2\right]$$

$$\leq \frac{1}{16}Q^2\mathbb{E}_{\pi_0}\left[\left(\left|\Phi - \Phi^{h,N}\right|\right)^2\right] + \frac{1}{8}Q^2\mathbb{E}_{\pi_0}\left[\left|\Phi - \Phi^{h,N}\right|\left|\Phi^{h,N} - \Phi_L^{h,N}\right|\right]$$

$$+ \frac{1}{16}Q^2\mathbb{E}_{\pi_0}\left[\left(\left|\Phi^{h,N} - \Phi_L^{h,N}\right|\right)^2\right]$$

$$+ \frac{1}{4}Q\mathbb{E}_{\pi_0}\left[\left(\left|\Phi - \Phi^{h,N}\right| + \left|\Phi^{h,N} - \Phi_L^{h,N}\right|\right)^2\right] + \frac{1}{4}Q\mathbb{E}_{\pi_0}\left[\left|e^{-\frac{1}{2}\Phi_L^{h,N}} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}}\right|^2\right]$$

$$+ \mathbb{E}_{\pi_0}\left[\left|e^{-\frac{1}{2}\Phi_L^{h,N}} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}}\right|^2\right]$$

$$\leq \left(\frac{1}{8}Q^2 + \frac{1}{2}Q\right)\mathbb{E}_{\pi_0}\left[\left|\Phi - \Phi^{h,N}\right|^2\right]$$

$$+ \left(\frac{1}{8}Q^2 + \frac{1}{2}Q\right)\mathbb{E}_{\pi_0}\left[\left|\Phi^{h,N} - \Phi_L^{h,N}\right|^2\right]$$

$$+ \left(1 + \frac{1}{4}Q\right)\mathbb{E}_{\pi_0}\left[\left|e^{-\frac{1}{2}\Phi_L^{h,N}} - e_\tau^{-\frac{1}{2}\Phi_L^{h,N}}\right|^2\right]. \tag{5.10}$$

From the reverse triangle inequality, (4.8) and (4.30), it follows

$$\frac{Z}{2}I \leq \left(\frac{1}{8}Q^2 + \frac{1}{2}Q\right)\mathbb{E}_{\pi_0}\left[\left|\Phi^{\frac{1}{2}} + (\Phi^{h,N})^{\frac{1}{2}}\right|^2\|\mathcal{O}\|_{\mathcal{L}(\mathcal{X},Y)}^2\|G - G^{h,N}\|_{L^2(X;\mathcal{X})}^2\right]$$

$$+ \left(\frac{1}{8}Q^2 + \frac{1}{2}Q\right)C\left(\sqrt{\sum_{k=1}^{M-1}\sum_{m=r_k+1}^{n_k}\sigma_k[m]^2} + L^{-s}\|\Phi\|_{H^s(\Xi)}\right)$$

$$+ \left(1 + \frac{1}{4}Q\right)\widehat{C}\left(L^{-\tilde{s}}\|e^{-\frac{1}{2}\Phi_L^{h,N}}\|_{H^{\tilde{s}}(\Xi)} + \tilde{\tau}^q\right). \tag{5.11}$$

By the uniform convergence of the forward operator, we can bound the maximum norm of the potential sum uniformly and independently of M and h, which yields

$$I \leq C_1\left(h^{2t} + N^{-2\left(\frac{1}{p}-1\right)} + \sqrt{\sum_{k=1}^{M-1}\sigma_k[r_k+1]^2} + L^{-s} + L^{-\tilde{s}} + \tilde{\tau}^q\right), \tag{5.12}$$

where

$$C_1 := \frac{2}{Z}\max\left\{\left(\frac{Q^2}{8} + \frac{Q}{2}\right)\sup_{y\in Y}\left|\Phi^{\frac{1}{2}} + \left(\Phi^{h,N}\right)^{\frac{1}{2}}\right|, \tag{5.13}\right.$$

$$C\left(\frac{Q^2}{8} + \frac{Q}{2}\max\{1, \|\Phi\|_{H^2(\Xi)}\}\right), \tag{5.14}$$

$$\left.\widehat{C}\left(1 + \frac{Q}{4}\right)\max\{1, \|e^{-\frac{1}{2}\Phi_L^{h,N}}\|_{H^s(\Xi)}\}\right\}. \tag{5.15}$$

The second term II can be bounded by multiplication with $1 = \left(\frac{Z - Z_L^{h,N}}{Z - Z_L^{h,N}} \right)^2$ and factoring out $\left(Z - Z_{L,\tau}^{h,N} \right)$,

$$
\begin{aligned}
II &= 2Z_{L,\tau}^{h,N} \left(\frac{1}{\sqrt{Z}} - \frac{1}{\sqrt{Z_{L,\tau}^{h,N}}} \right)^2 \\
&\leq 2Z_{L,\tau}^{h,N} \max\left\{ Z^{-3}, \left(Z_{L,\tau}^{h,N} \right)^{-3} \right\} \left(Z - Z_{L,\tau}^{h,N} \right)^2 \\
&= 2Z_{L,\tau}^{h,N} \max\left\{ Z^{-3}, \left(Z_{L,\tau}^{h,N} \right)^{-3} \right\} \left(\mathbb{E}_{\pi_0} \left[e^{-\Phi(\cdot;\delta)} - e_\tau^{-\Phi_L^{h,N}(\cdot;\delta)} \right] \right)^2 .
\end{aligned}
\tag{5.16}
$$

By Jensen's inequality and a similar argument as for I in (5.9), we obtain

$$
II \leq C_2 \left(h^{2t} + N^{-2\left(\frac{1}{p} + 1 \right)} + \sqrt{ \sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \sigma_k[m]^2 + L^{-s} + L^{-\tilde{s}} + \tilde{\tau}^q } \right),
\tag{5.17}
$$

with

$$
C_2 = 2Z_{L,\tau}^{h,N} C_1.
\tag{5.18}
$$

$\square$

Corollary 5.2. *With adaptively chosen interpolation nodes set such that the Bayesian potential is represented exactly, the error bound from Theorem 5.1 can be improved in the sense that*

$$
d_{Hell}(\pi_\delta, \pi_{\delta,L,\tau}^{h,N}) \leq \tilde{C} \left(h^{2t} + N^{-2\left(\frac{1}{p} + 1 \right)} + \sqrt{ \sum_{k=1}^{M-1} \sum_{m=r_k+1}^{n_k} \sigma_k[m]^2 + L^{-\tilde{s}} + \tilde{\tau}^q } \right).
\tag{5.19}
$$

5.2. **Adaptive low rank Bayesian inversion in tensor train format.** In addition to the presented a priori analysis of the preceding section, the proposed approach also allows for a completely *a posteriori adaptive* Bayesian inversion. To achieve this, there have to be computable error estimators or indicators for all approximations carried out in the numerical scheme. The adaptivity concerns the mesh, gPC and rank parameters of the forward model $G^{h,N}$, the interpolation degree L and the time step and rank parameters of the exponential calculation $\exp_\tau$.

The adaptive stochastic Galerkin FEM of [14] contains a reliable error estimator η_{SGFEM}, which guides the refinements of the discretizations and which yields the a posteriori estimate

$$
\| G - G^{h,N} \| \leq C_{\text{SGFEM}} \eta_{SGFEM}.
\tag{5.20}
$$

We summarize the adaptive algorithm for the forward problem with a desired accuracy $\varepsilon_{\text{SGFEM}}$ and weight parameter[4] $C_{\text{SGFEM}} > 0$ in the module $\texttt{Adapt-SGFEM}(\varepsilon_{\text{SGFEM}}, C_{\text{SGFEM}}) \to G^{h,N}$, which results in a tensor representation of the model problem satisfying the prescribed accuracy requirements.

In the same fashion, the module $\texttt{Adapt-}\Phi_L^{h,N}(\varepsilon_L, C_L, L, \Lambda_M) \to \Phi_L^{h,N}$ yields the Bayesian potential (4.6). Here, Λ_M is the adaptively constructed active index set (3.16) of $G^{h,N}$ and L acts as a threshold to cover indices in Λ_M leading to sufficiently large polynomial degrees, i.e. if there exists a $\mu \in \Lambda_M$ with $2\mu_j > L$ for some $j = 1, \dots, M$ we reduce μ_j to the next lower integer of $\frac{L}{2}$. The tolerance $\varepsilon_L > 0$ and the constant $C_L > 0$ are used to satisfy the error bound (4.8).

The adaptive calculation of the tensor exponential $\exp_\tau(-\Phi_L^{h,N})$ is summarized in the module $\texttt{Adapt-}\exp_\tau(\varepsilon_\tau, C_\tau, \tau_0) \to \exp_\tau(-\Phi_L^{h,N})$. Here, we consider $\varepsilon_\tau =: \text{tol}$ and $C_\tau =: \beta$ in (4.23). The input τ_0 is the initial step size.

Successively combining the results of these modules, Algorithm 3 illustrates the fully adaptive numerical method used in the numerical experiments in the next section.

---

[4]This is due to the unknown constants in the residual based error estimator, see [14] and [13] for a constant-free error estimator in the sparse setting.

Algorithm 3: A posteriori error control for the Bayesian inversion scheme

---

Require: Prior density π_0, measurements δ, tolerance $\varepsilon = (\varepsilon_{\mathrm{SGFEM}}, \varepsilon_L, \varepsilon_\tau)$, weight factors
   $C = (C_{\mathrm{SGFEM}}, C_L, C_\tau)$, interpolation threshold L, initial step size τ_0

1: `Adapt-SGFEM`$(\varepsilon_{\mathrm{SGFEM}}, C_{\mathrm{SGFEM}}) \to G^{h,N}$
2: `Adapt-`$\Phi_L^{h,N}(\varepsilon_L, C_L, L, \Lambda_M) \to \Phi_L^{h,N}$
3: `Adapt-exp`$_\tau(\varepsilon_\tau, C_\tau, \tau_0) \to \exp_\tau(-\Phi_L^{h,N})$
4: Evaluate normalization constant $Z_{L,\tau}^{h,N}$ (4.34)
5: **return** $\frac{d\pi_{\delta,L,\tau}^{h,N}}{d\pi_0}$ (4.37)

---

6. Numerical Experiments

In the numerical experiments of this section, the derived approach for the numerical computation of the posterior in tensor format is verified by looking at the error of the Bayesian potential and the normalization constant. The employed adaptive forward solver is described in detail in [14]. For the error computations, Monte Carlo sampling is carried out.

We consider the parametric stationary diffusion problem (2.16) on the unit square with deterministic forcing term $f \equiv 1$. The unknown coefficient admits an affine-parametric presentation analog to [14, Sec. 7.2],

$$u(x, y) = 2 + \sum_{m=1}^{M} \psi_m(x) y_m. \tag{6.1}$$

Here,

$$\psi_m(x) = \alpha_m \cos(2\pi \varrho_1(m) x_1) \cos(2\pi \varrho_2(m) x_2), \tag{6.2}$$

where α_m is of the form $\bar{\alpha} m^{-\sigma}$ for some $0 < \bar{\alpha} < \frac{1}{\zeta(\sigma)}$ with $\sigma > 0$ and ζ the Riemann zeta function. Moreover, we set

$$\varrho_1(m) := \frac{m - k(m)(k(m) + 1)}{2} \quad \text{and} \quad \varrho_2(m) := k(m) - \varrho_1(m), \tag{6.3}$$

with $k(m) = \lfloor -\frac{1}{2} + \sqrt{\frac{1}{4} + 2m} \rfloor$. This selection corresponds to an iteration of the planar Fourier sine modes in increasing total order. The random variables are assumed to be uniformly distributed $y_m \sim \mathcal{U}[-1, 1]$.

The goal is to estimate the distribution of u, resp. of the parameter y, from knowledge of noisy measurement data

$$\delta = (\mathcal{O} \circ G)(y) + \eta. \tag{6.4}$$

η is assumed to be Gaussian $\eta \sim \mathcal{N}(0, \Gamma)$ with covariance $\Gamma = \gamma I$ and $\gamma < 1$.

6.1. **Adaptive stochastic Galerkin FEM.** To get a good approximation of the parametric forward solution $G(y)$ to be used for the inverse problem, we employ the ASGFEM of [14] with conforming finite element spaces of order $p = 1, 2$. Based on computable a posteriori error estimators for all approximation components, the adaptive algorithm steers the mesh width h, the stochastic truncation parameter M, the anisotropic polynomial approximation order of the stochastic space, and the tensor rank $\mathbf{r} = (r_0, \dots, r_M)$ of the tensor representation. These discretization parameters are chosen such that the error contributions stay equilibrated. A suitable stopping criteria is given by the amount of information stored in the tensor object, represented by the compressed amount of degrees of freedom. For the system response (3.20) $G^{h,N}(x, y) = U[x, y]$ with coefficient tensor $\tilde{U} \in \mathbb{R}^{N_\mathcal{X} \times n_0 \times \dots \times n_M}$, the number of compressed degrees of freedom *TT-dofs* is defined by

$$\mathrm{ndofs} := \sum_{m=0}^{M-1} r_m n_m r_{m+1} - r_{m+1}^2 + r_M n_M. \tag{6.5}$$
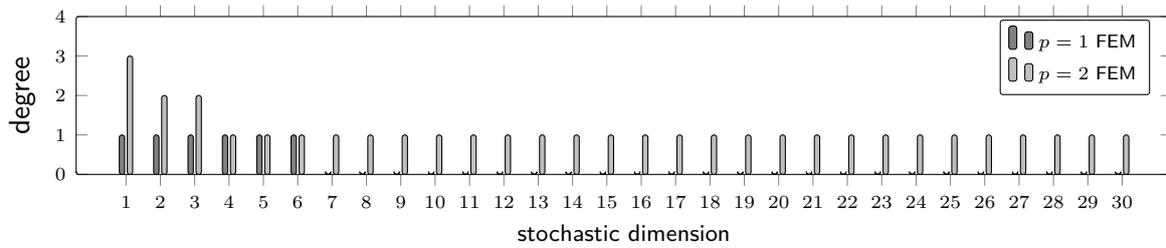
FIGURE 2. Resulting degrees of polynomials in the solution representation (3.20) after reaching the threshold ndofs $= 1.5 \cdot 10^5$ for finite elements of order $p = 1, 2$. Pictured is the polynomial degree for the stochastic dimensions.
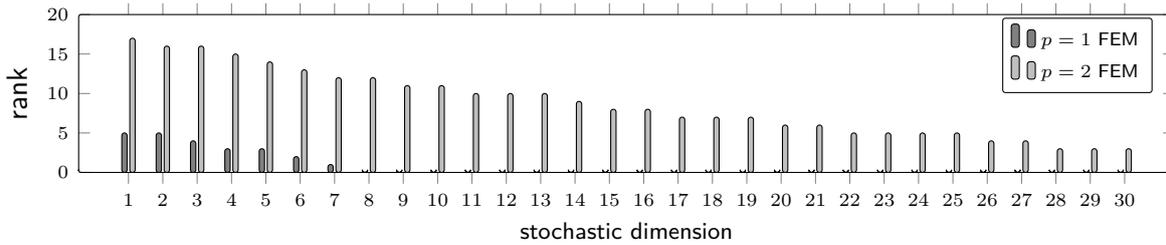
FIGURE 3. Resulting tensor ranks in the solution representation (3.20) after reaching the threshold ndofs $= 1.5 \cdot 10^5$ for finite elements of order $p = 1, 2$. Pictured is the tensor rank for each stochastic dimension.
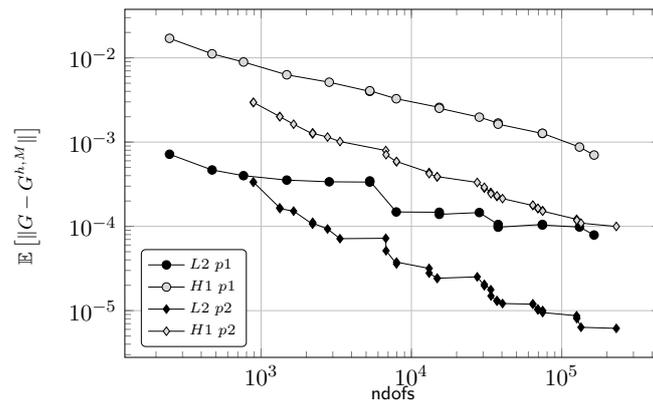
FIGURE 4. Sampled expectation of the mean square error of the stationary diffusion model problem in the $L^2(D)$ and $H_0^1(D)$ norms with FE approximations of degrees $p = 1, 2$ versus number of TT degrees of freedom (ndofs).

This quantity represents the amount of "true" (compressed) degrees of freedom in the tensor. Note that the first tensor rank r_0 for the ASGFEM solution (3.20) is nonzero. A detailed investigation of the performance of the adaptive algorithm can be found in [14] and, in a (non-tensor) sparse basis setting, in [11, 12].

The refinement process is stopped when a prescribed number of TT-dofs is reached. For the target threshold ndofs $= 1.5 \cdot 10^5$, the polynomial degrees for the finest solution can be seen in Figure 2. As one would expect from the experiments in [11, 12, 13, 14], the physical approximation error with a higher-order FE discretization is small even on a relatively coarse grid. Hence, the adaptive algorithm refines the stochastic space more often than with the lowest order FE method, which results in higher polynomial degrees and more active dimensions in the stochastic discretization. For $p = 2$, $M = 43$ active dimensions can be observed, where all dimensions $m \geq 4$ only consist of linear polynomials. The rank development in Figure 3 reflects the importance of the first dimensions.
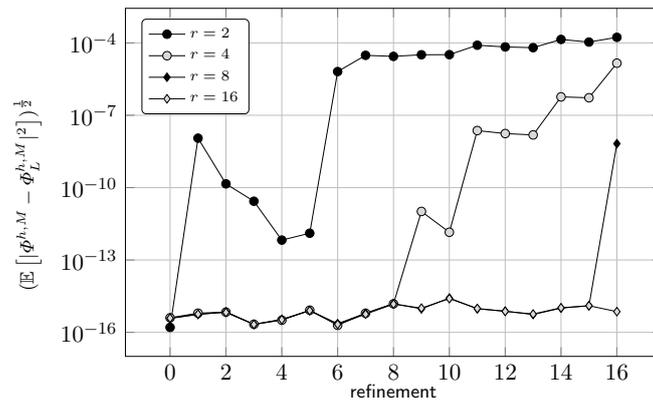
FIGURE 5. Sampled mean square error of the Bayesian potential using $p = 1$ FE approximations w.r.t. number of refinements in the ASGFEM.
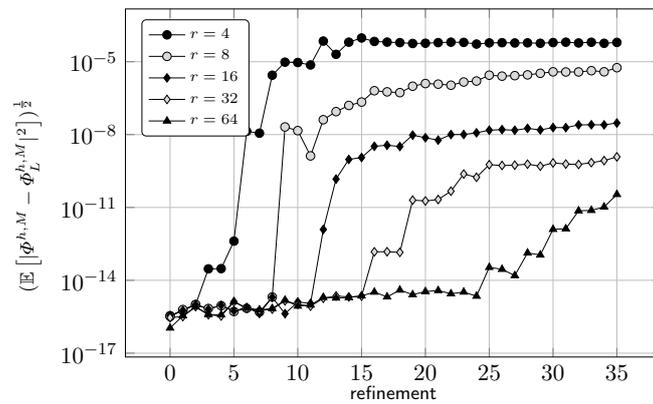
FIGURE 6. Sampled mean square error of the Bayesian potential using $p = 2$ FE approximations w.r.t. number of refinements in the ASGFEM.

The accuracy of the ASGFEM solution approximation $G^{h,N}$ in the $L^2(\Xi_M; L^2(D))$ and $L^2(\Xi_M; H_0^1(D))$ norms is depicted in Figure 4. The mean square error is determined by Monte Carlo sampling (with $N = 200$ samples) of the expectation with respect to a fine reference mesh. In accordance with the theory, it can be seen that the $p = 2$ FE discretization performs better than $p = 1$, which was already observed e.g. in [11, 14]. In particular, the error with the higher-order FEM is about one order of magnitude smaller than with the lowest-order FEM.

6.2. **Bayesian Potential Approximation.** For the proposed Bayesian method, the TT approximation of the forward model $G^{h,N}$ examined in the preceding section is used. The measurements $\delta \in \mathbb{R}^K$ are observed at $K = 9$ equidistantly selected nodes in D from the finest solution with ndofs $= 1.5 \cdot 10^5$, perturbed by uncorrelated uniform Gaussian noise with covariance $\Gamma = \gamma I_K$, $\gamma = 10^{-2}$.

As a precedent step for the Bayesian posterior measure, we examine the approximation of the Bayesian potential (2.12). This inner product is represented in the interpolation basis (4.5) with node parameter L chosen adaptively in every dimension, such that the index set Λ_L yields an exact interpolation. Obviously, the accuracy of this quantity depends on the quality of the forward model approximation and the accuracy of the low-rank tensor representation as described in (4.5). To illustrate the rank dependence, we vary the maximal TT rank and examine the resulting sampled mean square error of the Bayes potential in Figure 5 with respect to the refinement step of the ASGFEM solution. The corresponding TT-dofs can be deduced from (6.5) and become apparent in Figure 4.
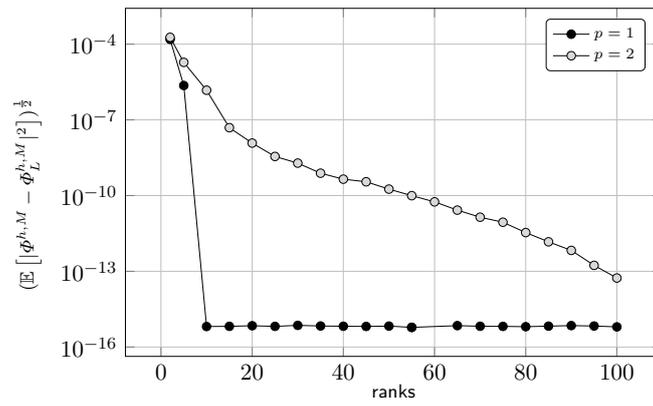
FIGURE 7. Sampled mean square error of the Bayesian potential for the finest refinement ($p = 1$ refinement 16, $p = 2$ refinement 35) versus different ranks used in the Bayesian potential calculation.
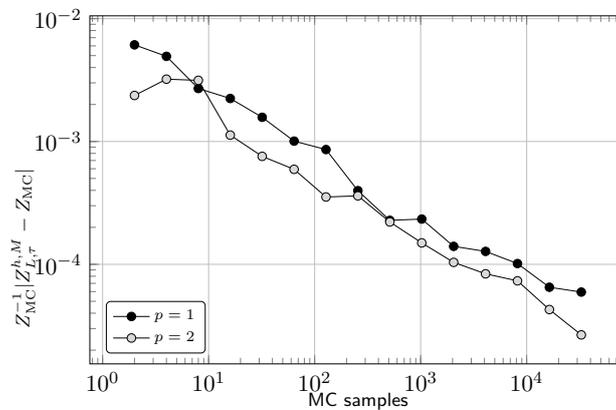
FIGURE 8. Sampled error of the approximated normalization constant and Monte Carlo sampling results for varying numbers of samples. The used $Z_{L,\tau}^{h,N}$ results from either the fines $p = 1$ forward solution approximation or the $p = 2$ forward solution approximation of the same error magnitude as visible in Figure 4.

As one would expect, the approximation is already accurate for small coefficient tensors. The effect of thresholding or tensor rounding approximation becomes dominant for increasing tensor sizes, i.e. for enlarged number of TT-dofs. The moment of the first tensor rounding is clearly visible and even more, one can see the dependence of the dimension (3.11). Rounding the $p = 1$ solution with only 6 dimensions to some small prescribed rank yields a more desirable error order than rounding the $p = 2$ tensor with 43 dimensions.

In Figure7 the convergence of the last refinement of the Galerkin solution with respect to various ranks used in the computation of the Bayesian potential is depicted. For linear finite elements ($p = 1$), the approximation is exact for small ranks due to the small tensors created in the forward calculation. Contrary to this, finite elements with $p = 2$ needs more ranks to display the amount of information contained in the solution tensor.

6.3. **Normalization Constant.**
Another important and in many numerical approaches difficult to obtain quantity is the normalization constant

$$Z = \int_{[-1,1]^M} e^{-\frac{1}{2\gamma}\|\delta-(\mathcal{O}\circ G)(y)\|^2} \, \mathrm{d}\pi_0(y). \qquad (6.6)$$

In the proposed method, the approximation $Z_{L,\tau}^{h,N}$ can be computed efficiently as in (4.34).

A Monte-Carlo integration with increasing sample sets $(y_i)_{i=1}^{N_{\mathrm{MC}}}$, $N_{\mathrm{MC}} \in \{2^i \mid i = 1, \ldots, 17\}$ and y_i drawn from $\mathcal{U}([-1,1]^M)$ is applied to (6.6), i.e.

$$Z_{\mathrm{MC}} := \frac{1}{N_{\mathrm{MC}}} \sum_{i=1}^{N_{\mathrm{MC}}} e^{-\frac{1}{2\gamma} \|\delta - (\mathcal{O} \circ G)(y_i)\|^2}. \tag{6.7}$$

In Figure 8 the relative error of the Monte-Carlo approximation Z_{MC} to $Z_{L,\tau}^{h,N}$ can be observed. The data is obtained from averaging 10 runs.

References

[1] Markus Bachmayr, Reinhold Schneider, and André Uschmajew. Tensor Networks and Hierarchical Tensors for the Solution of High-Dimensional Partial Differential Equations. *Found. Comput. Math.*, 16(6):1423–1472, 2016.

[2] Alex Bespalov, Catherine E Powell, and David Silvester. Energy norm a posteriori error estimation for parametric operator equations. *SIAM Journal on Scientific Computing*, 36(2):A339–A363, 2014.

[3] J. C. Butcher. *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, Ltd., Chichester, third edition, 2016. With a foreword by J. M. Sanz-Serna.

[4] Peng Chen and Christoph Schwab. *Adaptive Sparse Grid Model Order Reduction for Fast Bayesian Estimation and Inversion*, pages 1–27. Springer International Publishing, Cham, 2016.

[5] Abdellah Chkifa, Albert Cohen, and Christoph Schwab. Breaking the curse of dimensionality in sparse polynomial approximation of parametric {PDEs}. *Journal de Mathématiques Pures et Appliquées*, 103(2):400 – 428, 2015.

[6] Claudia Prévot, Michael Röckner. *A Concise Course on Stochastic Partial Differential Equations*. 2007.

[7] M. Dashti and A. M. Stuart. The Bayesian Approach to Inverse Problems. *ArXiv e-prints*, February 2013.

[8] J. Dick, R. N. Gantner, Q. T. Le Gia, and C. Schwab. Higher order Quasi-Monte Carlo integration for Bayesian Estimation. *ArXiv e-prints*, February 2016.

[9] Josef Dick, Frances Y. Kuo, Quoc T. Le Gia, and Christoph Schwab. Multilevel higher order QMC Petrov-Galerkin discretization for affine parametric operator equations. *SIAM J. Numer. Anal.*, 54(4):2541–2568, 2016.

[10] Sergey V Dolgov and Dmitry V Savostyanov. Alternating minimal energy methods for linear systems in higher dimensions. *SIAM Journal on Scientific Computing*, 36(5):A2248–A2271, 2014.

[11] Martin Eigel, Claude Jeffrey Gittelson, Christoph Schwab, and Elmar Zander. Adaptive stochastic Galerkin FEM. *Comput. Methods Appl. Mech. Engrg.*, 270:247–269, 2014.

[12] Martin Eigel, Claude Jeffrey Gittelson, Christoph Schwab, and Elmar Zander. A convergent adaptive stochastic Galerkin finite element method with quasi-optimal spatial meshes. *ESAIM: Mathematical Modelling and Numerical Analysis*, 49(5):1367–1398, 2015.

[13] Martin Eigel and Christian Merdon. Local equilibration error estimators for guaranteed error control in adaptive stochastic higher-order Galerkin finite element methods. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):1372–1397, 2016.

[14] Martin Eigel, Max Pfeffer, and Reinhold Schneider. Adaptive stochastic Galerkin FEM with hierarchical tensor representations. *Numerische Mathematik*, pages 1–39, 2016.

[15] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 6*, 2(11):559–572, 1901.

[16] Roger G. Ghanem and Pol D. Spanos. *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York, 1991.

[17] Viet A. Ha-Hoang, Christoph Schwab, and Andrew M. Stuart. Sparse MCMC gpc Finite Element Methods for Bayesian Inverse Problems. Technical report, Zürich, August 2012.

[18] Wolfgang Hackbusch. *Tensor Spaces and Numerical Tensor Calculus*. 2012.

[19] Wolfgang Hackbusch and Reinhold Schneider. *Tensor spaces and hierarchical tensor representations*, pages 237–261. Springer International Publishing, Cham, 2014.

[20] Wolfgang Hackbusch and Reinhold Schneider. Tensor spaces and hierarchical tensor representations. In *Extraction of Quantifiable Information from Complex Systems*, pages 237–261. Springer, 2014.

[21] E. Hairer, S. P. Nø rsett, and G. Wanner. *Solving Ordinary Differential Equations. I*, volume 8 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin, second edition, 1993. Nonstiff problems.

[22] Viet Ha Hoang and Christoph Schwab. Convergence rate analysis of MCMC-FEM for Bayesian inversion of log-normal diffusion problems. Technical Report 2016-19, Seminar for Applied Mathematics, ETH Zürich, 2016.

[23] Sebastian Holtz, Thorsten Rohwedder, and Reinhold Schneider. The Alternating Linear Scheme for tensor optimization in the Tensor Train format. *SIAM J. Sci. Comput.*, 34(2):A683–A713, 2012.

[24] Michel Loève. *Probability Theory. I*. Springer-Verlag, New York-Heidelberg, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.

[25] Gabriel J. Lord, Catherine E. Powell, and Tony Shardlow. *An Introduction to Computational Stochastic PDEs*. Cambridge Texts in Applied Mathematics. Cambridge University Press, New York, 2014.

[26] Christian Lubich, Thorsten Rohwedder, Reinhold Schneider, and Bart Vandereycken. Dynamical Approximation by Hierarchical Tucker and Tensor-Train Tensors. *SIAM J. Matrix Anal. Appl.*, 34(2):470–494, 2013.

[27] Cleve Moler and Charles Van Loan. Nineteen Dubious Ways to Compute the Exponential of a Matrix, Twenty-Five Years Later. *SIAM Review*, 45(1):3–49, 2003.

[28] I. V. Oseledets. Tensor-Train Decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011.

[29] M. Parno and Y. Marzouk. Transport map accelerated Markov chain Monte Carlo. *ArXiv e-prints*, December 2014.

[30] Claudia Schillings and Christoph Schwab. Scaling Limits in Computational Bayesian Inversion. *ESAIM: M2AN*, 50(6):1825–1856, 2016.

[31] Reinhold Schneider and André Uschmajew. Convergence results for projected line-search methods on varieties of low-rank matrices via Łojasiewicz inequality. *SIAM J. Optim.*, 25(1):622–646, 2015.

[32] C Schwab and A M Stuart. Sparse deterministic approximation of Bayesian inverse problems. *Inverse Problems*, 28(4):045003, 2012.

[33] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numer.*, 20:291–467, 2011.

[34] Christoph Schwab and Claude Jeffrey Gittelson. Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, 20:291–467, 005 2011.

[35] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 005 2010.

[36] A. M. Stuart and A. L. Teckentrup. Posterior Consistency for Gaussian Process Approximations of Bayesian Posterior Distributions. *ArXiv e-prints*, March 2016.