

MULTILEVEL ADAPTIVE SPARSE LEJA APPROXIMATIONS FOR BAYESIAN INVERSE PROBLEMS *

I-G. FARCAŞ[†], J. LATZ[‡], E. ULLMANN[‡], T. NECKEL[†], AND H.-J. BUNGARTZ[†]

Abstract. Deterministic interpolation and quadrature methods are often unsuitable to address Bayesian inverse problems depending on computationally expensive forward mathematical models. While interpolation may give precise posterior approximations, deterministic quadrature is usually unable to efficiently investigate an informative and thus concentrated likelihood. This leads to a large number of required expensive evaluations of the mathematical model. To overcome these challenges, we formulate and test a multilevel adaptive sparse Leja algorithm. At each level, adaptive sparse grid interpolation and quadrature are used to approximate the posterior and perform all quadrature operations, respectively. Specifically, our algorithm uses coarse discretizations of the underlying mathematical model to investigate the parameter space and to identify areas of high posterior probability. Adaptive sparse grid algorithms are then used to place points in these areas, and ignore other areas of small posterior probability. The points are weighted Leja points. As the model discretization is coarse, the construction of the sparse grid is computationally efficient. On this sparse grid, the posterior measure can be approximated accurately with few expensive, fine model discretizations. The efficiency of the algorithm can be enhanced further by exploiting more than two discretization levels. We apply the proposed multilevel adaptive sparse Leja algorithm in numerical experiments involving elliptic inverse problems in 2D and 3D space, in which we compare it with Markov chain Monte Carlo sampling and a standard multilevel approximation.

Key words. Bayesian inference, multilevel method, adaptive sparse grids, partial differential equation

AMS subject classifications. 35R30, 62F15, 65C60, 65D30, 65N30, 68T05

1. Introduction. Mathematical models in science and engineering often require input parameters which cannot be observed directly, yet these parameters are required for predictions based on the model. A standard procedure is to estimate the inputs from indirect observations, which is known as an *inverse problem*. In contrast, the corresponding *forward problem* maps from the parameter to the observation space.

In many applications, for instance the geosciences or medical sciences, the observations are noisy and their number is insufficient to identify a unique associated parameter value. The Bayesian approach to inverse problems [21, 41, 43] provides a consistent mechanism to combine noisy or incomplete data with prior knowledge, and to quantify the uncertainty in the parameter estimate. The prior knowledge is incorporated into a probability distribution over the parameter space; this is termed *prior (measure)* μ_0 . The Bayesian solution to the inverse problem is then the *posterior (measure)* μ^y arising from conditioning the prior μ_0 on the observations. Unfortunately, the posterior is often intractable in the sense that it does not admit closed form analytic expressions. Hence approximations have to be used in practice.

Sampling-based posterior approximations such as Markov chain Monte Carlo (MCMC) [1] or Sequential Monte Carlo (SMC) [7] do not rely on the smoothness of the parameter-to-observation map, and can be conducted in high-dimensional param-

*

Funding: This work was supported by the German Research Foundation (DFG) through the TUM International Graduate School of Science and Engineering (IGSSE) within Project 10.02 BAYES, and by the EPSRC under grant number EP/R014604/1.

[†]Department of Informatics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany (farcasi@in.tum.de, neckel@in.tum.de, bungartz@in.tum.de).

[‡]Department of Mathematics, Technical University of Munich, Boltzmannstr. 3, 85748 Garching, Germany (jonas.latz@ma.tum.de, elisabeth.ullmann@ma.tum.de).

eter spaces. The drawback is that without exploiting smoothness or low-dimensional structure in the parameter space often a prohibitive number of samples are required to obtain the desired accuracy. Since each sample entails the evaluation of the forward map, the total cost of the Bayesian inversion becomes prohibitive if the forward model is specified by a partial differential equation (PDE), and is thus computationally expensive. In recent years, many works suggested computationally feasible, yet accurate approximations of the posterior. In this work we focus on *sampling-free* approximations for computationally expensive problems for parameter spaces of small to moderate dimension. We assume that the prior and posterior have a probability density function with respect to (w.r.t.) the Lebesgue measure and approximate the posterior density using sparse grid interpolation and sparse grid quadrature.

Sampling-free approaches often involve *surrogates* for the forward response operator to decrease the computational cost. Typical surrogates are based on generalized polynomial chaos [10, 26, 30, 47], sparse grids [3, 5, 28, 36, 37, 38], Gaussian process regression [22, 42], model reduction [13, 24, 27], and combinations, e.g., sparse grids and reduced bases [3, 4]. To obtain an accurate (and convergent) approximation, these surrogates require certain types of smoothness of the response surface or likelihood function w.r.t. the input parameters. The smoothness assumptions can be weakened by using piecewise polynomial approximations together with Voronoi tessellations of the parameter space [31]. Of course, surrogates can also be used to accelerate sampling-based approximations such as MCMC, see e.g., [26, 33]. We remark that Quasi-Monte Carlo [8, 35] is in principle a sampling-free method which does not rely on surrogates, however, it requires again a smooth approximand, and is often used together with randomization.

Theoretical analysis shows that if the surrogate converges to the forward model at a specific rate w.r.t. the prior weighted L^2 -norm, then the approximate posterior converges to the exact posterior with at least the same rate [30, 41]. This result has been improved recently in [46] where it was showed that the convergence rate of the posterior approximation is at least twice as large as the convergence rate of the surrogate, for general priors. However, constructing an accurate surrogate over the entire support of the prior might not be feasible and is in fact often unnecessary. Indeed, in inference problems where the data is informative, the posterior differs significantly from the prior, and is supported only in a small part of the prior support. This suggests to adapt and localize the surrogate construction to the support of the posterior. We adopt this approach in our work and construct multilevel, adaptive surrogates with localized support using adaptive sparse grid approximations.

The idea of posterior-focused surrogates is not new, it has however received little attention to date in the literature. Li and Marzouk [26] borrow ideas from statistics and construct an efficient polynomial chaos surrogate associated with a density that minimizes the cross entropy between the posterior and a family of multivariate normal distributions. Jiang and Ou [20] suggest a two-stage surrogate based on generalized multiscale finite elements and least-squares stochastic collocation. Yan and Zhou [47] propose a multifidelity polynomial chaos surrogate which combines a large number of inexpensive low-fidelity model evaluations with a small number of expensive high-fidelity model evaluations, following the idea of multifidelity approximations [34].

One challenge of posterior-focused surrogates is the need to handle arbitrary densities which can deviate significantly from the prior which is usually a classical density such as uniform or Gaussian. We address this by constructing adaptive sparse grid approximations based on *weighted (L)-Leja sequences* (see, e.g., [16, 32]). Note that sparse grid approximations with Leja points have been devised for forward uncer-

tainty propagation in [11, 12, 14, 32]. In [11] the adaptive construction of the points is guided by sensitivity scores, and the strategy is applied in plasma microturbulence analysis. In [14] the Leja points are constructed with the help of an adjoint-based error indicator. The use of Leja points to approximate posterior densities is a novel contribution to the literature. Leja points offer further computational advantages since they are *nested* and thus allow to reuse (expensive) model evaluations.

We address the possible high-dimensional parameter space in Bayesian inversion by the use of multilevel approximations. At each level, *dimension-adaptive* sparse grids [6, 15, 32] are employed, either in standard form, or using directional variances to better exploit anisotropies in the parameter space. In particular, at the first level, our algorithm uses a coarse discretization of the given model to investigate the parameter space and to identify areas of high posterior probability. Adaptive algorithms are then used to place weighted (L)-Leja points in these areas, and ignore other areas of small posterior probability. Starting with the second level, we sequentially update the prior such that the previous sparse grid approximation of the posterior is reused. In this way, the current posterior measure can be approximated accurately with few expensive, finer model discretizations. We point out that sparse grid approximations are based on point sequences in one dimension. However, starting with the second level the posterior densities are in general not separable and hence we cannot rely on a simple tensorization of univariate Leja points. Instead we construct Leja points w.r.t. a *Gaussian approximation* of the posterior, which is separable, and we then correct the bias introduced by this approximation in quadrature computations.

The remainder of this paper is structured as follows. In [section 2](#) we provide the necessary background information. In particular, in [subsection 2.1](#) we formulate the Bayesian inverse problem and discuss the computation of posterior expectations by importance sampling. [Subsection 2.2](#) reviews multilevel approaches, and [subsection 2.3](#) discusses generalized sparse grids. [Section 3](#) contains the major contribution of our work, the multilevel adaptive sparse Leja approximation to the posterior density in a Bayesian inverse problem. This method is independent of the specific implementation of the adaptive sparse grid approximations. Details on the used dimension-adaptive approximations are given in [section 4](#). In [section 5](#), we present numerical results, comparing our multilevel algorithm with sampling methods and the classical multilevel approach based on telescoping sums. Finally, [section 6](#) offers concluding remarks.

2. Background.

2.1. Bayesian inversion. To begin we formulate the Bayesian inverse problem. Let $\mathbf{X} \subset \mathbb{R}^{N_{\text{sto}}}$ denote the parameter space. In addition, let $\mathbf{Y} = \mathbb{R}^{N_{\text{obs}}}$ be a separable Banach space that denotes the data space. N_{sto} is the dimension of the data space, and N_{obs} is the number of observations. Notably, the parameter and data space are *finite-dimensional*. This allows us to work with densities w.r.t. the Lebesgue measure. The underlying mathematical model is formalized by a function $\mathcal{G} : \mathbf{X} \rightarrow \mathbf{Y}$, which maps from the parameter space to the data space. Noisy observations $\mathbf{y} \in \mathbf{Y}$ are obtained. To model the noise we assume that \mathbf{y} is a realisation of the random variable $\mathcal{G}(\boldsymbol{\theta}^{\text{true}}) + \boldsymbol{\eta}$ where $\boldsymbol{\eta} \sim \text{N}(0, \Gamma)$ is non-degenerate Gaussian noise, and $\boldsymbol{\theta}^{\text{true}} \in \mathbf{X}$ is the true parameter. In an inverse problem we wish to identify the parameter $\boldsymbol{\theta}^{\text{true}}$, i.e., solve the equation

$$(2.1) \quad \mathcal{G}(\boldsymbol{\theta}) + \boldsymbol{\eta} = \mathbf{y}$$

for $\boldsymbol{\theta}$. This problem is typically ill-posed in the sense of Hadamard [18], due to noise, and since the low-dimensional data space is often not sufficiently rich to allow

the identification of a unique parameter in the high-dimensional space \mathbf{X} . The ill-posedness can be cured by reformulating (2.1) as a Bayesian inverse problem. Next, we introduce our notation and briefly discuss Bayesian inversion, and refer to [41] for more details.

We assume that $\boldsymbol{\theta}$ is an \mathbf{X} -valued random variable that is distributed according to a *prior measure* μ_0 with Lebesgue density π_0 on the parameter space \mathbf{X} . Moreover, we assume that $\boldsymbol{\theta}$ is stochastically independent of the noise $\boldsymbol{\eta}$. The density π_0 reflects our knowledge about $\boldsymbol{\theta}$ before we make an observation \mathbf{y} . The information provided by \mathbf{y} is modelled by the (data) likelihood. Since the noise $\boldsymbol{\eta}$ is Gaussian by assumption, the likelihood is given by

$$(2.2) \quad \begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}) &:= \exp(-\Phi(\boldsymbol{\theta}; \mathbf{y})), \\ \Phi(\boldsymbol{\theta}; \mathbf{y}) &:= \frac{1}{2} \|\Gamma^{-1/2}(\mathbf{y} - \mathcal{G}(\boldsymbol{\theta}))\|^2. \end{aligned}$$

The function Φ is called potential or negative log-likelihood. The solution of the Bayesian inverse problem is the *posterior (measure)* $\mu^{\mathbf{y}}$, i.e., the conditional measure of $\boldsymbol{\theta}$ given that the event $\{\mathcal{G}(\boldsymbol{\theta}) + \boldsymbol{\eta} = \mathbf{y}\}$ occurred. The posterior measure $\mu^{\mathbf{y}}$ has also a density $\pi^{\mathbf{y}}$ which can be computed using Bayes's formula:

$$(2.3) \quad \begin{aligned} \pi^{\mathbf{y}}(\boldsymbol{\theta}) &= \frac{L(\boldsymbol{\theta}|\mathbf{y})\pi_0(\boldsymbol{\theta})}{Z(\mathbf{y})}, \quad \boldsymbol{\theta} \in X, \mathbf{y} \in Y, \\ Z(\mathbf{y}) &= \int_{\mathbf{X}} L(\boldsymbol{\theta}|\mathbf{y})\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}, \end{aligned}$$

provided that $0 < Z(\mathbf{y}) < \infty$. In the given setting (non-degenerate Gaussian additive noise, finite dimensional data space), one can show that $Z(\mathbf{y})$ is always finite and bounded away from 0. This implies existence of the posterior measures, see [23]. The work [23] also establishes that Bayesian inverse problems of this type are always *well-posed*.

Finally, let $g : \mathbf{X} \rightarrow \mathbb{R}$ denote a quantity of interest (QoI) depending on the parameter $\boldsymbol{\theta}$. Since $\boldsymbol{\theta}$ is a random variable, one is typically interested in the forward propagation of uncertainties through the action of g . For instance, we wish to evaluate integrals of g w.r.t. the posterior measure:

$$(2.4) \quad \mathbb{E}_{\mu^{\mathbf{y}}}[g] = \int_{\mathbf{X}} g(\boldsymbol{\theta})\pi^{\mathbf{y}}(\boldsymbol{\theta})d\boldsymbol{\theta}.$$

In practice, we approximate integrals of this type via numerical quadrature. Note that we can write the expected value in (2.4) in terms of a ratio of two expected values w.r.t. the prior measure:

$$(2.5) \quad \mathbb{E}_{\mu^{\mathbf{y}}}[g] = \frac{1}{Z(\mathbf{y})} \int g(\boldsymbol{\theta})L(\boldsymbol{\theta}|\mathbf{y})\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta} = \frac{\mathbb{E}_{\mu_0}[g(\cdot)L(\cdot|\mathbf{y})]}{\mathbb{E}_{\mu_0}[L(\cdot|\mathbf{y})]}.$$

Note that the prior is typically much more accessible compared to the posterior, via i.i.d. sampling or a closed form probability density function (pdf). If the two expected values in (2.5) are approximated with samples from μ_0 , we refer to this method by *importance sampling*. If the expected values are approximated with other numerical quadrature rules, e.g., Quasi-Monte Carlo [35] or sparse grid quadrature [36, 38], we refer to any of these methods as *importance-sampling-based methods*.

2.2. Multilevel approximation of the quantity of interest. The approximation of the expected value $\mathbb{E}_{\mu^{\mathbf{y}}}[g]$ involves two sources of error: (i) the quadrature error associated with the approximation of the measure $\mu^{\mathbf{y}}$, and (ii) the discretization error associated with the approximation of g . The quadrature error is typically controlled by the number of particles or grid points in the parameter domain. The discretization error in g is often controlled by the mesh size in the physical domain. In our setting, the evaluation of g typically involves a computationally expensive model, e.g., a partial differential equation (PDE). If we wish to construct accurate approximations to $\mathbb{E}_{\mu^{\mathbf{y}}}[g]$, then g has to be discretized with a high resolution on a fine grid in physical space, and g has to be evaluated for a large number of parameter values. For these reasons the approximation of $\mathbb{E}_{\mu^{\mathbf{y}}}[g]$ is computationally demanding.

Multilevel methods provide a framework to approximate high-resolution problems efficiently by combining evaluations from fine and coarse grid approximations. Specifically, approximations based on quadrature and physical domain discretization grids of complementary resolution are combined such that most computations are done on coarse grids while fine grid approximations are evaluated only a limited number of times. In this way, the overall computational cost is reduced, while the accuracy is preserved. A large number of multilevel approaches for Bayesian inverse problems proceed by using a telescoping sum based on the linearity of the expectation operator, see, e.g., [9]. Alternatively, it is possible to construct a multilevel approximation without relying on such a telescoping sum. Examples are the multifidelity preconditioned MCMC method in [33], Multilevel Sequential² Monte Carlo [25], and our multilevel sparse Leja approximation presented in [section 3](#).

2.3. Approximation with generalized sparse grids. We aim to approximate posterior density functions with sparse grid interpolation and quadrature at each level in our multilevel approach. To this end, we employ *generalized, adaptive Smolyak approximations* [39]. Smolyak's algorithm, also known as the combination scheme [17], is a strategy to construct multivariate sparse grid approximations by weakening the assumed coupling between the input dimensions (see, e.g., [6, 11]). We briefly summarize Smolyak's algorithm. For a more detailed overview of this approach and its application in scientific computing, see [2, 44] and the references therein.

Let $f^i: X_i \subset \mathbb{R} \rightarrow \mathbb{R}$ denote univariate functions, where $i = 1, 2, \dots, N_{\text{sto}}$. In addition, let $f^{N_{\text{sto}}}: \mathbf{X} \rightarrow \mathbb{R}$ denote a multivariate function with a scalar output. For example, $f^{N_{\text{sto}}}$ could be the potential function, $\Phi(\boldsymbol{\theta}; \mathbf{y})$, or the QoI, $g(\boldsymbol{\theta})$. Let $\mathcal{U}^i[f^i]$ for $i = 1, 2, \dots, N_{\text{sto}}$ denote either univariate interpolation or integration operators defined w.r.t. a weight function $w: X_i \rightarrow \mathbb{R}_+$, which in our context is the i th component of the prior density. Further, consider approximations $\mathcal{U}_k^i[f^i]$ that converge as $k \rightarrow \infty$, where k is typically referred to as *level*. Starting from one-dimensional *difference* or *hierarchical surplus operators*,

$$(2.6) \quad \Delta_k^i[f^i] := \mathcal{U}_k^i[f^i] - \mathcal{U}_{k-1}^i[f^i], \quad i = 1, 2, \dots, N_{\text{sto}},$$

with the convention $\Delta_1^i[f^i] := \mathcal{U}_1^i[f^i]$, Smolyak's approximation formula reads

$$(2.7) \quad \mathcal{U}_{\mathcal{K}}[f^{N_{\text{sto}}}] = \sum_{\mathbf{k} \in \mathcal{K}} (\Delta_{k_1}^1 \otimes \Delta_{k_2}^2 \otimes \dots \otimes \Delta_{k_{N_{\text{sto}}}}^{N_{\text{sto}}})[f^{N_{\text{sto}}}] = \sum_{\mathbf{k} \in \mathcal{K}} \Delta_{\mathbf{k}}[f^{N_{\text{sto}}}],$$

where $\mathbf{k} := (k_1, k_2, \dots, k_{N_{\text{sto}}}) \in \mathbb{N}^{N_{\text{sto}}}$ is a *multiindex* and \mathcal{K} is a *finite set of multi-indices*. Note that by construction, (2.7) requires the underlying multivariate space, \mathbf{X} , as well as the corresponding weight function to be separable. When the weight function is a density the stochastic parameters need to be independent.

Since (2.7) is written in terms of tensorizations of univariate difference operators (2.6) the set \mathcal{K} must be constructed such that the summation in (2.7) telescopes correctly. Suitable sets of multiindices \mathcal{K} are called *admissible* (or *downward closed*); see [15]. In particular, for an admissible set \mathcal{K} it holds that $\mathbf{k} \in \mathcal{K} \Rightarrow \mathbf{k} - \mathbf{e}_i \in \mathcal{K}$ for $i = 1, 2, \dots, N_{\text{sto}}$, where \mathbf{e}_i denotes the i th unit vector in $\mathbb{R}^{N_{\text{sto}}}$.

To construct the approximations $\mathcal{U}_k^i[f^i]$, we employ *weighted (L)-Leja sequences* (see, e.g., [16, 32]). Given the weight function $w : X_i \rightarrow \mathbb{R}_+$, weighted (L)-Leja sequences are constructed recursively as follows:

$$(2.8) \quad \begin{aligned} \theta_1 &= \operatorname{argmax}_{\theta \in X_i} w(\theta), \\ \theta_n &= \operatorname{argmax}_{\theta \in X_i} w(\theta) \prod_{m=1}^{n-1} (\theta - \theta_m), \quad n = 2, 3, \dots \end{aligned}$$

When w is the standard uniform density with support $[0, 1]$, we choose $\theta_1 = 0.5$. Note that the above point sequence is in general not uniquely defined, because (2.8) might have multiple maximizers. In that case we simply pick one of the maximizers. Weighted (L)-Leja sequences allow the construction of sparse grid approximations for arbitrary probability densities. Moreover, they lead to accurate approximations with low cardinality (see, e.g., [32]). To fully define (2.7), we need to specify the multiindex set \mathcal{K} as well. We construct \mathcal{K} adaptively based on the *dimension-adaptive* algorithm of [15, 19], which we outline in section 4.

2.4. Assumptions. As discussed in subsection 2.3, sparse grid approximations require a tensor domain and a tensorized prior measure. Hence, we assume:

A1. The parameter space is a tensor product space, i.e.,

$$\mathbf{X} = \bigotimes_{i=1}^{N_{\text{sto}}} X_i,$$

where $X_i \subset \mathbb{R}$ for $i = 1, \dots, N_{\text{sto}}$.

A2. The prior density π_0 is separable, i.e.,

$$(2.9) \quad \pi_0(\boldsymbol{\theta}) = \prod_{i=1}^{N_{\text{sto}}} \pi_{0,i}(\theta_i),$$

where $\pi_{0,i} : X_i \rightarrow \mathbb{R}$, $i = 1, \dots, N_{\text{sto}}$.

Assumption A1 can always be satisfied by embedding a non-tensorized parameter space into a hyperrectangle of suitable dimension. Assumption A2 is fulfilled if the components of $\boldsymbol{\theta}$ are stochastically independent under the prior measure. If Assumption A2 is not satisfied, the Gaussian approximation of π_0 is needed (see section 3).

3. Multilevel sparse Leja algorithm. This section contains the major contribution of our work. Our goal is to address the challenges of Bayesian inversion in computationally expensive problems. To this end, we formulate a deterministic, multilevel, sampling-free methodology based on sparse grids in which we sequentially update the prior information as the level in the multilevel hierarchy increases.

Most computations in Bayesian inversion involve evaluations of the forward operator, $\mathcal{G}(\boldsymbol{\theta})$, which in this paper is assumed to be computationally expensive. When

a large number of such evaluations is needed, the corresponding computational cost is prohibitive. To reduce the cost, we employ multilevel approximations. At each level, we construct sparse grid surrogates of the potential function $\Phi(\boldsymbol{\theta}; \mathbf{y})$. Our motivation is two-fold. First, evaluating $\Phi(\boldsymbol{\theta}; \mathbf{y})$ means evaluating the forward model, $\mathcal{G}(\boldsymbol{\theta})$ (see (2.2)), hence the computationally expensive part. Second, even if $\mathcal{G}(\boldsymbol{\theta})$ is vector valued, $\Phi(\boldsymbol{\theta}; \mathbf{y})$ is a scalar. Sparse grid approximations can be constructed for vector-valued functions, however, a separate approximation is usually needed for each output component, which can be infeasibly expensive. After the surrogate is obtained, all other single level operations, which typically involve integration, employ this surrogate, making them computationally cheap. In the following, we use the superscripts *in* and *qu* to specifically refer to interpolation and quadrature, respectively, whereas superscript *op* is used to refer to either of the two operations.

In this paper, the surrogates of the potential function are constructed via adaptive sparse grid interpolation, whereas all integration operations are performed using adaptive sparse grid quadrature. The specific implementations do not influence the formulation of our multilevel approach. Thus we assume we have two adaptive strategies, $\text{AdaptSGInterp}(tol^{\text{in}}, K_{\max}^{\text{in}}, g, \pi)$ and $\text{AdaptSGQuad}(tol^{\text{qu}}, K_{\max}^{\text{qu}}, g, \pi)$, which depend on a *tolerance*, tol^{op} . The other inputs are a maximum reachable sparse grid level, K_{\max} , the target function, g , and the density function w.r.t. which the approximation is performed, π . Note that specific implementations might have additional input arguments, however the four inputs considered here are sufficient to illustrate these algorithms. The adaptive strategies are summarized in [section 4](#).

3.1. General setup. Let $J > 1$, $J \in \mathbb{N}$, denote the number of *levels* in our multilevel formulation. Further, let $A \in \{\mathcal{G}, \Phi, L, Z, \pi^{\mathbf{y}}\}$ denote a generic quantity depending on both physical and stochastic parameters. Let $j = 1, 2, \dots, J$. By h_j we characterize the discretization of the physical domain of the forward response operator $\mathcal{G}(\boldsymbol{\theta})$, where h_1 is the coarsest and h_J the finest discretization level. Hence, by A_j we denote the semi-discrete approximation of A depending on h_j , whereas A_{δ_j} denotes either $A_j - A_{j-1}$ or A_j/A_{j-1} . In addition, tol_j^{op} denotes the tolerance employed in the adaptive sparse grid approximations of A_j such that

$$tol_1^{\text{op}} \leq tol_2^{\text{op}} \leq \dots \leq tol_J^{\text{op}}.$$

Thus by $A_{j,s}$ we denote the sparse grid approximation of A_j depending on tol_s^{op} .

In our multilevel formulation, we determine $A_{j,J-j+1}$ for all $j = 1, 2, \dots, J$. To simplify the notation we use the subscript $\ell(j)$ to refer to $(h_j, tol_{j-j+1}^{\text{op}})$ and the subscript $\ell(\delta j)$ to denote approximations $A_{\ell(\delta j)} \approx A_{\delta j}$. Hence we refer to the *level* j in our multilevel approach by $\ell(j)$ or $\ell(\delta j)$. Note that levels are used to characterize both sparse grid and multilevel formulations. To avoid confusion, we will explicitly specify what is meant by level in each context.

3.2. Level $\ell(1)$. At $\ell(1)$ we compute the approximation $\Phi_{\ell(1)}(\boldsymbol{\theta}; \mathbf{y}) \approx \Phi_1(\boldsymbol{\theta}; \mathbf{y})$ using adaptive sparse grid interpolation w.r.t. the prior π_0 . This is possible since π_0 has a product structure by Assumption A2, see (2.9). The potential's approximation $\Phi_{\ell(1)}(\boldsymbol{\theta}; \mathbf{y})$ is used for $L_1(\boldsymbol{\theta}|\mathbf{y}) \approx L_{\ell(1)}(\boldsymbol{\theta}|\mathbf{y}) := \exp(-\Phi_{\ell(1)}(\boldsymbol{\theta}; \mathbf{y}))$, the level one likelihood surrogate. Afterwards, we employ $L_{\ell(1)}(\boldsymbol{\theta}|\mathbf{y})$ to compute the evidence $Z_{\ell(1)}(\mathbf{y}) \approx Z_1(\mathbf{y})$ via adaptive sparse grid quadrature w.r.t. the prior density π_0 . Having computed $L_{\ell(1)}(\boldsymbol{\theta}|\mathbf{y})$ and $Z_{\ell(1)}(\mathbf{y})$ we apply formula (2.3) and obtain the posterior at level $\ell(1)$, $\pi_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta})$. In addition, we also compute the posterior expectation,

$\mathbf{m}_{\ell(1)} \in \mathbb{R}^{N_{\text{sto}}}$, where for $i = 1, \dots, N_{\text{sto}}$,

$$\mathbf{m}_{\ell(1)}^i := Z_{\ell(1)}^{-1}(\mathbf{y}) \int_{\mathbf{X}} \theta_i L_{\ell(1)}(\boldsymbol{\theta}|\mathbf{y}) \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

and covariance matrix, $\mathbf{C}_{\ell(1)} \in \mathbb{R}^{N_{\text{sto}} \times N_{\text{sto}}}$, where for $n, p = 1, \dots, N_{\text{sto}}$,

$$\mathbf{C}_{\ell(1)}^{np} := Z_{\ell(1)}^{-1}(\mathbf{y}) \int_{\mathbf{X}} \theta_n \theta_p L_{\ell(1)}(\boldsymbol{\theta}|\mathbf{y}) \pi_0(\boldsymbol{\theta}) d\boldsymbol{\theta} - \mathbf{m}_{\ell(1)}^n \mathbf{m}_{\ell(1)}^p,$$

which we need to construct the *Gaussian approximation* $\hat{\pi}_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta}) := \mathbf{N}(\mathbf{m}_{\ell(1)}, \mathbf{C}_{\ell(1)})$ of $\pi_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta})$ at the second level $\ell(2)$ (see [subsection 3.3.1](#)).

These steps are summarized in [Algorithm 3.1](#). The first three inputs are the spatial discretization parameter h_1 and the tolerances for sparse grid interpolation and quadrature, tol_1^{in} and tol_1^{qu} . Moreover, $\mathbf{K}_{\text{max}} := (K_{\text{max}}^{\text{in}}, K_{\text{max}}^{\text{qu}})$ comprises the maximum attainable sparse grid levels for the two adaptive operations. The last two inputs are the potential function, $\Phi(\boldsymbol{\theta}; \mathbf{y})$, and the prior density, $\pi_0(\boldsymbol{\theta})$.

Algorithm 3.1 Level One Adaptive Sparse Leja Algorithm for Bayesian Inversion

- 1: **procedure** LEVEL1SPARSELEJA($h_1, tol_1^{\text{in}}, tol_1^{\text{qu}}, \mathbf{K}_{\text{max}}, \Phi(\boldsymbol{\theta}; \mathbf{y}), \pi_0(\boldsymbol{\theta})$)
- 2: Compute the potential's surrogate via adaptive sparse grid interpolation

$$\Phi_{\ell(1)}(\boldsymbol{\theta}; \mathbf{y}) = \text{AdaptSGInterp}(tol_1^{\text{in}}, K_{\text{max}}^{\text{in}}, \Phi_1, \pi_0)$$

- 3: Construct the likelihood surrogate $L_{\ell(1)}(\boldsymbol{\theta}|\mathbf{y}) := \exp(-\Phi_{\ell(1)}(\boldsymbol{\theta}; \mathbf{y}))$
- 4: Compute the evidence via adaptive sparse grid quadrature

$$Z_{\ell(1)}(\mathbf{y}) = \text{AdaptSGQuad}(tol_1^{\text{qu}}, K_{\text{max}}^{\text{qu}}, L_{\ell(1)}, \pi_0)$$

- 5: Compute the posterior

$$\pi_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta}) := \frac{\pi_0(\boldsymbol{\theta}) L_{\ell(1)}(\boldsymbol{\theta}|\mathbf{y})}{Z_{\ell(1)}(\mathbf{y})}$$

- 6: Compute the expectation $\mathbf{m}_{\ell(1)}$ of $\pi_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta})$ w.r.t. $\pi_0(\boldsymbol{\theta})$
 - 7: Compute the covariance $\mathbf{C}_{\ell(1)}$ of $\pi_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta})$ w.r.t. $\pi_0(\boldsymbol{\theta})$
 - 8: **return** $\pi_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta}), \mathbf{m}_{\ell(1)}, \mathbf{C}_{\ell(1)}$
 - 9: **end procedure**
-

3.3. Level $\ell(j)$ with $j \geq 2$.

3.3.1. Gaussian approximation. At levels $\ell(j)$ with $j \geq 2$, we sequentially update the prior density such that the previous level posterior, $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$, is used as the prior; we detail the sequential update in [subsection 3.3.2](#). To be able to construct sparse grid approximations w.r.t. $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$, the underlying stochastic space needs to have a separable density (recall Assumption A1). However, $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ is usually not separable. Therefore, we approximate $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ with a density $\hat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ that allows us to obtain the required product structure. In this paper, we approximate $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ with the Gaussian density $\hat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ defined as

$$(3.1) \quad \hat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta}) := \mathbf{N}(\mathbf{m}_{\ell(j-1)}, \mathbf{C}_{\ell(j-1)}),$$

where $\mathbf{m}_{\ell(j-1)}$ and $\mathbf{C}_{\ell(j-1)}$ are the expectation and covariance matrix of $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$.

In most cases $\mathbf{C}_{\ell(j-1)}$ is not diagonal, i.e., $\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ does not have a product structure. Nevertheless, from the spectral decomposition $\mathbf{C}_{\ell(j-1)} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, we have $\mathbf{C}_{\ell(j-1)}^{1/2} = \mathbf{V}\mathbf{D}^{1/2}\mathbf{V}^{-1}$. We arrive at

$$(3.2) \quad \boldsymbol{\theta} = T_{\ell(j-1)}(\boldsymbol{\zeta}) := \mathbf{m}_{\ell(j-1)} + \mathbf{C}_{\ell(j-1)}^{1/2}\boldsymbol{\zeta} \Rightarrow \boldsymbol{\theta} \sim \mathbf{N}(\mathbf{m}_{\ell(j-1)}, \mathbf{C}_{\ell(j-1)}),$$

where $\boldsymbol{\zeta}$ is a standard Gaussian random variable, i.e., $\boldsymbol{\zeta} \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$.

Formula (3.2) allows to write a general multivariate Gaussian random variable with correlated components as a mapping of a standard multivariate Gaussian random variable, which has the desired product structure since the components of $\boldsymbol{\zeta}$ are uncorrelated and thus independent. In our context, we use (3.2) as follows. We first generate 1D (L)-Leja points weighted w.r.t. the standard normal density, that is, $w(\theta) := \exp(-\theta^2/2)/\sqrt{2\pi}$ in (2.8). Moreover, since the maximization defined in (2.8) is typically performed over a compact domain, we consider $X_i := \mathbb{R} \approx [-4, 4]$. For quadrature, we compute 1D quadrature weights w.r.t. normalized Hermite polynomials. We extend these constructions to N_{sto} dimensions via tensorization and employ (3.2) to obtain the desired weighted (L)-Leja points. Note that $T_{\ell(j-1)}$ in (3.2) can be seen as an *affine transport map* (see [29]).

3.3.2. Level update on tensor domain. We assume that $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ for $j \geq 2$ is not separable, hence we employ the Gaussian approximation (3.2) for all adaptive sparse grid operations; if $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ is separable, everything that follows is computed directly using $\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ for all levels greater than two. Thus, we sequentially update the prior in Bayes' formula (2.3) such that we reuse the Gaussian approximation of the posterior from the previous level, i.e.,

$$(3.3) \quad \begin{aligned} \pi_j^{\mathbf{y}}(\boldsymbol{\theta}) &:= \frac{L_j(\boldsymbol{\theta}|\mathbf{y})\pi_0(\boldsymbol{\theta})}{Z_j(\mathbf{y})} = \frac{L_j(\boldsymbol{\theta}|\mathbf{y})\pi_0(\boldsymbol{\theta})}{Z_j(\mathbf{y})} \frac{L_{j-1}(\boldsymbol{\theta}|\mathbf{y})}{L_{j-1}(\boldsymbol{\theta}|\mathbf{y})} \frac{Z_{j-1}(\mathbf{y})}{Z_{j-1}(\mathbf{y})} \\ &= \frac{L_{j-1}(\boldsymbol{\theta}|\mathbf{y})\pi_0(\boldsymbol{\theta})}{Z_{j-1}(\mathbf{y})} \frac{L_j(\boldsymbol{\theta}|\mathbf{y})}{L_{j-1}(\boldsymbol{\theta}|\mathbf{y})} \frac{Z_{j-1}(\mathbf{y})}{Z_j(\mathbf{y})} = \frac{\pi_{j-1}^{\mathbf{y}}(\boldsymbol{\theta}) \frac{L_j(\boldsymbol{\theta}|\mathbf{y})}{L_{j-1}(\boldsymbol{\theta}|\mathbf{y})}}{\frac{Z_j(\mathbf{y})}{Z_{j-1}(\mathbf{y})}} \\ &\approx \frac{\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})L_{\delta_j}(\boldsymbol{\theta}|\mathbf{y})}{Z_{\delta_j}(\mathbf{y})} = \frac{\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})L_{\delta_j}(\boldsymbol{\theta}|\mathbf{y}) \frac{\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})}{\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})}}{Z_{\delta_j}(\mathbf{y})}, \end{aligned}$$

where $L_{\delta_j}(\mathbf{y}) := L_j(\mathbf{y})/L_{j-1}(\mathbf{y})$ and $Z_{\delta_j}(\mathbf{y}) := Z_j(\mathbf{y})/Z_{j-1}(\mathbf{y})$. Note that in (3.3) we correct the bias introduced by the Gaussian approximation of the posterior from the previous level, $\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}$, with the ratio $\pi_{\ell(j-1)}^{\mathbf{y}}/\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}$.

First, we construct an adaptive sparse grid interpolation surrogate $\Phi_{\ell(\delta_j)}(\boldsymbol{\theta}; \mathbf{y})$ of $\Phi_{\delta_j}(\boldsymbol{\theta}; \mathbf{y})$ w.r.t. the density $\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ because it holds

$$L_{\delta_j}(\boldsymbol{\theta}|\mathbf{y}) := \frac{\exp(-\Phi_j(\boldsymbol{\theta}; \mathbf{y}))}{\exp(-\Phi_{j-1}(\boldsymbol{\theta}; \mathbf{y}))} = \exp(-\Phi_{\delta_j}(\boldsymbol{\theta}; \mathbf{y})).$$

Recall that the mapping $T_{\ell(j-1)}$ defined in (3.2) allows us to use adaptive sparse grid interpolation w.r.t. the Gaussian approximation $\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$. To construct the surrogate for the potential function $\Phi_{\delta_j}(\boldsymbol{\theta}; \mathbf{y})$ we employ $T_{\ell(j-1)}$ and obtain

$$\Phi_{\delta_j}(\boldsymbol{\theta}; \mathbf{y}) \approx \Phi_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta}); \mathbf{y}).$$

$\Phi_{\ell(\delta_j)}$ gives the approximation $L_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta})|\mathbf{y}) := \exp(-\Phi_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta}); \mathbf{y}))$.
To evaluate the ratio of evidences $Z_{\delta_j}(\mathbf{y})$ we make use of (3.3), i.e.,

$$(3.4) \quad \begin{aligned} Z_{\delta_j}(\mathbf{y}) &= \int_{\mathbf{X}} \widehat{\pi}_{j-1}^{\mathbf{y}}(\boldsymbol{\theta}) L_{\delta_j}(\boldsymbol{\theta}|\mathbf{y}) \frac{\pi_{j-1}^{\mathbf{y}}(\boldsymbol{\theta})}{\widehat{\pi}_{j-1}^{\mathbf{y}}(\boldsymbol{\theta})} d\boldsymbol{\theta} \\ &\approx \int_{\mathbf{X}} \widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(T_{\ell(j-1)}(\boldsymbol{\zeta})) \cdot L_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta})|\mathbf{y}) \cdot \frac{\pi_{\ell(j-1)}^{\mathbf{y}}(T_{\ell(j-1)}(\boldsymbol{\zeta}))}{\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(T_{\ell(j-1)}(\boldsymbol{\zeta}))} d\boldsymbol{\zeta} \end{aligned}$$

and numerically integrate $(L_{\ell(\delta_j)}\pi_{\ell(j-1)}^{\mathbf{y}}/\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}) \circ T_{\ell(j-1)}$ w.r.t. the density $\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta})$ via adaptive sparse grid quadrature to obtain the approximation $Z_{\ell(\delta_j)}(\mathbf{y}) \approx Z_{\delta_j}(\mathbf{y})$. Thus, at each level $\ell(j)$ with $j \geq 2$ we obtain the posterior approximation

$$\pi_j^{\mathbf{y}}(\boldsymbol{\theta}) \approx \pi_{\ell(j)}^{\mathbf{y}}(\boldsymbol{\theta}) := Z_{\ell(\delta_j)}^{-1}(\mathbf{y}) \pi_{\ell(j-1)}^{\mathbf{y}}(T_{\ell(j-1)}(\boldsymbol{\zeta})) L_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta})|\mathbf{y}).$$

For all other quadrature computations, we proceed analogously to (3.4). To simplify the notation, denote $R_{\ell(j-1)} := \pi_{\ell(j-1)}^{\mathbf{y}}/\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}$. Given an integrable function $g(\boldsymbol{\theta})$, we integrate $\int_{\mathbf{X}} g(\boldsymbol{\theta}) \pi_j^{\mathbf{y}}(\boldsymbol{\theta}) d\boldsymbol{\theta}$, which reads as

$$\int_{\mathbf{X}} g(T_{\ell(j-1)}(\boldsymbol{\zeta})) L_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta})|\mathbf{y}) R_{\ell(j-1)}(T_{\ell(j-1)}(\boldsymbol{\zeta})) \widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}(T_{\ell(j-1)}(\boldsymbol{\zeta})) d\boldsymbol{\zeta},$$

via adaptive sparse grid quadrature w.r.t. $\widehat{\pi}_{\ell(j-1)}^{\mathbf{y}}$. The above formula is used to assess the expectation and covariance matrix of $\pi_j^{\mathbf{y}}(\boldsymbol{\theta})$. Moreover, at level J of our approach, we integrate the QoI $g \approx g_{\ell(J)}$ w.r.t. $\widehat{\pi}_{\ell(J)}^{\mathbf{y}}(\boldsymbol{\theta})$. In this way, we perform the multilevel decomposition implicitly, different from standard multilevel methods in which the QoI is assessed explicitly via telescoping sums. Note that at the end of our multilevel algorithm we also obtain a surrogate for the posterior density which can be further used, for example, in an uncertainty propagation setting.

We summarize the steps for levels greater than two in Algorithm 3.2. The first three inputs are the number of levels, J , the sequence of mesh sizes, \mathbf{h} , and \mathbf{tol} , which comprises the tolerances for adaptive sparse grid interpolation and quadrature at all levels. The next input denotes the maximum reachable levels for the adaptive algorithms, $\mathbf{K}_{\max} := (K_{\max}^{\text{qu}}, K_{\max}^{\text{in}})$. Finally, π_0 is the prior density, $\Phi(\boldsymbol{\theta}; \mathbf{y})$ is the potential function and g is the QoI. We combine Algorithms 3.1 and 3.2 and depict all steps in our proposed multilevel approach in Figure 1.

3.4. Computational cost. The largest computational effort in the proposed multilevel methodology is spent in finding the interpolation surrogates since this involves evaluations of the forward operator. Thus, for $j = 1, 2, \dots, J$, let C_{h_j} denote the cost of the evaluating once the forward model discretized using a mesh depending on h_j . Additionally, let $N_{\text{tol}_{j-j+1}^{\text{in}}}$ denote the number of forward operator solves to achieve tolerance $\text{tol}_{j-j+1}^{\text{in}}$ for adaptive sparse interpolation. Then, the *total interpolation cost* of our approach reads

$$\text{cost}_{\text{MLSL}}^{\text{in}} = \sum_{j=1}^{J-1} (N_{\text{tol}_{j-j+1}^{\text{in}}} + N_{\text{tol}_{j-j}^{\text{in}}}) C_{h_j} + N_{\text{tol}_1^{\text{in}}} C_{h_J}.$$

We obtain the above number because for each level except the last one, the same potential function and thus the same forward model enters two different likelihood

Algorithm 3.2 Multilevel Adaptive Sparse Leja Algorithm

-
- 1: **procedure** MULTILEVELADAPTSPARSELEJA($J, \mathbf{h}, \mathbf{tol}, \mathbf{K}_{\max}, \Phi(\boldsymbol{\theta}; \mathbf{y}), \pi_0(\boldsymbol{\theta}), g$)
 - 2: Use [Algorithm 3.1](#) to obtain

$$\pi_{\ell(1)}^{\mathbf{y}}(\boldsymbol{\theta}), \mathbf{m}_{\ell(1)}, \mathbf{C}_{\ell(1)} = \text{Level1SparseLeja}(h_1, \mathbf{tol}_1^{\text{in}}, \mathbf{tol}_1^{\text{qu}}, \mathbf{K}_{\max}, \Phi, \pi_0)$$

- 3: **for** $j \leftarrow 2, J$ **do**

- 4: Construct the Gaussian approximation [\(3.1\)](#) and the mapping [\(3.2\)](#)

$$\hat{\pi}_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta}) := \text{N}(\mathbf{m}_{\ell(j-1)}, \mathbf{C}_{\ell(j-1)}), \quad T_{\ell(j-1)}(\boldsymbol{\zeta}) := \mathbf{m}_{\ell(j-1)} + \mathbf{C}_{\ell(j-1)}^{1/2} \boldsymbol{\zeta}$$

- 5: Compute the potentials ratio surrogate via adaptive interpolation

$$\Phi_{\ell(\delta_j)}(\boldsymbol{\theta}; \mathbf{y}) = \text{AdaptSGInterp}(\mathbf{tol}_j^{\text{in}}, \mathbf{K}_{\max}^{\text{in}}, \Phi_{\delta_j} \circ T_{\ell(j-1)}(\boldsymbol{\zeta}), \hat{\pi}_{\ell(j-1)}^{\mathbf{y}})$$

- 6: Construct $L_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta})|\mathbf{y}) := \exp(-\Phi_{\ell(\delta_j)}(T_{\ell(j-1)}(\boldsymbol{\zeta}); \mathbf{y}))$
- 7: Compute the evidence ratio via adaptive quadrature

$$Z_{\ell(\delta_j)}(\mathbf{y}) = \text{AdaptSGQuad}(\mathbf{tol}_j^{\text{qu}}, \mathbf{K}_{\max}^{\text{qu}}, L_{\ell(\delta_j)}, \hat{\pi}_{\ell(j-1)}^{\mathbf{y}})$$

- 8: Compute the updated posterior

$$\pi_{\ell(j)}^{\mathbf{y}}(\boldsymbol{\theta}) := \frac{\pi_{\ell(j-1)}^{\mathbf{y}}(\boldsymbol{\theta}) L_{\ell(\delta_j)}(\boldsymbol{\theta}|\mathbf{y})}{Z_{\ell(\delta_j)}(\mathbf{y})}$$

- 9: Compute the expectation $\mathbf{m}_{\ell(j)}$ of $\pi_{\ell(j)}^{\mathbf{y}}(\boldsymbol{\theta})$ w.r.t. $\hat{\pi}_{\ell(j-1)}^{\mathbf{y}}(T_{\ell(j-1)}(\boldsymbol{\zeta}))$
 - 10: Compute the covariance $\mathbf{C}_{\ell(j)}$ of $\pi_{\ell(j)}^{\mathbf{y}}(\boldsymbol{\theta})$ w.r.t. $\hat{\pi}_{\ell(j-1)}^{\mathbf{y}}(T_{\ell(j-1)}(\boldsymbol{\zeta}))$
 - 11: **end for**
 - 12: Compute $g_{\ell(J)} \approx g$ and integrate w.r.t. $\pi_{\ell(J)}^{\mathbf{y}} \rightarrow \mathcal{I}_{\ell(J)}$
 - 13: **return** $\mathcal{I}_{\ell(J)}$
 - 14: **end procedure**
-

ratios (see [\(3.3\)](#)). However, because we update the prior as the level increases, we expect the number of forward model solves to decrease significantly with the level.

All other costs are due to computations depending on the interpolation surrogates. These costs are however insignificant compared to the evaluation cost of a computationally expensive forward operator.

4. Dimension adaptivity with sparse grids. In this section we discuss the construction of the multiindex set \mathcal{K} used in the sparse grid approximations defined in [subsection 2.3](#) via adaptive refinement. For interpolation we consider a standard adaptive strategy as well as an enhanced approach that employs directional variance information. For quadrature we employ a standard adaptive strategy. In the following, our notation is similar to [\[11\]](#).

4.1. Standard dimension-adaptive interpolation and quadrature. Adaptive refinement is preferred especially when the underlying problem has a richer structure, such as anisotropic coupling of the input parameters or lower intrinsic dimensionality – which is typically the case in most problems (see, e.g., [\[6, 11, 12, 45\]](#)). The standard strategy is based on the dimension-adaptive algorithm of [\[15, 19\]](#). The

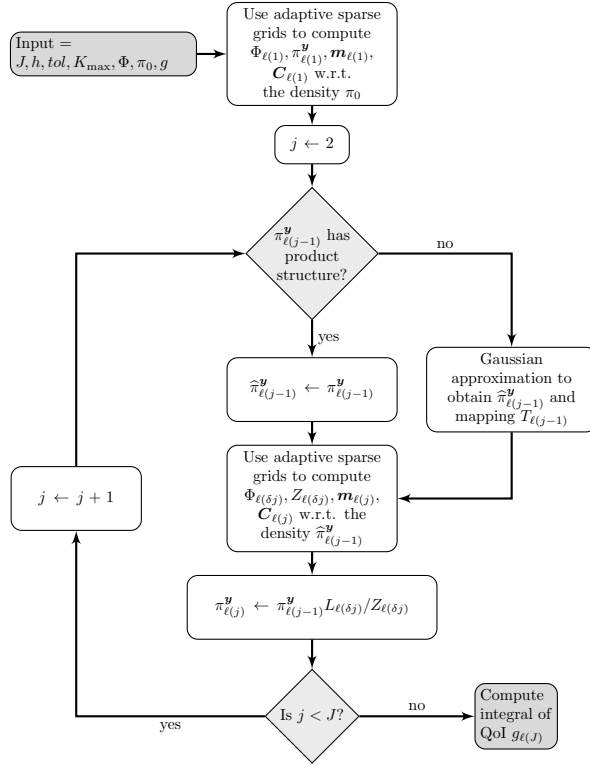


FIG. 1. Flowchart of the multilevel adaptive sparse Leja algorithm. The goal is to approximate the posterior density $\pi^{\mathbf{y}}$ and $\mathbb{E}_{\pi^{\mathbf{y}}}[g]$ for an output QoI g at level J given the prior density π_0 .

algorithm is described, e.g., in [6, 15, 32]. We summarize only the basic idea below. $\mathcal{K} = \{\mathbf{1}\}$ initially. Each refinement step is performed using the following principle: if a current multiindex contributes significantly to the approximation, its adjacent neighbours are likely to contribute as well. Therefore, the forward neighbours of the multiindex with the largest contribution are added to \mathcal{K} provided that \mathcal{K} remains admissible. The contribution of each \mathbf{k} is assessed via a *refinement indicator* $\epsilon(\mathbf{k})$, whose choice has a crucial impact on the performance of the adaptive algorithm.

We define

$$(4.1) \quad \epsilon(\mathbf{k}) := s^{\text{op}}(\Delta_{\mathbf{k}}^{\text{op}}[f^{\mathbf{N}_{\text{sto}}}] / \delta N_{\mathbf{k}}),$$

where s^{op} is a function depending on the multivariate surplus $\Delta_{\mathbf{k}}^{\text{op}}[f^{\mathbf{N}_{\text{sto}}}]$ and $\delta N_{\mathbf{k}}$ is the number model evaluations needed to assess $\Delta_{\mathbf{k}}^{\text{op}}[f^{\mathbf{N}_{\text{sto}}}]$. Note that $\delta N_{\mathbf{k}}$ penalizes subspaces with a large number of points.

For sparse grid quadrature, we consider

$$s^{\text{qu}}(\Delta_{\mathbf{k}}^{\text{qu}}[f^{\mathbf{N}_{\text{sto}}}]) := \|\Delta_{\mathbf{k}}^{\text{qu}}[f^{\mathbf{N}_{\text{sto}}}]\|_{L^1} = |\Delta_{\mathbf{k}}^{\text{qu}}[f^{\mathbf{N}_{\text{sto}}}]|,$$

which is a surrogate for the local quadrature error. For sparse interpolation, we use

$$(4.2) \quad s^{\text{in}}(\Delta_{\mathbf{k}}^{\text{in}}[f^{\mathbf{N}_{\text{sto}}}]) := \|\Delta_{\mathbf{k}}^{\text{in}}[f^{\mathbf{N}_{\text{sto}}}]\|_{L^2}.$$

As in [11], we employ $\|\Delta_{\mathbf{k}}^{\text{in}}[f^{\mathbf{N}_{\text{sto}}}]\|_{L^2}$ in the standard refinement indicator (4.2) because it yields the local variance contribution of the surplus to the total variance.

4.2. Directional variance dimension-adaptive sparse interpolation. Dimension adaptivity based on error indicators such as (4.1) does not inherently distinguish between the individual input parameters. Since in most problems the input parameters are anisotropically coupled, we wish to exploit this structure and tune the adaptive process such that it stops refining directions that are rendered unimportant. This is particularly important in our proposed approach since we update the prior starting with level $\ell(2)$ and thus we have updated information about the model's stochastic input parameters. To this end, for sparse interpolation, we enhance the standard adaptive strategy such that we additionally compute a global measure of importance of each input parameter via *total directional variances*, and we stop refining the directions having insignificant total directional variances.

To enhance the standard adaptive approach for interpolation, we proceed analogously to [11, 45] and perform a Sobol' decomposition [40] of the active set \mathcal{A} to obtain directional variance surpluses, that is,

$$\left\| \sum_{\mathbf{k} \in \mathcal{A}} \Delta_{\mathbf{k}}^{\text{in}} [f^{N_{\text{sto}}}] \right\|_{L^2}^2 = \Delta V_{\mathcal{A}}^0 + \sum_{i=1}^{N_{\text{sto}}} \Delta V_{\mathcal{A}}^i + \Delta V_{\mathcal{A}}^{\text{int}},$$

where $\Delta V_{\mathcal{A}}^0 := \gamma_{\mathbf{0}}^2$ refers to the expectation contribution,

$$\Delta V_{\mathcal{A}}^i := \sum_{\ell \in \mathcal{J}_i} \Delta \gamma_{\ell}^2, \quad i = 1, \dots, N_{\text{sto}},$$

where $\mathcal{J}_i = \{\mathbf{0} < \ell \leq \mathbf{I}_{\mathbf{k}} : \ell_i \neq 0 \wedge \ell_j = 0, \forall j \neq i\}$, are surplus contributions to all individual variances, and $\Delta V_{\mathcal{A}}^{\text{int}} := \sum_{\mathbf{m} \in \mathcal{J}_{\text{int}}} \Delta \gamma_{\mathbf{m}}^2$ refers to the variance surplus due to all possible interactions, where $\mathcal{J}_{\text{int}} = \bigcup_{i=1}^{N_{\text{sto}}} \mathcal{J}_{i,\text{int}}$, $\mathcal{J}_{i,\text{int}} = \{\mathbf{0} < \mathbf{m} \leq \mathbf{I}_{\mathbf{k}} : \mathbf{m}_i \neq 0\}$. Further, we compute *total directional variance surpluses* as

$$\Delta V_{\mathcal{A}}^{i,\text{tot}} := \Delta V_{\mathcal{A}}^i + \Delta V_{\mathcal{A}}^{i,\text{int}},$$

where $\Delta V_{\mathcal{A}}^{i,\text{int}} := \sum_{\mathbf{m} \in \mathcal{J}_{i,\text{int}}} \Delta \gamma_{\mathbf{m}}^2$ denotes the contribution due to all interactions involving direction i . Note that $\Delta V_{\mathcal{A}}^{i,\text{tot}}$ can be seen as a global measure of importance for each stochastic input: a large $\Delta V_{\mathcal{A}}^{i,\text{tot}}$ implies that the i th parameter is significant from a stochastic perspective. To this end, we prescribe N_{sto} user-defined *directional tolerances* $\boldsymbol{\tau}^{\text{in}} := (\tau_1^2, \tau_2^2, \dots, \tau_{N_{\text{sto}}}^2)$ and ascertain the importance of each input directions by comparing $\Delta V_{\mathcal{A}}^{i,\text{tot}}$ with τ_i^2 for $i = 1, 2, \dots, N_{\text{sto}}$. When the stochastic direction i is rendered unimportant, we simply stop adding multiindices whose i th component exceeds the maximum i th index in the current multiindex set \mathcal{K} . In this way, the algorithm preferentially refines the most important directions, thus decreasing the overall computational cost. When neither of the directional tolerances are met, the enhanced algorithm reduces to the standard approach in subsection 4.1.

5. Numerical experiments. In this section we present the numerical results obtained using our proposed multilevel approach for Bayesian inversion.

5.1. Simple quadrature showcase. In this test case we investigate the behaviour of weighted (L)-Leja points in integration problems of the form

$$(5.1) \quad \mathcal{I}(g) := \int g(\boldsymbol{\theta}) \pi^{\mathbf{y}}(\boldsymbol{\theta}) d\boldsymbol{\theta},$$

where $g(\boldsymbol{\theta})$ is an integrable function and $\pi^{\mathbf{y}}(\boldsymbol{\theta})$ is the posterior density; we outline the setup used to compute $\pi^{\mathbf{y}}(\boldsymbol{\theta})$ below. We assess (5.1) via quadrature w.r.t. two different weight functions. In the first case, we employ a standard importance-sampling-based strategy (recall (2.5)). Specifically, adaptive sparse grid quadrature w.r.t. the prior density, $\pi_0(\boldsymbol{\theta})$, with tolerance $tol_{\pi_0}^{\text{qu}}$ is used:

$$(5.2) \quad \mathcal{I}(g) = \int g(\boldsymbol{\theta}) \frac{L(\boldsymbol{\theta}|\mathbf{y})\pi_0(\boldsymbol{\theta})}{Z(\mathbf{y})} d\boldsymbol{\theta} \approx Z_{N_{\text{pr}}}^{-1}(\mathbf{y}) \left(\sum_{n=1}^{N_{\text{pr}}} g(\boldsymbol{\theta}_{n,\text{pr}}) L(\boldsymbol{\theta}_{n,\text{pr}}|\mathbf{y}) w_{n,\text{pr}} \right),$$

where $\{\boldsymbol{\theta}_{n,\text{pr}}\}_{n=1}^{N_{\text{pr}}}$ are (L)-Leja nodes computed w.r.t. $\pi_0(\boldsymbol{\theta})$ and

$$Z_{N_{\text{pr}}}(\mathbf{y}) = \sum_{n=1}^{N_{\text{pr}}} \boldsymbol{\theta}_{n,\text{pr}} L(\boldsymbol{\theta}_{n,\text{pr}}|\mathbf{y}) w_{n,\text{pr}}.$$

In the second strategy, we compute (5.1) using our proposed approach. We integrate (5.1) numerically via adaptive sparse grid quadrature w.r.t. the Gaussian approximation $\hat{\pi}^{\mathbf{y}}(\boldsymbol{\theta})$ of the posterior density (recall (3.1)), using a tolerance $tol_{\hat{\pi}^{\mathbf{y}}}^{\text{qu}}$:

$$(5.3) \quad \mathcal{I}(g) = \int g(\boldsymbol{\theta}) \frac{\pi^{\mathbf{y}}(\boldsymbol{\theta})}{\hat{\pi}^{\mathbf{y}}(\boldsymbol{\theta})} \hat{\pi}^{\mathbf{y}}(\boldsymbol{\theta}) d\boldsymbol{\theta} \approx \sum_{n=1}^{N_{\text{post}}} g(T(\boldsymbol{\zeta}_{n,\text{post}})) \frac{\pi^{\mathbf{y}}(T(\boldsymbol{\zeta}_{n,\text{post}}))}{\hat{\pi}^{\mathbf{y}}(T(\boldsymbol{\zeta}_{n,\text{post}}))} w_{n,\text{post}},$$

where $\{\boldsymbol{\zeta}_{n,\text{post}}\}_{n=1}^{N_{\text{post}}}$ are (L)-Leja nodes computed w.r.t. the standard multivariate normal density, $N(\mathbf{0}, I)$, and $T(\boldsymbol{\zeta}) := \mathbf{m} + \mathbf{C}^{1/2}\boldsymbol{\zeta}$, where \mathbf{m} and \mathbf{C} are the expectation and covariance matrix associated with the density $\hat{\pi}^{\mathbf{y}}(\boldsymbol{\theta})$.

We consider the following forward model

$$\mathcal{G}(\boldsymbol{\theta}) := \frac{A(\theta_1)}{(w(\theta_2)\pi)^2} (\sin(w(\theta_2)\pi x) - \sin(w(\theta_2)\pi x)),$$

where $x \in [0, 1]$, $A(\theta_1) = 20\theta_1 + 1$ and $w(\theta_2) = \theta_2 + 1.2$. We employ Bayesian inversion to infer (θ_1, θ_2) . The prior is the uniform density in $[0, 1]^2$, i.e., $\pi_0 = U(0, 1)^2$. The observation data \mathbf{y} are generated synthetically using $(\theta_1, \theta_2)_{\text{true}} = (0.45, 0.65)$. We take $N_{\text{obs}} = 9$ measurements at locations $o_j = 0.1j$, $j = 1, \dots, 9$, assumed to be corrupted by additive Gaussian noise $\eta \sim N(\mathbf{0}, 0.1^2 I)$. We depict the prior and posterior densities in the left figure in Figure 2. Observe that the posterior is unimodal and non-symmetric, but it can be well approximated with a Gaussian density.

In the numerical experiments, we let $g(\boldsymbol{\theta}) := \exp(-\theta_1 - \theta_2)$ in (5.1). We compute a reference solution using $3 \cdot 10^5$ Metropolis-Hastings MCMC (MH) samples obtained from a random walk Gaussian proposal with initial sample $\boldsymbol{\theta}_0 = (1, 1)$ and covariance matrix $C_{\text{MH}} = 7 \cdot 10^{-3} I$. The acceptance rate is 44%. Additionally, we employ a tolerance $tol_{\pi_0}^{\text{qu}} = 10^{-11}$ in (5.2) and a tolerance $tol_{\hat{\pi}^{\mathbf{y}}}^{\text{qu}} = 10^{-5}$ in (5.3).

The results are summarized in Table 1. The employed tolerances in the two (L)-Leja sparse grid quadrature approaches are sufficient to match four digits of the reference results. However, integrating w.r.t. the prior requires 1603 nodes, whereas our approach which uses the Gaussian approximation of the posterior as weight function requires only 49 quadrature nodes, i.e., almost 33 times fewer points. This is because the support of the prior density, $\pi_0(\boldsymbol{\theta})$, is significantly larger than the support of the posterior: when integrating w.r.t. the prior density, the adaptive algorithm places a large number of quadrature points outside of the support of the posterior.

TABLE 1

Results for the quadrature problem (5.1) using a reference MH solution with $3 \cdot 10^5$ samples, integration w.r.t. the prior density as in (5.2) and our proposed approach in which we integrate w.r.t. the Gaussian approximation of the posterior, as showed in (5.3).

Method	No. quadrature points	Result
MH	$3 \cdot 10^5$	0.33813
Integration w.r.t. π_0	1603	0.33813
Integration w.r.t. $\hat{\pi}^{\mathbf{y}}(\boldsymbol{\theta})$	49	0.33811

We visualize the quadrature nodes corresponding to (5.2) and (5.3) in the center and right figures in Figure 2, respectively.

Remark 5.1. Adaptive sparse grid quadrature w.r.t. an uninformative prior can sometimes stop early since the quadrature points will fall outside of the support of the integrated function, yielding null evaluations and thus null error indicators. To overcome this, one could employ non-adaptive quadrature with a sufficiently large, a priori chosen number of nodes to cover the support of the integrated function.

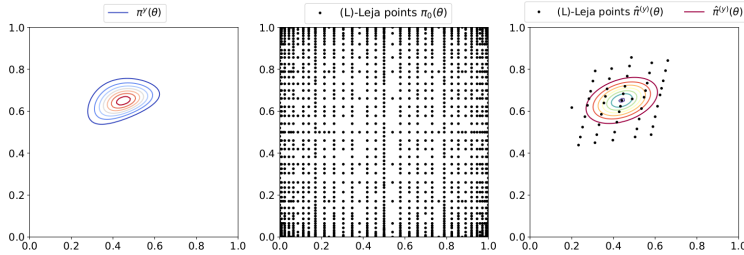


FIG. 2. Left: Posterior density $\pi^{\mathbf{y}}(\boldsymbol{\theta})$. Center: (L)-Leja points computed w.r.t. the uniform prior density used in the integration problem (5.2). Right: Gaussian approximation $\hat{\pi}^{\mathbf{y}}(\boldsymbol{\theta})$ of the posterior and the associated weighted (L)-Leja points used in (5.3).

5.2. Source inversion with one source in a 2D spatial domain. Consider a two dimensional Bayesian inverse problem in which the forward model $\mathcal{G}(\boldsymbol{\theta})$ is an elliptic PDE defined on $\Omega := [0, 1]^2$,

$$(5.4) \quad \begin{aligned} -(u_{xx} + u_{yy})(x, y) &= A(\alpha) \exp(-[(x - \theta_1)^2 + (y - \theta_2)^2]/2\alpha^2), & (x, y) \in \Omega \\ u(x, y) &= 0, & (x, y) \in \partial\Omega, \end{aligned}$$

with $A(\alpha) = 5/(2\pi\alpha^2)$ and $\alpha = 0.2$.

The goal is to infer the coordinates (θ_1, θ_2) of the source term in the right-hand side, i.e., we seek the solution to a source inversion problem. We perform the multilevel Bayesian inversion as described in Algorithms 3.1 and 3.2 with three levels, i.e., $J = 3$. Thus $j = 1, 2, 3$. The employed multilevel setup is summarized in Table 2. Standard triangular finite elements (FEs) with mesh widths h_j are used for spatial discretization. To find surrogates for the potential function at each level in our proposed multilevel approach we employ both adaptive sparse interpolation variants summarized in section 4. Recall that for standard adaptive interpolation we have tolerances $tol_1^{\text{in}}, tol_2^{\text{in}}, tol_3^{\text{in}}$ (see subsection 4.1) whereas for directional variances-based adaptivity from subsection 4.2 we additionally have the directional tolerances $\{\tau_j^{\text{in}}\}_{j=1}^3$. We choose the FE mesh widths and adaptive interpolation tolerances such that the approximation errors are quantitatively similar. At level $\ell(1)$ we combine h_1

with tol_3^{in} and τ_3^{in} , at level $\ell(2)$, h_2 is combined with tol_2^{in} and τ_2^{in} , and at level $\ell(3)$ we employ h_3 together with tol_1^{in} and τ_1^{in} . For adaptive sparse grid quadrature we employ small tolerances $tol_1^{\text{qu}}, tol_2^{\text{qu}}, tol_3^{\text{qu}}$ to prevent the adaptive algorithm to stop too early especially when integrating w.r.t. the prior density (recall [Remark 5.1](#)).

TABLE 2
Multilevel setup for the 2D inversion problem with forward model (5.4).

Level	h	tol^{in}	τ^{in}	tol^{qu}
$\ell(1)$	$h_1 = \sqrt{2}/2^4$	$tol_3^{\text{in}} = 10^{-5}$	$\tau_3^{\text{in}} = (10^{-7}, 10^{-7})$	$tol_3^{\text{qu}} = 10^{-12}$
$\ell(2)$	$h_2 = \sqrt{2}/2^5$	$tol_2^{\text{in}} = 10^{-4}$	$\tau_2^{\text{in}} = (10^{-6}, 10^{-6})$	$tol_2^{\text{qu}} = 10^{-11}$
$\ell(3)$	$h_3 = \sqrt{2}/2^6$	$tol_1^{\text{in}} = 10^{-3}$	$\tau_1^{\text{in}} = (10^{-5}, 10^{-5})$	$tol_1^{\text{qu}} = 10^{-10}$

Since the source locations need to reside inside Ω , the prior is the uniform density in $[0, 1]^2$, i.e., $\pi_0 = U(0, 1)^2$. We consider 16 sensor locations at $(0.2i, 0.2j)$ for $i, j = 1, 2, 3, 4$. The measurements are obtained synthetically by discretizing the forward model on a finer mesh, i.e., $h = \sqrt{2}/2^7$ to avoid committing an “inverse crime”. Moreover, $\theta_{\text{true}} = (0.35, 0.65)$ and the additive Gaussian noise $\eta \sim N(\mathbf{0}, 0.2^2 I)$.

The QoI is the posterior mean $\mathbb{E}_{\pi^y}[\theta]$. We compute a reference solution using $2 \cdot 10^5$ samples obtained from a random walk Metropolis-Hastings algorithm with Gaussian proposal having covariance matrix $C_{\text{MH}} = 4 \cdot 10^{-3} I$, started from $\theta_0 = (1, 1)$. The acceptance rate of the chain is 64%. To obtain a comprehensive overview of the accuracy and cost of our approach, we compare it with the standard three-level approach in which all adaptive sparse grid operations are performed w.r.t. the prior density. Moreover, the QoI is assessed using the classical telescoping sum. To simplify the notation, in the following we use the abbreviation StdML to refer to the standard multilevel approach. MLLejaStd refers to our approach in which standard dimension-adaptive interpolation is used at each level and MLLejaDV refers to our approach combined with directional variance-based adaptive interpolation summarized in [subsection 4.2](#). The results are presented in [Table 3](#). Observe that all multilevel methods yield results very close to the reference estimate. Thus, the two variants of our proposed approach, MLLejaStd and MLLejaDV, are comparably accurate as the sampling-based and the standard multilevel solutions.

TABLE 3
Comparison of estimates of $\mathbb{E}[\pi^y(\theta)]$ for the source inversion problem with forward model (5.4). We first compute a reference solution using $2 \cdot 10^5$ MH samples. Afterwards, we employ StdML and the two variants of our proposed multilevel approach, MLLejaStd and MLLejaDV.

Method	$\mathbb{E}_{\pi^y}[\theta]$
MH	(0.3628, 0.6370)
StdML	(0.3631, 0.6368)
MLLejaStd	(0.3630, 0.6369)
MLLejaDV	(0.3630, 0.6369)

In [Figure 3](#), we visualize the results for all employed multilevel methods as follows. The left subplots show the results for StdML, whereas the center and right subplots depict the results for MLLejaStd and MLLejaDV respectively. Furthermore, at each level, the prior as well as the corresponding (L)-Leja points used to find the interpolation surrogate are visualized in the top part. In the bottom plots, we depict the resulting posterior densities. At level $\ell(1)$ we obtain the same three posteriors since the prior is the same in all cases. However, starting with level $\ell(2)$ the sequential update of the prior in the proposed approach leads to significantly fewer interpolation

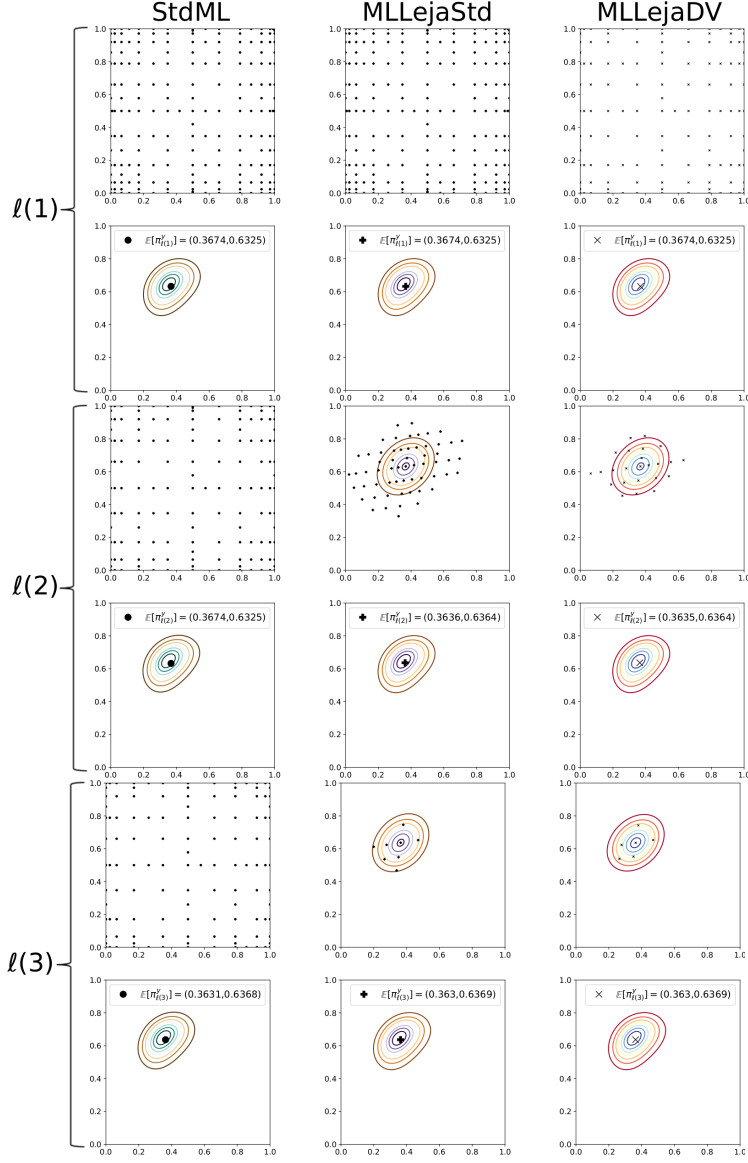


FIG. 3. Results obtained using StdML (left), MLEjaStd (center) and MLEjaDV (right) to find the adaptive sparse grid interpolation surrogate for the potential function in the source inversion problem with forward model (5.4). At each of the three levels, in the top plots we depict the prior density and the corresponding weighted (L)-Leja points used to find the surrogate. At levels $\ell(2)$ and $\ell(3)$, the prior is the Gaussian approximation of the posterior from the posterior level. In the bottom plots we depict the corresponding posterior density solution.

points compared to StdML, which places a large number weighted (L)-Leja points outside of the support of the corresponding posterior. Moreover, comparing the two variants of the proposed approach, MLEjaDV requires fewer (L)-Leja points than MLEjaStd. This is because at both levels $\ell(2)$ and $\ell(3)$ in MLEjaDV, the two total directional variances fall below the imposed tolerances. Thus MLEjaDV discovers and exploits more structure in the underlying approximation problem.

We visualize the multiindex sets for the three multilevel variants in Figure 4. At level $\ell(1)$ the multiindex sets corresponding to StdML and MLEjaStd are symmetric; this is because the two stochastic parameters have equal importance w.r.t. the prior density, which is the 2D (symmetric) uniform density. MLEjaDV leads to a smaller multiindex set since the directional variances τ_3^{in} fall below $(10^{-7}, 10^{-7})$, but there is no clear distinction between the two input directions as well. However, at level $\ell(2)$ we observe a different behaviour in the two variants of our proposed approach, MLEjaStd and MLEjaDV. Recall that in these two variants the prior density is the Gaussian approximation of the posterior density from level $\ell(1)$. Computing the eigenvalues (λ_1, λ_2) of its covariance matrix, we obtain $\lambda_1 = 0.0097$ and $\lambda_2 = 0.0055$. Therefore, the first direction is more important than the second, which is reflected in the two multiindex sets. On the other hand, the multiindex set for the standard approach remains symmetric because the prior density is unchanged. Finally, at level $\ell(3)$ we see low-cardinality multiindex sets for both MLEjaStd and MLEjaDV. This is because at this stage we have an informative prior, thus the likelihood ratio is close 1, which requires little approximation effort. Hence going beyond level $\ell(3)$ is not necessary for our approach.

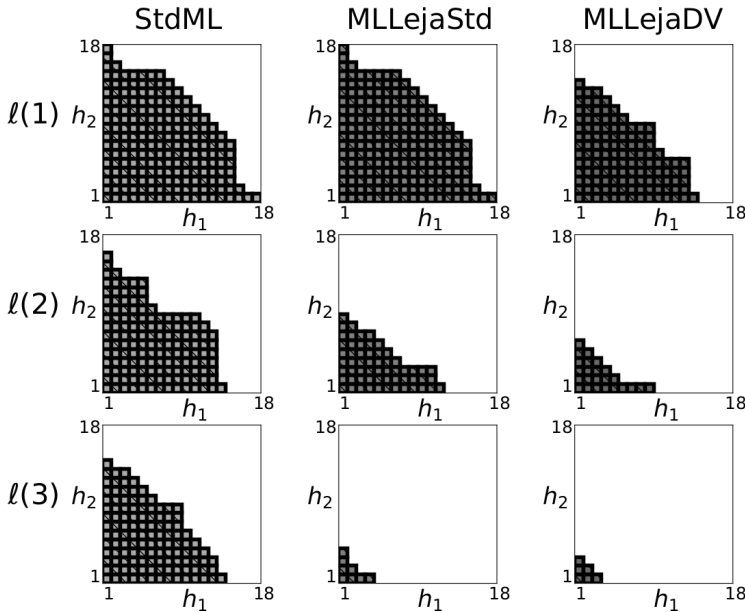


FIG. 4. Multiindex sets corresponding to adaptive sparse grid interpolation in the source inversion problem (5.4) for StdML (left), MLEjaStd (center) and MLEjaDV (right).

The costs of all multilevel methods are visualized in Figure 5. The number of forward model evaluations needed to find the adaptive sparse grid surrogate of the potential function are shown on the left side. In the right plot we depict the number of evaluations of the surrogate in all quadrature computations. Note in all multilevel variants we need quadrature to assess the evidences and expectations at all three levels. Additionally, in MLEjaStd and MLEjaDV we need to compute the covariance matrices at levels $\ell(1)$ and $\ell(2)$ as well, which are needed in the Gaussian approximation of the associated posteriors and affine mapping (recall (3.1) and (3.2)). However, since we integrate w.r.t. the same weight function, we keep all surrogate evaluations

in a look-up table and reuse them whenever the same grid points are used for different evaluations. We observe that at level $\ell(1)$ our proposed approach is slightly more expensive for interpolation which is due to the need to evaluate the FE solver on level $\ell(1)$ at level $\ell(2)$ as well: recall that at level $\ell(2)$ we construct a sparse grid surrogate for the ratio of potential functions. However, the increased cost is not significant since it involves evaluations of the coarsest FE solver, which are very fast. Starting with level $\ell(2)$ we see significant cost savings for both interpolation and quadrature. On the one hand, at level $\ell(2)$, MLLejaStd leads to about 2 times fewer forward model evaluations for sparse grid interpolation and around 12.5 fewer sparse grid quadrature evaluations. Moreover, at level $\ell(3)$ we obtain about 15 times fewer interpolation points and 7.5 times fewer quadrature nodes. On the other hand, MLLejaDV leads to about 5 times fewer interpolation nodes and about 12.5 times fewer quadrature evaluations. Furthermore, we obtain 20 times fewer interpolation nodes and 9.5 times fewer quadrature nodes at level $\ell(3)$. These results clearly show that updating the prior information in our multilevel approach for Bayesian Inversion leads to significant cost reduction in finding and evaluating sparse grid surrogates.

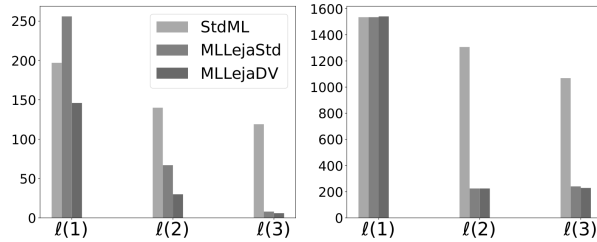


FIG. 5. *Left: total number of forward model evaluations needed in the adaptive sparse interpolation of the potential using the three multilevel variants in the source inversion problem (5.4). Right: total number of quadrature nodes.*

5.3. Source inversion with two sources in a 2D spatial domain. For a more comprehensive overview of the proposed approach, we consider now a test case with multimodal observation data. In particular, we consider another source inversion test case in which we use two sources to generate the data – to have bimodal observation data – and only one source to perform the Bayesian inference.

The elliptic forward operator defined on $\Omega := [0, 1]^2$ reads:

$$(5.5) \quad \begin{aligned} -(u_{xx} + u_{yy})(x, y) &= A(\alpha) \left(\exp(-[(x - \theta_1)^2 + (y - \theta_2)^2]/2\alpha^2) \right. \\ &\quad \left. + b \exp(-[(x - \theta_3)^2 + (y - \theta_4)^2]/2\alpha^2) \right), \quad (x, y) \in \Omega \\ u(x, y) &= 0, \quad (x, y) \in \partial\Omega, \end{aligned}$$

where $A(\alpha) = 5/(2\pi\alpha^2)$, $\alpha = 0.15$ and the binary parameter $b = 1$ when generating the data and $b = 0$ when performing the inference. Therefore, we are solving a source inversion similar to the one in [subsection 5.2](#) but starting from bimodal data.

To generate the data we choose the locations of the two sources far apart, i.e., $(\theta_1, \theta_2)_{\text{true}} = (0.15, 0.15)$ and $(\theta_3, \theta_4)_{\text{true}} = (0.85, 0.85)$. For Bayesian inference we employ StdML, MLLejaStd and MLLejaDV using three levels. The multilevel setup is outlined in [Table 4](#).

We visualize in [Figure 6](#) the bimodal posterior density obtained via standard Bayes' formula [\(2.3\)](#) for which we used $50^2 = 2500$ Gauss-Legendre points to assess the evidence. Note that the two peaks are symmetric around $(0.5, 0.5)$. Therefore,

TABLE 4
Multilevel setup for the 2D inversion problem with forward model (5.5).

Level	h	tol^{in}	τ^{in}	tol^{qu}
$\ell(1)$	$h_1 = \sqrt{2}/2^5$	$tol_3^{\text{in}} = 10^{-6}$	$\tau_3^{\text{in}} = (10^{-8}, 10^{-8})$	$tol_3^{\text{qu}} = 10^{-13}$
$\ell(2)$	$h_2 = \sqrt{2}/2^6$	$tol_2^{\text{in}} = 10^{-5}$	$\tau_2^{\text{in}} = (10^{-7}, 10^{-7})$	$tol_2^{\text{qu}} = 10^{-12}$
$\ell(3)$	$h_3 = \sqrt{2}/2^7$	$tol_1^{\text{in}} = 10^{-4}$	$\tau_1^{\text{in}} = (10^{-6}, 10^{-6})$	$tol_1^{\text{qu}} = 10^{-11}$

Bayesian inference using only once source in the forward model can yield, in the best case, a posterior density centered at $(0.5, 0.5)$.

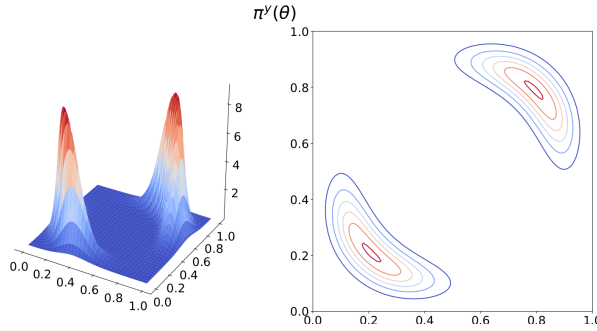


FIG. 6. Posterior density for the source inversion problem with forward model (5.5). We used $50^2 = 2500$ Gauss-Legendre points to assess the evidence.

The QoI is again $\mathbb{E}[\pi^{\mathbf{y}}(\boldsymbol{\theta})]$. In Table 5 we show the obtained estimates. First, a reference Metropolis-Hastings estimate with $2 \cdot 10^5$ samples is computed using a random walk Gaussian proposal with initial sample $\boldsymbol{\theta}_0 = (1.0, 1.0)$ and covariance matrix $C_{\text{MH}} = 10^{-1}I$. The acceptance rate is 45%. We observe that the MH and StdML solutions yield estimates close to the center, $(0.5, 0.5)$. However, the estimates given by MLEjaStd and MLEjaDV are far away from this value.

TABLE 5
Comparison of estimates of $\mathbb{E}[\pi^{\mathbf{y}}(\boldsymbol{\theta})]$ for the source inversion problem with forward model (5.5). We first compute a reference solution using $2 \cdot 10^5$ MH samples. Afterwards, we employ StdML and the two variants of our proposed multilevel approach, MLEjaStd and MLEjaDV.

Method	$\mathbb{E}_{\pi^{\mathbf{y}}}[\boldsymbol{\theta}]$
MH	(0.5032, 0.5068)
StdML	(0.5002, 0.5002)
MLEjaStd	(0.6688, 0.6548)
MLEjaDV	(0.6648, 0.6548)

We depict in Figure 7 the prior and posterior density as well as the weighted (L)-Leja points used to construct the adaptive sparse grid interpolation surrogate of the potential function for all employed multilevel methods. We observe that the Gaussian approximation used in our proposed approach at levels $\ell(2)$ and $\ell(3)$ is very spread and quite different from the approximated posterior. Hence the bias-correcting ratio in quadrature operations, $\pi_{\ell(j)}^{\mathbf{y}}/\widehat{\pi}_{\ell(j)}^{\mathbf{y}}$ for $j = 2, 3$ (recall (3.3)) is different from 1 in most regions of the domain. The large variations of this ratio lead to large variations of the error indicators in adaptive sparse grid quadrature which prevent the algorithm to converge and hence to yield accurate estimates. Therefore, the estimates of the expectation and covariance matrix of the posteriors from levels $\ell(1)$ and $\ell(2)$, and with

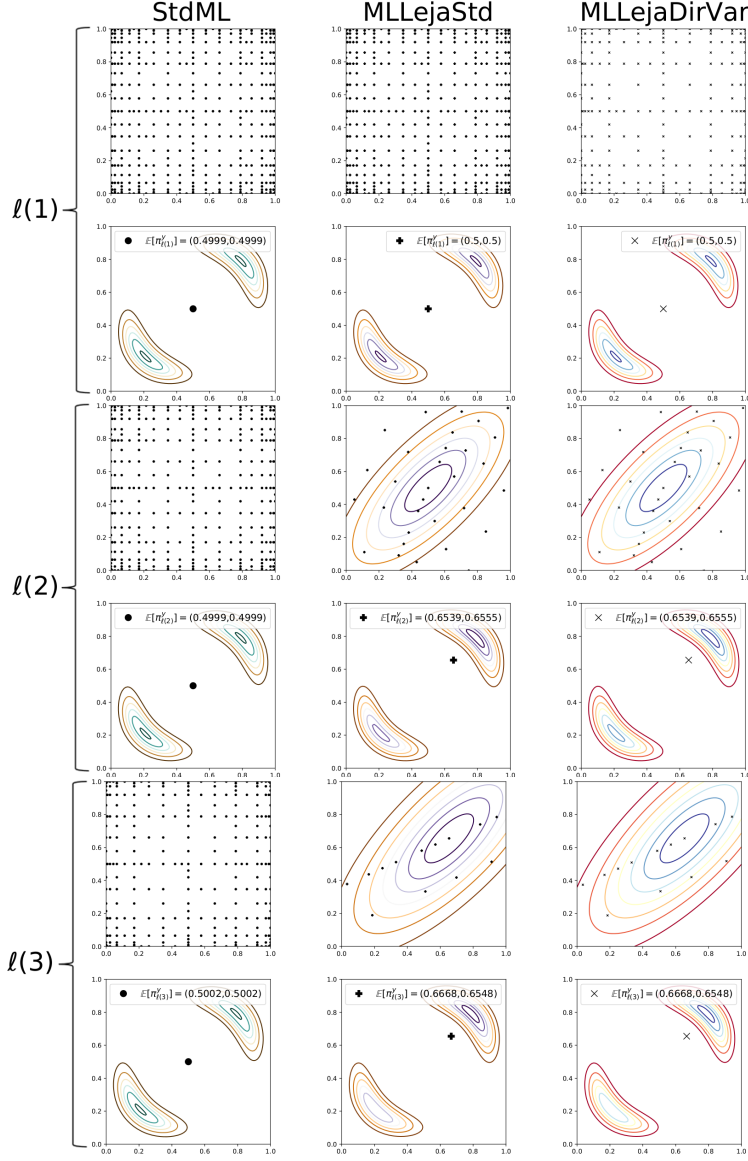


FIG. 7. Results obtained using StdML (left), MLEjaStd (center) and MLEjaDV (right) to find the adaptive sparse grid interpolation surrogate for the potential function in the source inversion problem with forward model (5.5). At each of the three levels, in the top plots we depict the prior density and the corresponding weighted (L)-Leja points used to find the surrogate. At levels $\ell(2)$ and $\ell(3)$, the prior is the Gaussian approximation of the posterior from the posterior level. In the bottom plots we depict the corresponding posterior density solution.

that, the Gaussian approximations employed at levels $\ell(2)$ and $\ell(3)$, are inaccurate. Note that the large spread of the Gaussian approximations leads to weighted (L)-Leja points outside of the domain of the uniform prior, which coincides with the domain of the forward operator, Ω (see (5.5)). Whenever this happens, we impose the corresponding likelihood evaluation to be zero.

5.4. Higher-dimensional problem in a 3D spatial domain. In the final test case, we apply the proposed approach in a more challenging and computationally more expensive problem. We consider an elliptic forward model defined on $\Omega := [0, 1]^3$ with a permeability field projected onto a Fourier basis:

$$(5.6) \quad \begin{aligned} -\nabla \cdot (k(x, y, z, \boldsymbol{\theta}) \nabla u(x, y, z)) &= f, & (x, y, z) &\in \Omega \\ u(x, y, z) &= 0, & (x, y, z) &\in \partial\Omega, \end{aligned}$$

where $f \equiv 5$ and

$$(5.7) \quad k(x, y, z, \boldsymbol{\theta}) := \exp \left(\sum_{n=1}^8 s_n \theta_n \sin(p_{n,1} \pi x) \sin(p_{n,2} \pi y) \sin(p_{n,3} \pi z) \right),$$

where $s_1 = 0.1785$, $s_2 = s_3 = s_4 = 0.1428$, $s_5 = s_6 = s_7 = 0.1071$, $s_8 = 0.0714$ are normalized scaling factors, i.e., $\sum_{n=1}^8 s_n = 1$, and $(p_{n,1}, p_{n,2}, p_{n,3}) \in \{1, 2\}^3$. Note that with the chosen setup, θ_1 is the most important parameter, $\theta_2, \theta_3, \theta_4$ are the second most important parameters etc under the prior density.

Bayesian inference is carried out for the weights $(\theta_1, \theta_2, \dots, \theta_8)$ of the permeability field $k(x, y, z, \boldsymbol{\theta})$. Thus, we are solving an 8D inversion problem. These weights follow a standard normal prior distribution, i.e., $\mu_0 = \mathbf{N}(\mathbf{0}, I)$. To perform the Bayesian inference we employ both the standard and our proposed multilevel approach with the two variants, MLLejaStd and MLLejaDV, considering $J = 3$. The employed multilevel setup is outlined in Table 6. The forward model (5.6) is discretized via standard tetrahedral FE meshes h_j . Moreover, the sparse grid interpolation tolerances tol_j^{in} and τ_j^{in} are chosen to yield quantitatively similar errors to the FE approximation for $j = 1, 2, 3$. Finally, we choose small tolerances for quadrature to prevent the adaptive algorithm from stopping too early.

TABLE 6
Multilevel setup for the 8D inversion problem with forward model (5.6).

Level	h	tol^{in}	τ^{in}	tol^{qu}
$\ell(1)$	$h_1 = \sqrt{3}/2^4$	$tol_3^{\text{in}} = 10^{-5}$	$\tau_3^{\text{in}} = 10^{-7} \cdot \mathbf{1}_8$	$tol_3^{\text{qu}} = 10^{-9}$
$\ell(2)$	$h_2 = \sqrt{3}/2^5$	$tol_2^{\text{in}} = 10^{-4}$	$\tau_2^{\text{in}} = 10^{-6} \cdot \mathbf{1}_8$	$tol_2^{\text{qu}} = 10^{-8}$
$\ell(3)$	$h_3 = \sqrt{3}/2^6$	$tol_1^{\text{in}} = 10^{-3}$	$\tau_1^{\text{in}} = 10^{-5} \cdot \mathbf{1}_8$	$tol_1^{\text{qu}} = 10^{-7}$

The observation data consists of 729 measurements at $\{0.1, 0.2, \dots, 0.9\}^3 \in \Omega$, stemming from the FE solution of the forward model discretized using a finer mesh width $h = \sqrt{3}/2^7$ and assuming measurement noise $\eta \sim \mathbf{N}(\mathbf{0}, 0.1^2 I)$. In addition, $\boldsymbol{\theta}_{\text{true}}$ is drawn from the standard multivariate Gaussian density, i.e.,

$$\boldsymbol{\theta}_{\text{true}} = (0.3015, 0.6578, -0.5002, 0.4608, 1.1345, 0.5447, -1.5353, -0.1689).$$

The QoI is again the expectation of the posterior density, $\mathbb{E}_{\pi_y}[\boldsymbol{\theta}]$. We begin with level $\ell(1)$ in the standard multilevel approach. Both the expectation and the covariance matrix of the corresponding posterior are computed since we need these evaluations at $\ell(2)$. We obtain, however, an indefinite covariance matrix with a negative variance for θ_1 . This is mainly due to the limitations of the standard approach: adaptive sparse grid quadrature w.r.t. the prior density becomes challenging when the complexity of the underlying Bayesian inverse problem increases. To overcome this limitation, we employ instead standard sparse grids of a priori fixed levels having sufficiently many points to guarantee a positive definite covariance matrix. In

particular, we consider an interpolation grid of level 10 for interpolation (24310 grid points) and a quadrature grid of the same level comprising 598417 points. Since our goal is to compare multilevel methods based on adaptive sparse grid algorithms, we do not perform the standard multilevel approach on the remaining two levels using a priori chosen sparse grids, but rather focus only on the two variants of our proposed approach starting from the Gaussian approximation of the posterior at level $\ell(1)$ obtained with the aforementioned standard sparse grids. Moreover, evaluating the FE discretizations depending on h_2 and h_3 on the standard sparse grid of level 10 (24310 evaluations) requires a significant computational cost.

We first compute a reference MH solution with 10^5 samples. The Gaussian proposal has covariance matrix $C_{\text{MH}} = 0.5I$. To reduce the burn-in, the chain is started from $\boldsymbol{\theta}_{\text{true}}$. The acceptance rate is 24%. Afterwards, we employ MLLejaStd and MLLejaDV as described above. The results are showed in Table 7. The two variants of our approach produce results comparable to the reference solution, thus making our proposed approach competitive with sampling methods in this test case as well. Observe, however, that the accuracy of all estimates deteriorates compared with $\boldsymbol{\theta}_{\text{true}}$. This is because the likelihood becomes less informative as the index increases from 1 to 8: the likelihood updates the prior very well for the first direction, relatively well for the next three, and almost not at all for the last four directions. Nevertheless, since the last four directions are the least important by construction (5.7) we expect that having not accurate corresponding mean estimates will not be too significant.

TABLE 7

Estimation of the posterior's mean value for the fourth test case using a referece MH solution with 10^5 samples and the two variants of our proposed multilevel approach for Bayesian inversion.

Method	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_1]$	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_2]$	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_3]$	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_4]$	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_5]$	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_6]$	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_7]$	$\mathbb{E}_{\pi_{\mathbf{y}}}[\theta_8]$
MH	0.2532	0.2123	-0.1363	0.1326	0.1486	0.0753	-0.1584	0.0066
MLStd	0.2642	0.2111	-0.1630	0.1539	0.1429	0.0670	-0.1816	-0.0053
MLDV	0.2620	0.2114	-0.1600	0.1542	0.1448	0.0689	-0.1808	-0.0050

To assess the quality of the expectation estimates, we use them to represent the permeability field as $k(x, y, z, \mathbb{E}_{\pi_{\mathbf{y}}}[\boldsymbol{\theta}])$ and we compare the results with $k(x, y, z, \boldsymbol{\theta}_{\text{true}})$. In Figure 8 we depict 2D slices of the field in which the spatial coordinates x (top), y (middle) and z (bottom) are fixed respectively to 0.5. We observe that having inaccurate estimates for the latter four components of $\boldsymbol{\theta}$ does not significantly affect the estimation of the true permeability field. This is due to having good estimates for the first components of $\boldsymbol{\theta}_{\text{true}}$, which are the most important by construction.

In Figure 9 the costs for interpolation (left) and quadrature (right) are shown. Note that for interpolation we show costs for level $\ell(1)$ as well because evaluations of the forward PDE discretized using h_1 are needed at $\ell(2)$. MLLejaDV is cheaper than MLLejaStd at all three levels, requiring about 4.3 times fewer evaluations at level $\ell(1)$, 4.8 times fewer evaluations on level $\ell(2)$ and 7 times fewer evaluations at level $\ell(3)$. Observe that the overall interpolation costs are very small given that we have an 8D inversion problem at hand. For example, at level $\ell(3)$, MLLejaDV requires only 22 PDE evaluations. The maximum reached level in the corresponding multiindex set is 4 and all its multiindices have components larger than 1 only in the first four directions. Indeed, the directional variances-based algorithm detects that the latter four directions are unimportant, thus invests effort only in the first four directions. Therefore, we see once again that using the enhanced adaptive algorithm for adaptive sparse grid interpolation leads to significant cost savings. For quadrature, the total

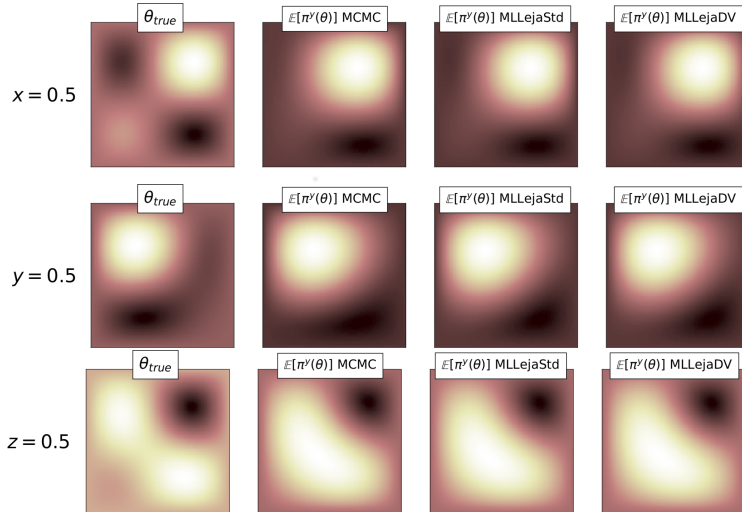


FIG. 8. 2D slices of the 3D permeability field from (5.7) parametrized using θ_{true} and the mean estimates from Table 7. Top: yz slice taking $x = 0.5$. Middle: xz slice when $y = 0.5$. Bottom: xy slice fixing $z = 0.5$.

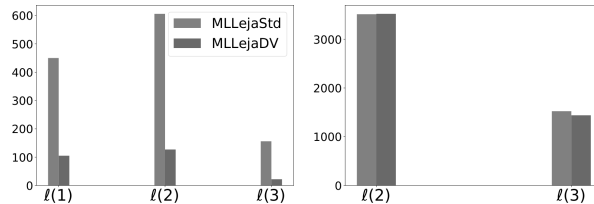


FIG. 9. Left: total number of forward model evaluations needed by the two variants of our proposed approach, MLEjaStd and MLEjaDV, at all three levels in the 8D inversion problem with forward model (5.6). Right: total number of quadrature nodes used in all three in MLEjaStd and MLEjaDV at levels $l(2)$ and $l(3)$.

number of evaluations of the interpolation surrogates are similar.

6. Conclusions. We proposed a novel multilevel Leja algorithm for computing posterior approximations in computationally expensive, higher-dimensional Bayesian inverse problems. At each level adaptive sparse grid interpolation is employed to find a surrogate of the potential function, and adaptive sparse grid quadrature is then used to perform all integration operations with respect to the posterior. We considered two adaptive strategies for interpolation: (i) a standard method and (ii) an enhanced adaptive algorithm in which directional variances are used to ensure that only the most important stochastic directions are refined. The backbone of the proposed approach is the sequential update of the prior density. In this way, we can create weighted (L)-Leja points in areas of high posterior probability. Numerical experiments with elliptic inverse problems in 2D and 3D space show that the sequential update of the prior leads to considerably fewer model evaluations compared to the standard multilevel approach which employs the prior density at all levels. We remark that the proposed approach is not designed to handle well multimodal posterior densities. In future research we will extend it such that it employs more general nonlinear mappings, e.g., transport maps, to accurately approximate arbitrary, multimodal posterior densities.

Acknowledgements. IGF thankfully acknowledges the support of the German Academic Exchange Service (DAAD). IGF, JL, and EU would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the programme *Uncertainty quantification for complex systems: theory and methodologies* when work on this paper was undertaken.

REFERENCES

- [1] S. BROOKS, A. GELMAN, G. L. JONES, AND X.-L. MENG, *Handbook of Markov chain Monte Carlo*, Chapman & Hall/CRC Handbooks of Modern Statistical Methods, CRC Press, Boca Raton, FL, 2011.
- [2] H.-J. BUNGARTZ AND M. GRIEBEL, *Sparse grids*, *Acta Numer.*, 13 (2004), pp. 147–269.
- [3] P. CHEN AND C. SCHWAB, *Sparse-grid, reduced-basis Bayesian inversion*, *Comput. Methods Appl. Mech. Engrg.*, 297 (2015), pp. 84–115.
- [4] P. CHEN AND C. SCHWAB, *Adaptive sparse grid model order reduction for fast Bayesian estimation and inversion*, in *Sparse grids and applications—Stuttgart 2014*, vol. 109 of *Lect. Notes Comput. Sci. Eng.*, Springer, Cham, 2016, pp. 1–27.
- [5] P. CHEN, U. VILLA, AND O. GHATTAS, *Hessian-based adaptive sparse quadrature for infinite-dimensional Bayesian inverse problems*, *Comput. Methods Appl. Mech. Engrg.*, 327 (2017), pp. 147–172.
- [6] P. CONRAD AND Y. MARZOUK, *Adaptive Smolyak pseudospectral approximations*, *SIAM J. Sci. Comput.*, 35 (2013), pp. A2643 – A2670.
- [7] P. DEL MORAL, A. DOUCET, AND A. JASRA, *Sequential Monte Carlo samplers*, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 68 (2006).
- [8] J. DICK, R. N. GANTNER, Q. T. LE GIA, AND C. SCHWAB, *Multilevel higher-order quasi-monte carlo bayesian estimation*, *Math. Models Methods Appl. Sci.*, 27 (2017), pp. 953–995.
- [9] T. DODWELL, C. KETELSEN, R. SCHEICHL, AND A. TECKENTRUP, *A hierarchical multilevel Markov Chain Monte Carlo algorithm with applications to uncertainty quantification in subsurface flow*, *SIAM/ASA J. Uncertain. Quantif.*, 3 (2015), pp. 1075–1108.
- [10] M. EIGEL, M. MARSCHALL, AND R. SCHNEIDER, *Sampling-free Bayesian inversion with adaptive hierarchical tensor representations*, *Inverse Probl.*, 34 (2018), pp. 035010, 29.
- [11] I.-G. FARÇAŞ, T. GÖRLER, H.-J. BUNGARTZ, F. JENKO, AND T. NECKEL, *Sensitivity-driven adaptive sparse stochastic approximations in plasma microinstability analysis*, arXiv e-prints, (2018), arXiv:1812.00080, p. arXiv:1812.00080, <https://arxiv.org/abs/1812.00080>.
- [12] I.-G. FARÇAŞ, P. C. SÂRBU, H.-J. BUNGARTZ, T. NECKEL, AND B. UEKERMANN, *Multilevel adaptive stochastic collocation with dimensionality reduction*, in *Sparse Grids and Applications - Miami 2016*, J. Garcke, D. Pflüger, C. G. Webster, and G. Zhang, eds., Cham, 2018, Springer International Publishing, pp. 43–68.
- [13] M. FRANGOS, Y. MARZOUK, K. WILLCOX, AND B. VAN BLOEMEN WAANDERS, *Surrogate and reduced-order modeling: a comparison of approaches for large-scale statistical inverse problems*, in *Large-scale inverse problems and quantification of uncertainty*, Wiley Ser. Comput. Stat., Wiley, Chichester, 2011, pp. 123–149.
- [14] N. GEORG, D. LOUKREZIS, U. RÖMER, AND S. SCHÖPS, *Uncertainty quantification for an optical grating coupler with an adjoint-based leja adaptive collocation method*, *CoRR*, abs/1807.07485 (2018), <https://arxiv.org/abs/1807.07485>.
- [15] T. GERSTNER AND M. GRIEBEL, *Dimension-adaptive tensor-product quadrature*, *Computing*, 71 (2003), pp. 65–87.
- [16] M. GRIEBEL AND J. OETTERSCHAGEN, *On tensor product approximation of analytic functions*, *J. Complex.*, 207 (2016), pp. 348 – 379.
- [17] M. GRIEBEL, M. SCHNEIDER, AND C. ZENGER, *A combination technique for the solution of sparse grid problems*, in *Iterative Methods in Linear Algebra*, P. de Groen and R. Beauwens, eds., Elsevier, North Holland, 1992, IMACS, pp. 263–281.
- [18] J. HADAMARD, *Sur les problèmes aux dérivées partielles et leur signification physique*, *Princeton University Bulletin*, 13 (1902), pp. 49–52.
- [19] M. HEGLAND, *Adaptive sparse grids*, in *Proc. of 10th Computational Techniques and Applications Conference CTAC-2001*, K. Burrage and R. B. Sidje, eds., vol. 44, 2003, pp. C335–C353.
- [20] L. JIANG AND N. OU, *Bayesian inference using intermediate distribution based on coarse multiscale model for time fractional diffusion equations*, *Multiscale Model. Simul.*, 16 (2018), pp. 327–355.

- [21] J. KAIPIO AND E. SOMERSALO, *Statistical and computational inverse problems*, vol. 160 of Applied Mathematical Sciences, Springer-Verlag, New York, 2005.
- [22] M. C. KENNEDY AND A. O'HAGAN, *Bayesian calibration of computer models*, J. R. Stat. Soc. Ser. B Stat. Methodol., 63 (2001), pp. 425–464.
- [23] J. LATZ, *On the well-posedness of Bayesian inverse problems*, arXiv e-prints, (2019), arXiv:1902.10257, <https://arxiv.org/abs/1902.10257>.
- [24] J. LATZ, M. EISENBERGER, AND E. ULLMANN, *Fast sampling of parameterised Gaussian random fields*, Comput. Methods Appl. Mech. Eng., 348 (2019), pp. 978–1012.
- [25] J. LATZ, I. PAPAIOANNOU, AND E. ULLMANN, *Multilevel sequential² Monte Carlo for Bayesian inverse problems*, J. Comput. Phys., 368 (2018), pp. 154 – 178.
- [26] J. LI AND Y. M. MARZOUK, *Adaptive construction of surrogates for the Bayesian solution of inverse problems*, SIAM J. Sci. Comput., 36 (2014), pp. A1163–A1186.
- [27] C. LIEBERMAN, K. WILLCOX, AND O. GHATTAS, *Parameter and state model reduction for large-scale statistical inverse problems*, SIAM J. Sci. Comput., 32 (2010), pp. 2523–2542.
- [28] X. MA AND N. ZABARAS, *An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method*, Inverse Probl., 25 (2009), p. 035013.
- [29] Y. MARZOUK, T. MOSELHY, M. PARNO, AND A. SPANTINI, *Sampling via Measure Transport: An Introduction*, Springer International Publishing, Cham, 2016, pp. 1–41.
- [30] Y. MARZOUK AND D. XIU, *A stochastic collocation approach to Bayesian inference in inverse problems*, Commun. Comput. Phys., 6 (2009), pp. 826–847.
- [31] S. A. MATTIS AND B. WOHLMUTH, *Goal-oriented adaptive surrogate construction for stochastic inversion*, Comput. Methods Appl. Mech. Engrg., 339 (2018), pp. 36–60.
- [32] A. NARAYAN AND J. D. JAKEMAN, *Adaptive Leja sparse grid constructions for stochastic collocation and high-dimensional approximation*, SIAM J. Sci. Comput., 36 (2014), pp. A2952–A2983.
- [33] B. PEHERSTORFER AND Y. MARZOUK, *A transport-based multifidelity preconditioner for Markov chain Monte Carlo*, ArXiv e-prints, (2018), arXiv:1808.09379, <https://arxiv.org/abs/1808.09379>.
- [34] B. PEHERSTORFER, K. WILLCOX, AND M. GUNZBURGER, *Survey of Multifidelity Methods in Uncertainty Propagation, Inference, and Optimization*, SIAM Rev., 60 (2018), pp. 550–591.
- [35] R. SCHEICHL, A. STUART, AND A. TECKENTRUP, *Quasi-Monte Carlo and multilevel Monte Carlo methods for computing posterior expectations in elliptic inverse problems*, SIAM/ASA J. Uncertain. Quantif., 5 (2017), pp. 493–518.
- [36] C. SCHILLINGS AND C. SCHWAB, *Sparse, adaptive Smolyak quadratures for Bayesian inverse problems*, Inverse Probl., 29 (2013), pp. 065011, 28.
- [37] C. SCHILLINGS AND C. SCHWAB, *Sparsity in Bayesian inversion of parametric operator equations*, Inverse Probl., 30 (2014), pp. 065007, 30.
- [38] C. SCHWAB AND A. M. STUART, *Sparse deterministic approximation of Bayesian inverse problems*, Inverse Probl., 28 (2012), pp. 045003, 32.
- [39] S. SMOLYAK, *Quadrature and interpolation formulas for tensor products of certain classes of functions*, Soviet Math. Dokl., 4 (1963), pp. 240–243.
- [40] I. M. SOBOL', *Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates*, Math. Comput. Simul., 55 (2001), pp. 271 – 280.
- [41] A. STUART, *Inverse problems: a Bayesian perspective*, Acta Numer., 19 (2010), pp. 451–559.
- [42] A. STUART AND A. TECKENTRUP, *Posterior consistency for Gaussian process approximations of Bayesian posterior distributions*, Math. Comp., 87 (2018), pp. 721–753.
- [43] A. TARANTOLA, *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2005.
- [44] R. TEMPONE AND S. WOLFERS, *Smolyak's algorithm: A powerful black box for the acceleration of scientific computations*, in Sparse Grids and Applications - Miami 2016, J. Garcke, D. Pfüger, C. G. Webster, and G. Zhang, eds., Cham, 2018, Springer International Publishing, pp. 201 – 228.
- [45] J. WINOKUR, D. KIM, F. BISETTI, O. P. LE MAÎTRE, AND O. M. KNIO, *Sparse Pseudo Spectral Projection Methods with Directional Adaptation for Uncertainty Quantification*, J. Sci. Comput., 68 (2016), pp. 596–623.
- [46] L. YAN AND Y.-X. ZHANG, *Convergence analysis of surrogate-based methods for Bayesian inverse problems*, Inverse Probl., 33 (2017), pp. 125001, 20.
- [47] L. YAN AND T. ZHOU, *Adaptive multi-fidelity polynomial chaos approach to Bayesian inference in inverse problems*, J. Comput. Phys., 381 (2019), pp. 110–128.