SCALING LIMITS IN COMPUTATIONAL BAYESIAN INVERSION

CLAUDIA SCHILLINGS[1] AND CHRISTOPH SCHWAB[2]

Abstract. Computational Bayesian inversion of operator equations with distributed uncertain input parameters is based on an infinite-dimensional version of Bayes' formula established in M. Dashti and A.M. Stuart [*Handbook of Uncertainty Quantification*, edited by R. Ghanem, D. Higdon and H. Owhadi. Springer (2015).] and its numerical realization in C. Schillings and Ch. Schwab [*Inverse Problems* **29** (2013) 065011; *Inverse Problems* **30** (2014) 065007.] Based on the sparsity of the posterior density shown in C. Schillings and Ch. Schwab [*Inverse Problems* **29** (2013) 065011; *Inverse Problems* **30** (2014) 065007.]; C. Schwab and A.M. Stuart [*Inverse Problems* **28** (2012) 045003.], dimension-adaptive Smolyak quadratures can afford higher convergence rates than MCMC in terms of the number M of solutions of the forward (parametric operator) equation in C. Schillings and Ch. Schwab [*Inverse Problems* **29** (2013) 065011; *Inverse Problems* **30** (2014) 065007.]. The error bounds and convergence rates obtained in C. Schillings and Ch. Schwab [*Inverse Problems* **29** (2013) 065011; *Inverse Problems* **30** (2014) 065007.] are independent of the parameter dimension (in particular free from the curse of dimensionality) but depend on the (co)variance $\Gamma > 0$ of the additive, Gaussian observation noise as $\exp(b\Gamma^{-1})$ for some constant $b > 0$. It is proved that the Bayesian estimates admit asymptotic expansions as $\Gamma \downarrow 0$. Sufficient (nondegeneracy) conditions for the existence of finite limits as $\Gamma \downarrow 0$ are presented. For Gaussian priors, these limits are shown to be related to MAP estimators obtained from Tikhonov regularized least-squares functionals. Quasi-Newton (QN) methods with symmetric rank-1 updates are shown to identify the concentration points in a non-intrusive way, and to obtain second order information of the posterior density at these points. Based on the theory, two novel computational Bayesian estimation algorithms for Bayesian estimation at small observation noise covariance $\Gamma > 0$ with performance independent of $\Gamma \downarrow 0$ are proposed: first, dimension-adaptive Smolyak quadrature from C. Schillings and Ch. Schwab [*Inverse Problems* **29** (2013) 065011; *Inverse Problems* **30** (2014) 065007.] combined with a reparametrization of the parametric Bayesian posterior density near the MAP point (assumed unique) and, second, generalized Richardson extrapolation to the limit of vanishing observation noise variance. Numerical experiments are presented which confirm Γ-independent convergence of the curvature-rescaled, adaptive Smolyak algorithm. Dimension truncation of the posterior density is justified by a general compactness result for the posterior's Hessian at the MAP point.

Mathematics Subject Classification. 65M32, 65N35, 65K10.

Received October 16, 2014. Revised July 29, 2015. Accepted January 20, 2016.

[1] Mathematics Institute, University of Warwick, Coventry CV4 7AL, England. `c.schillings@warwick.ac.uk`

[2] Seminar for Applied Mathematics, ETH, CH-8092 Zurich, Switzerland. `christoph.schwab@sam.math.ethz.ch`

1. INTRODUCTION

Efficient computational Bayesian inversion of partial differential equations has attracted considerable attention in recent years, in statistics in the context of "big data", and in computational uncertainty quantification, as documented in [13, 23, 26] and the references therein. While, historically, numerical analysis focused on finite-dimensional, parametric problems, recent years have seen a heightened interest in numerical treatment of partial differential equations (PDEs) with "distributed uncertainty", such as random field inputs from an infinite-dimensional space X. Upon choosing an (unconditional) basis of the space X the Bayesian inversion and estimation formally become infinite-dimensional, parametric deterministic *quadrature* problems, with quadrature understood with respect to the Bayesian posterior measure. Under appropriate sparsity and smoothness of the Bayesian posterior density, deterministic, dimension-adaptive quadrature approaches have been shown to achieve higher convergence rates than the widely used MCMC methods; we refer to [29,30] and the references therein. It is well-known, however, that the Bayesian posterior exhibits concentration effects for small observation noise covariance Γ; in high-dimensional parameter spaces, which commonly arise in problems with uncertain, distributed input data, most contributions to the Bayesian estimate therefore stem from a "small" subset of the parameter space. This subset is, generally, data dependent, and its efficient computational localization, *e.g.* during the "burn-in" of MCMC samplers, is key to efficient computational Bayesian inversion.

To develop efficient numerical treatment of a class of concentrating posterior densities in the sparse, dimension-adaptive deterministic quadrature methods from [29, 30] is the purpose of the present paper. We propose to numerically identify so-called MAP points of posterior concentration in parameter spaces by Quasi-Newton methods with symmetric rank-1 (SR1) update, [10] applied to the (co)variance-weighted Bayesian misfit functional. These methods afford locally *superlinear convergence of iterates and Hessians*. We propose to use the second order information on the Bayesian posterior density at the MAP point for posterior "desingularization" via *curvature-based, affine reparametrization* of the posterior density. We prove that this reparametrization renders the dimension-adaptive Smolyak quadratures from [29,30] robust with respect to the observation noise variance Γ. We also obtain asymptotic expansions of the Bayesian estimate with respect to observation noise variance Γ which generalize the results in [24] for the linear Gaussian case to nonlinear forward problems.

The outline of this paper is as follows: in Section 2 we present (nonparametric) Bayesian estimation problems on function space for operator equations with Lipschitz dependence on distributed uncertainty $u \in X$, a separable Banach space admitting an unconditional basis. To this end, we recapitulate the formalism of [29,30] whereby the Bayesian estimate can be expressed as a formally infinite-dimensional integral wr. to the Bayesian posterior density. Section 3 addresses the case when the observation noise (co)variance $\Gamma \downarrow 0$. Under nondegeneracy assumptions on the (co)variance-weighted LSQ functions, the Bayesian estimate admits a finite (possibly weak, *cf.* [14], Sect. 3) limit which equals the QoI evaluated at the point $\boldsymbol{y}_0$ in the (possibly infinite-dimensional) parameter domain. Section 3 presents an asymptotic analysis which reveals, in particular, the need for the numerical solution of a nonlinear, high-dimensional (co)variance-weighted least-squares problem to determine the $\Gamma = 0$ limit as well as its asymptotic expansion with respect to Γ. Section 4 uses the structure of the leading terms in the asymptotic expansion of the Bayesian estimate to develop a curvature-rescaling coordinate transformation at the MAP point. We prove that under some nondegeneracy assumption, our approach removes the posterior concentration as $\Gamma \downarrow 0$. The asymptotic expansions obtained in Section 3 also justify a new (generalized) "extrapolation to the limit" approach developed in Section 4.3, which can be combined with curvature rescaling. The required second order information at the MAP point $\boldsymbol{y}_0$ can be obtained "for free" by Quasi-Newton methods with SR1 updates (we refer to [31], Appendix C or to [10, 18]). The asymptotic expansion of the Bayesian estimate w.r. to small covariances justifies the use of generalized Richardson extrapolation to small observation (co)variance Γ, and its stable, algorithmic realization by the T-transformation based on [34]. In Section 5 numerical experiments are presented which confirm the theoretical results. Appendix A collects known results on Laplace's method for the asymptotic analysis of integrals depending on a parameter, and Appendix B establishes the compactness of the Hessian, for bounded parameter ranges and uniform prior π_0.

2. Bayesian inversion of operator equations

We consider a class of Bayesian inverse problems for partial differential equations (PDES) with "distributed" uncertain input data u taking values in infinite-dimensional spaces, in the setting outlined in [13].

2.1. Bayesian inversion in infinite dimension

By $G : X \to \mathcal{X}$ we denote a "forward" response map from the separable Banach space X of uncertain, distributed parameters u into some state (reflexive Banach) space $\mathcal{X}$ of responses. We equip X and $\mathcal{X}$ with norms $\| \cdot \|_X$ and with $\| \cdot \|_{\mathcal{X}}$, respectively. We think of the forward response map as a solution map of an operator equation acting on a state $q \in \mathcal{X}$ and taking values in the dual of a reflexive Banach space $\mathcal{Y}$, i.e. we consider an operator equation of the following form:

$$\text{Given } u \in X, \ f \in \mathcal{Y}', \text{ find } q \in \mathcal{X}: \quad A(u; q) = f \tag{2.1}$$

where, for each instance $u \in X$, the uncertain operator $A(u; \cdot) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ is assumed to be boundedly invertible, at least locally sufficiently close to a nominal input $u_0 \in X$, i.e. for $\|u - u_0\|_X$ small enough. Then, for any instance of the uncertainty u and for known forcing $f \in \mathcal{Y}'$, the response q, i.e. the solution $q \in \mathcal{X}$ of the forward problem (2.1) is the image of a map $G : X \times \mathcal{Y}'$ to $\mathcal{X}$, i.e.

$$q = G(u; f) \in \mathcal{X}.$$

We omit the dependence of the response on f and simply write $q = q(u) = G(u)$ for the uncertainty-to-solution map $G(\cdot)$. We also assume given K bounded, linear *observation functionals* $\mathcal{O}(\cdot) = (o_1, \ldots, o_K) : \mathcal{X} \to Y = \mathbb{R}^K$. Specifically, we shall assume that $\mathcal{O}(\cdot)$ comprises K *bounded linear observation operators* on the space $\mathcal{X}$ of system responses, i.e. $\mathcal{O} \in (\mathcal{X}')^K$, the dual space of the space $\mathcal{X}$ of system responses. We assume that there are a finite number K of observables, taking values in the data space $Y = \mathbb{R}^K$ equipped with the Euclidean norm, denoted by $\| \cdot \|_2$. The data δ is assumed to consist of observations of QoI system responses corrupted by additive, centered Gaussian noise, i.e.

$$\delta = \mathcal{O}(G(u)) + \eta \in Y \tag{2.2}$$

where $\eta \in Y = \mathbb{R}^K$ is Gaussian observation noise and where the observation functional is $\mathcal{O}(\cdot) = (o_k(\cdot))_{k=1}^K \in (\mathcal{X}')^K$. In the present paper, we assume that the noise process η is Gaussian, i.e. a random vector $\eta \sim \mathcal{N}(0, \Gamma)$, for a positive definite covariance operator Γ on $\mathbb{R}^K$ (i.e., a symmetric, positive definite $K \times K$ covariance matrix Γ) which we assume to be known. Then, the *uncertainty-to-observation map* $\mathcal{G} = \mathcal{O} \circ G : X \to Y = \mathbb{R}^K$ reads

$$\delta = \mathcal{G}(u) + \eta = (\mathcal{O} \circ G)(u) + \eta : X \mapsto L_\Gamma^2(Y; g_\Gamma) \tag{2.3}$$

where $L_\Gamma^2(Y; g_\Gamma)$ denotes random vectors taking values in $Y = \mathbb{R}^K$ which are square integrable with respect to the centered Gaussian measure g_Γ on Y with positive definite covariance matrix $\Gamma > 0$. In view of Bayes' formula (e.g. [13], Thm. 3.3), we define the least-squares functional (also referred to as "potential" in [13], or as "mismatch" resp. "misfit" functional in the literature, e.g. [3]) $\Phi_\Gamma : X \times Y \to \mathbb{R}$ by $\Phi_\Gamma(u; \delta) = \frac{1}{2} R(u)^\top \Gamma^{-1} R(u)$, with the residual R at data δ and uncertainty $u \in X$ given by

$$R(u) := \mathcal{G}(u) - \delta = (\mathcal{O} \circ G)(u) - \delta.$$

For observation noise (co)variance $\Gamma > 0$, the Bayesian potential $\Phi_\Gamma(u; \delta) : X \times Y \to \mathbb{R}$ is a covariance-weighted model-data misfit least squares function, given by

$$\Phi_\Gamma(u; \delta) = \frac{1}{2} R(u)^\top \Gamma^{-1} R(u) = \frac{1}{2} \left((\delta - (\mathcal{O} \circ G)(u))^\top \Gamma^{-1} (\delta - (\mathcal{O} \circ G)(u)) \right). \tag{2.4}$$

In ([13], Thm. 3.4), an infinite-dimensional version of Bayes' rule is shown to hold in the present setting. It states that, under appropriate continuity conditions on the uncertainty-to-observation map

$\mathcal{G} = (\mathcal{O} \circ G)(\cdot) : X \mapsto Y = \mathbb{R}^K$ and on the prior measure π_0 on the space X of uncertain parameters u, the posterior distribution π^δ is absolutely continuous with respect to the prior π_0. While MCMC methods sample from π^δ, the deterministic quadrature approach from [29, 30] is based on parametrizing the (bounded) posterior density $\Theta_\Gamma = \frac{\mathrm{d}\pi^\delta}{\mathrm{d}\pi_0}$ and to use Θ_Γ to integrate adaptively against the posterior π^δ; *cf.* Proposition 2.3 ahead.

2.2. Uncertainty parametrization

We parametrize the uncertain datum u in the forward equation (2.1). In parametric statistical estimation, u is a (low-dimensional) vector containing a few unknown parameters $(y_j)_{j \in \mathbb{J}}$, for a finite index set $\mathbb{J} = \{1, 2, \ldots, J\}$ with small cardinality J so that $X \simeq \mathbb{R}^J$. In the present context of PDEs, $u \in X$, an infinite-dimensional, separable Banach space is of interest in which case $\mathbb{J} = \mathbb{N}$. We assume that there exists a Schauder basis $\{\psi_j\}_{j \in \mathbb{J}}$ of X such that, for some "nominal" value $\langle u \rangle \in X$ of the uncertain datum u, and for some coefficient sequence $\boldsymbol{y} = (y_j)_{j \in \mathbb{J}}$ (uniquely associated with $u - \langle u \rangle \in X$) the uncertainty u is parametrized by the sequence $\boldsymbol{y}$ in the sense that there holds

$$u = u(\boldsymbol{y}) := \langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j \in X \tag{2.5}$$

with unconditional convergence. We refer to $u - \langle u \rangle$ as "fluctuation" of u about the nominal value $\langle u \rangle \in X$. So far, the parametrization (2.5) is deterministic. In the case of a uniform prior, in order to place (2.2), (2.5) into the (probabilistic) Bayesian setting of [13], we introduce (after possibly rescaling the fluctuations) a "reference" parameter domain $U = [-1, 1]^{\mathbb{J}} = \prod_{j \in \mathbb{J}}[-1, 1]$, and equip this cartesian product of sets with the product sigma-algebra $\mathcal{B} = \bigotimes_{j \in \mathbb{J}} \mathcal{B}^1$, with $\mathcal{B}^1$ denoting the sigma-algebra of Borel sets on $[-1, 1]$. On the measurable space $(U, \mathcal{B})$ thus obtained, we introduce a probability measure π_0 (which will serve a Bayesian prior in what follows), and which we shall choose as $\pi_0 = \bigotimes_{j \in \mathbb{J}} \frac{1}{2}\lambda^1$ with λ^1 denoting the Lebesgue measure on $[-1, 1]$. Then $(U, \mathcal{B}, \pi_0)$ becomes (as countable product of probability spaces) a probability space on the set U of all sequences of coefficient vectors $\boldsymbol{y}$ in the uncertainty parametrization (2.5). The uncertain datum u in (2.5) becomes a random field, with π_0 charging the possible realizations of u. As indicated in [8, 29, 33], *analyticity of uncertainty parametrization* (2.5) *with respect to the parameter sequence* $\boldsymbol{y}$ can be used to derive sparsity results for this posterior. In the case of a Gaussian prior, a parametrization of the form (2.5) of the Gaussian random field can be obtained *via* a Karhunen−Loève expansion, where $(y_j)_{j \in \mathbb{J}}$ is an iid. sequence of $\mathcal{N}(0, 1)$ random variables. We refer to ([35], Chap. 6) and the references therein for details.

Throughout the remainder of this paper, we will assume that the uncertain input u in the forward problem (2.1) *is parametrized as in* (2.5). We write $R(\boldsymbol{y})$ and $\mathcal{G}(\boldsymbol{y})$ in place of $R(u(\boldsymbol{y}))$ and of $\mathcal{G}(u(\boldsymbol{y}))$, respectively.

2.3. Forward models

We recapitulate classes of abstract, countably-parametric operator equations considered in [29, 30]. Throughout, we denote by $\mathcal{X}$ and $\mathcal{Y}$ two separable and reflexive Banach spaces over $\mathbb{R}$ (for some of the technical arguments which follow, we shall require also extensions of these spaces to Banach spaces over the coefficient field $\mathbb{C}$; we shall use these without distinguishing these extensions notationally) with (topological) duals $\mathcal{X}'$ and $\mathcal{Y}'$, respectively. By $\mathcal{L}(\mathcal{X}, \mathcal{Y}')$, we denote the set of bounded linear operators $A : \mathcal{X} \to \mathcal{Y}'$. *Via* the Riesz representation theorem, we associate to each $A \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ in a one-to-one correspondence a bilinear form (with $_\mathcal{Y}\langle \cdot, \cdot \rangle_{\mathcal{Y}'}$ denoting the $\mathcal{Y} \times \mathcal{Y}'$-duality pairing) *via* $\mathfrak{a}(v, w) := {}_\mathcal{Y}\langle w, Av \rangle_{\mathcal{Y}'}$ for all $v \in \mathcal{X}, w \in \mathcal{Y}$.

2.3.1. Affine-parametric operator equations. Uniform prior π_0

The assumption on affine parametrization of the distributed system uncertainty by the sequence $\boldsymbol{y} = (y_j)_{j \in \mathbb{J}} \in U$ of (possibly countably many) parameters results in a parametric operator equation of the form

$$A(\boldsymbol{y}) = A_0 + \sum_{j \in \mathbb{J}} y_j A_j \in \mathcal{L}(\mathcal{X}, \mathcal{Y}'). \tag{2.6}$$

Here, either $\mathbb{J} = \{1, 2, \ldots, J\}$ for some $J < \infty$ or $\mathbb{J} = \mathbb{N}$. In the latter case, the forward models admit dimension truncations with error bounds which are addressed in Section 2.5 ahead.

In (2.6), $\boldsymbol{y} = (y_j)_{j \in \mathbb{J}}$ can be, for example, an iid. sequence of real-valued random variables $y_j \sim \mathcal{U}(-1, 1)$, A_0 is a "nominal operator" (representing the non-perturbed system) and $\{A_j\}_{j \in \mathbb{J}} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ denotes a sequence of "fluctuations" about the "nominal operator" $A_0 = A(0)$. We impose the following assumptions on the sequence $\{A_j\}_{j \geq 0} \subset \mathcal{L}(\mathcal{X}, \mathcal{Y}')$.

Assumption 2.1. The operator family $\{A_j\}_{j \geq 0} \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ in (2.6) satisfies:

(1) The "nominal" or "mean field" operator $A_0 \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ is boundedly invertible.
(2) The "fluctuation" operators $\{A_j\}_{j \geq 1}$ are small relative to A_0 in the following sense: there exists a constant $0 < \kappa < 1$ such that

$$\sum_{j \in \mathbb{J}} b_j \leq \kappa < 1, \quad \text{where} \quad b_j := \|A_0^{-1} A_j\|_{\mathcal{L}(\mathcal{X}, \mathcal{X})}, \tag{2.7}$$

(3) (p summability) For some $0 < p < 1$, the operators $B_j = A_0^{-1} A_j$ are p-summable, in the sense that with the sequence $b = (b_j)_{j \in \mathbb{J}}$ as in (2.7) holds

$$\|b\|_{\ell^p(\mathbb{J})}^p = \sum_{j \in \mathbb{J}} b_j^p < \infty. \tag{2.8}$$

Condition (2.7) (and, hence, Assumption 2.1) is sufficient for the bounded invertibility of $A(\boldsymbol{y})$, uniformly with respect to the parameter sequence $\boldsymbol{y} \in U = [-1, 1]^{\mathbb{J}}$. The next result from [30] makes this precise.

Theorem 2.2. *Under Assumption 2.1, for every realization $\boldsymbol{y} \in U$ of the parameters, the affine parametric operator family $A(\boldsymbol{y})$ is boundedly invertible, uniformly with respect to the parameter sequence $\boldsymbol{y} \in U$: for every $f \in \mathcal{Y}'$ and for every $\boldsymbol{y} \in U$, the parametric operator equation*

$$\text{find } q(\boldsymbol{y}) \in \mathcal{X} : \quad \mathfrak{a}(\boldsymbol{y}; q(\boldsymbol{y}), v) = \langle f, v \rangle_{\mathcal{Y}' \times \mathcal{Y}} \quad \forall v \in \mathcal{Y}$$

admits a unique solution $q(\boldsymbol{y}) = (A(\boldsymbol{y}))^{-1} f$ which is uniformly bounded over U, i.e.

$$\sup_{\boldsymbol{y} \in U} \|q(\boldsymbol{y})\|_{\mathcal{X}} \leq \frac{\|f\|_{\mathcal{Y}'}}{\mu}.$$

In the case that the observation functional $\mathcal{O} : \mathcal{X} \to Y = \mathbb{R}^K$ comprises K continuous, linear functionals $o_k \in \mathcal{X}'$, $k = 1, \ldots, K$,

$$\forall \boldsymbol{y} \in U : \quad \|\mathcal{G}(\boldsymbol{y})\|_2 = \|\mathcal{O}(q(\boldsymbol{y}))\|_2 \leq \frac{\|f\|_{\mathcal{Y}'}}{\mu} \left(\sum_{k=1}^K \|o_k\|_{\mathcal{X}'}^2 \right)^{\frac{1}{2}}.$$

The forward maps $q : U \to \mathcal{X}$ and $\mathcal{G} : U \to \mathbb{R}^K$ are globally Lipschitz and admit analytic continuations wr. to the parameters y_j into the complex domain. Specifically (see [33], Lem. 3.3 and Thm. 3.4) if q and $\tilde{q}$ are solutions of (2.1) with the same right hand side f with operators $A(\boldsymbol{y})$ and $A(\boldsymbol{y}')$, respectively, then the forward solution map $\boldsymbol{y} \to q(\boldsymbol{y}) = (A(\boldsymbol{y}))^{-1} f$ is Lipschitz as a mapping from U into $\mathcal{X}$, i.e. there exists a constant $C > 0$ (depending only on κ and π_0 in Assumption 2.1) such that for every $\boldsymbol{y}, \tilde{\boldsymbol{y}} \in U$ holds

$$\|q(\boldsymbol{y}) - q(\tilde{\boldsymbol{y}})\|_{\mathcal{X}} \leq C \|\boldsymbol{y} - \tilde{\boldsymbol{y}}\|_{\ell^\infty} \|f\|_{\mathcal{Y}'}.$$

Moreover, the uncertainty-to-observation map $U \ni \boldsymbol{y} \to \mathcal{G}(\boldsymbol{y}) := (\mathcal{O} \circ q)(\boldsymbol{y})$ is globally Lipschitz as a mapping from $\ell^\infty(\mathbb{N})$ into $Y = \mathbb{R}^K$, in the sense that

$$\|\mathcal{G}(\boldsymbol{y}) - \mathcal{G}(\tilde{\boldsymbol{y}})\|_2 \leq C \left(\sum_{k=1}^K \|o_k\|_{\mathcal{X}'}^2 \right)^{\frac{1}{2}} \|\boldsymbol{y} - \tilde{\boldsymbol{y}}\|_{\ell^\infty(\mathbb{J})} \|f\|_{\mathcal{Y}'}.$$

The prototypical example for (2.6) is the linear, elliptic diffusion problem

$$-\nabla \cdot (u(x, \boldsymbol{y})\nabla q) = f \quad \text{in} \quad D, \quad q(\cdot, \boldsymbol{y})|_{\partial D} = 0. \tag{2.9}$$

Here, $D \subset \mathbb{R}^d$ is a bounded Lipschitz domain, $f \in L^2(D)$ a known, deterministic source term, and the uncertain diffusion coefficient u is given by the affine-parametric function

$$u(x, \boldsymbol{y}) := \langle u \rangle(x) + \sum_{j \geq 1} y_j \psi_j(x) \tag{2.10}$$

where $\langle u \rangle, \psi_j \in C^{0,\alpha}(\overline{D})$ for some $0 < \alpha < 1$, and $|y_j| \leq 1$ so that $\boldsymbol{y} = (y_j)_{j \geq 1} \in U$. Convergence of (2.10) for $\boldsymbol{y} \in U$ is ensured by assuming that for some $0 < p \leq 1$

$$\boldsymbol{b} = (b_j)_{j \geq 1} \in \ell^p(\mathbb{N}), \quad b_j := \|\psi_j\|_{C^{0,\alpha}(\overline{D})},$$

i.e. we have (2.5) with $X = C^{0,\alpha}(\overline{D})$. Uniform (with respect to the parameter vector $\boldsymbol{y}$) inf-sup conditions are implied by uniform ellipticity of (2.9). This, in turn, is ensured by imposing that there exist constants $\mu_0 > 0$ and $0 < \kappa < 1$ such that

$$0 < \mu_0 \leq \text{ess} \inf_{x \in D} \langle u \rangle(x), \qquad \kappa := \|\boldsymbol{b}\|_{\ell^1(\mathbb{N})} < 1. \tag{2.11}$$

For possibly nonlinear operator equations admissible in the present theory see [7, 8, 21, 22, 32].

2.3.2. *Lognormal diffusion models. Gaussian Prior π_0.*

Gaussian priors arise, for example, in UQ for subsurface flow problems, with unknown permeability coefficient u, $\mathbb{J} = \mathbb{N}$ and $U = \mathbb{R}^{\mathbb{J}}$, (*cf. e.g.* [5], [13], Sect. 3.4) In a bounded Lipschitz domain $D \subset \mathbb{R}^d$, we consider once more the diffusion problem (2.9), with

$$u(x, y) = \exp\left(\langle u \rangle(x) + \sum_{j \geq 1} y_j \psi_j(x)\right), \tag{2.12}$$

with $\langle u \rangle(x) + \sum_{j \geq 1} y_j \psi_j(x)$ an isotropic, Gaussian random field in D as in (2.10). Then, (2.11) is *not* required. The prior π_0 is the normalized, centered Gaussian measure $\pi_0 = \mathcal{N}(0, B^{-\theta})$ on $U = \mathbb{R}^{\mathbb{N}}$. If $B = -\Delta$ denotes the Dirichlet Laplacian on D, the covariance operator $\Gamma = B^{-\theta}$ is a trace class operator on the Hilbert space $H = L^2(D)$ if $\theta > d/2$. If $\mathbb{J} = \{1, \ldots, J\}$ and $U = \mathbb{R}^J$, *i.e.* if the parameter space dimension J is finite, $\theta = 0$ is admissible (for finite J, $\Gamma = I$ is trace-class, and the prior $\pi_0 = \mathcal{N}(0, I)$ on $U = \mathbb{R}^J$ is admissible). The case of infinite dimensional observational data will not be investigated in this work and we refer to [14] for more details in this setting. Furthermore, Bayesian posterior consistency for the MAP estimator, which can be defined as the minimizer of an Onsager–Machlup functional on the Cameron–Martin space of the prior in the infinite dimensional case, is established in [14]. In both the small noise limit and large sample size limit, the MAP estimator for Gaussian prior π_0 is shown in [14] to concentrate on the truth.

2.4. Parametric Bayesian posterior density

Motivated by [29, 33], the basis for the presently proposed, adaptive deterministic quadrature approaches for Bayesian estimation *via* the computational realization of Bayes' formula is a *parametric, deterministic representation* of the derivative of the posterior measure with respect to the uniform prior measure π_0. The prior measure π_0 in Section 2.3.1 being uniform, for this prior we admit in (2.5) sequences $\boldsymbol{y}$ which take values in the parameter domain $U = [-1, 1]^{\mathbb{J}}$. In the lognormal case in Section 2.3.2, we admit as priors $\pi_0 \sim \mathcal{N}(0, C)$ Gaussian measures on $U = \mathbb{R}^{\mathbb{N}}$ with trace class covariance operator C. As explained in Section 2.2, this leads

to parametric, deterministic forward problems in the probability space $(U, \mathcal{B}, \pi_0)$. With the parameter domain U, the parametric forward map $\Xi : U \to \mathbb{R}^K$ is given by

$$\Xi(\boldsymbol{y}) = \mathcal{G}(u)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j}.$$

The mathematical foundation of Bayesian inversion is Bayes' theorem. It addresses the structure of the mathematical expectation of the QoI ϕ, over all realizations of the uncertain datum u which are distributed according to the prior π_0, given data δ. We present a version of it, from [13] and, in its parametric version, from [33].

Proposition 2.3. *Assume that $\Xi : U \to \mathbb{R}^K$ is continuous, $\pi_0(U) = 1$ and*

$$\int_U \exp\big(-\Phi_\Gamma(u; \delta)\big)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} \pi_0(\mathrm{d}\boldsymbol{y}) > 0.$$

Then $\pi^\delta(\mathrm{d}\boldsymbol{y})$, the distribution of $y \in U$ given δ, is absolutely continuous with respect to $\pi_0(\mathrm{d}\boldsymbol{y})$, i.e.

$$\frac{\mathrm{d}\pi^\delta}{\mathrm{d}\pi_0}(\boldsymbol{y}) = \frac{1}{Z_\Gamma}\Theta_\Gamma(\boldsymbol{y}), \quad \Theta_\Gamma(\boldsymbol{y}) := \exp\big(-\Phi_\Gamma(u; \delta)\big)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j}, \tag{2.13}$$

with the parametric Bayesian posterior $\Theta_\Gamma(\boldsymbol{y})$ and with the Bayesian potential Φ_Γ defined in (2.4). The normalization constant Z_Γ in (2.13) is given by

$$Z_\Gamma = \mathbb{E}^{\pi^\delta}[1] = \int_U \Theta_\Gamma(\boldsymbol{y})\pi_0(\mathrm{d}\boldsymbol{y}). \tag{2.14}$$

Given (noisy) observation data δ, computational Bayesian inversion is concerned with computational approximation of a "most likely" *system response* $\phi : X \to \mathcal{S}$ of a QoI ϕ which may take values in a Banach space $\mathcal{S}$ (possibly distinct from the space Y of observations). With the QoI ϕ we associate the parametric map

$$\Psi_\Gamma(\boldsymbol{y}) = \Theta_\Gamma(\boldsymbol{y})\phi(u) \big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} = \exp\big(-\Phi_\Gamma(u; \delta)\big)\phi(u)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} : U \to \mathcal{S}. \tag{2.15}$$

Then the Bayesian estimate of the QoI ϕ, given noisy data δ, takes the form

$$\mathbb{E}^{\pi^\delta}[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} = \frac{1}{Z_\Gamma}\int_{\boldsymbol{y} \in U} \Psi_\Gamma(\boldsymbol{y})\pi_0(\mathrm{d}\boldsymbol{y}) = \frac{1}{Z_\Gamma}\int_{\boldsymbol{y} \in U} \exp\big(-\Phi_\Gamma(u; \delta)\big)\phi(u)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} \pi_0(\mathrm{d}\boldsymbol{y}) \tag{2.16}$$

where we introduced the integral

$$Z'_\Gamma := \int_{\boldsymbol{y} \in U} \Psi_\Gamma(\boldsymbol{y})\pi_0(\mathrm{d}\boldsymbol{y}) = \int_{\boldsymbol{y} \in U} \exp\big(-\Phi_\Gamma(u; \delta)\big)\phi(u)\Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} \pi_0(\mathrm{d}\boldsymbol{y}).$$

Based on (2.16), in [29, 30] we approximated Z'_Γ and Z_Γ which, in the parametrization with respect to $\boldsymbol{y} \in U$, take the form of infinite-dimensional integrals with respect to the prior $\pi_0(\mathrm{d}\boldsymbol{y})$. In [29, 30], we proposed the use of dimension adaptive Smolyak quadrature to the numerical evaluation of the integrals Z_Γ and Z'_Γ. The definitions (2.14) and (2.16) imply that $0 < \Gamma \ll 1$ entails concentration of the posterior density Θ_Γ in (2.13).

The Bayesian approach to inverse problems of the form (2.2) as well as the adaptive quadrature approach proposed in [29, 30] are both well defined in the infinite dimensional setting, *i.e.* in the case that the parameter space X is infinite dimensional. However, the analysis based on Laplace's method presented below, which will be used to quantify the concentration effect of the posterior, requires the truncation to a finite number of parameters. Due to the ill posedness inherent in the underlying parameter identification problem, it is well known, and observed in many numerical experiments, that the data informs only a lower dimensional (finite) subspace of the parameter space resulting in a small effective dimension of the problem. This observed effect is theoretically underpinned by the compactness result of the Hessian of the uncertainty-to-observation map (*cf.* Prop. B.1).

2.5. Dimension truncation

The uncertainty parametrization (2.5) renders the forward map $A(u;q)$ parametric, which we express by writing $A(\boldsymbol{y};q) := (A(u;q))|_{u=u(\boldsymbol{y})}$. We consider now the case $J = \infty$, *i.e.*, that $\mathbb{J} = \mathbb{N}$, so that $A(\boldsymbol{y};q)$ depends on the sequence $\boldsymbol{y} = (y_j)_{j\in\mathbb{N}}$ of countably many variables. The asymptotic analysis based on Laplace's method requires the *dimensional truncation* of the forward map to a finite number $J < \#(\mathbb{J})$ of parameters, *i.e.* we replace $u \in X$ in (2.5) by its J-term truncation

$$u^{(J)} = \langle u \rangle + \sum_{j=1}^{J} y_j \psi_j \in X. \tag{2.17}$$

We assume that these truncations converge towards $u \in X$ in the norm of X, for every $u \in X$, at rate $s > 0$: there exists $C(s) > 0$ such that for all J and for every admissible uncertainty $u \in X_s \subset X$ there holds

$$\forall J \in \mathbb{N}: \quad \|u - u^{(J)}\|_X \leq CJ^{-s}. \tag{2.18}$$

Condition (2.18) is related to a smoothness assumption on the admissible unknown data: *e.g.* assume given a "smoothness scale" $X = X_0 \supset X_1 \supset \ldots \supset X_s$ of Hilbert spaces and that $\{\psi_j\}_{j\geq 1}$ is, properly rescaled, a Riesz basis for each X_s, (2.18) is implied by assuming $u \in X_s \subset X$. For Gaussian priors, (2.17) is a truncated Karhunen−Loève expansion, and (2.18) is implied by assuming that u belongs to the domain of a suitable power of the covariance operator. We refer to ([13], Sect. 2) for further details. Assumption 2.1, item (3), implies (2.18) with $s = 1/p - 1$. We also remark that (2.18) is closely connected to compactness of Hessian of the uncertainty-to-observation map (*cf.* condition (B.1) in Prop. B.1). Under the assumption that the dependence of the forward operator $A(u;q)$ on the uncertain parameter $u \in X$ is Lipschitz, *i.e.* there exists $L > 0$ such that for $u_1, u_2 \in X_0$

$$\sup_{q\in\mathcal{X}} \frac{\|A(u_1;q) - A(u_2;q)\|_{\mathcal{Y}'}}{\|q\|_{\mathcal{X}}} \leq L\|u_1 - u_2\|_X,$$

then, choosing $u_1 = u \in X_0$ as in (2.5) and $u_2 = u^{(J)}$ as in (2.17), we find that for given $f \in \mathcal{Y}'$ the corresponding solutions $q = (A(u;\cdot))^{-1}f$ and $q^{(J)} = (A(u^{(J)};\cdot))^{-1}f$ satisfy the estimate

$$\|q - q^{(J)}\|_{\mathcal{X}} \leq CJ^{-s} \tag{2.19}$$

where $C > 0$ is possibly different from the constant in (2.18). It follows from (2.19) and from ([13], Sect. 3.4, Thm. 4.7, Rem. 4.9), that the Bayesian estimates obtained with the dimensionally truncated forward solution $q^{(J)}$ in place of q likewise admit the bound CJ^{-s}. *We therefore note that in the ensuing developments, it suffices to assume that the uncertain input $u \in X$ is parametrized according to (2.17) with a sufficiently large, but finite number J of parameters.*

3. ASYMPTOTIC ANALYSIS OF Z_Γ, Z'_Γ AS $\Gamma \downarrow 0$

We are interested in Bayesian prediction (2.16) in the case that the (co)variance Γ of the (assumed Gaussian) noise η in the observation data δ in (2.2) concentrates, *i.e.* when $\Gamma \downarrow 0$. This will induce concentration of the density (2.15) in u. Based on the integral representation (2.16), (2.14), and bearing in mind the definition (2.4) of the Bayesian potential, the asymptotic behaviour as $\Gamma \downarrow 0$ of the Bayesian estimate (2.16) follows from an asymptotic analysis of the integrals Z_Γ and Z'_Γ in (2.16) by Laplace's method which we present next; necessary results and references are collected for convenience in Section 6. We remark that Laplace's method can be also used to compute approximations of posterior moments by using the explicit expression of the first term in the expansion for the numerator Z_Γ and denominator Z'_Γ, see *e.g.* [36]. In the present work, we derive an expansion of the Bayesian estimate w.r. to the observational noise covariance Γ, which allows to design computational methods numerically stable in the small noise limit.

Throughout, we assume that (possibly after dimension-truncating the parameter space as in Sect. 2.5) that $J < \infty$. We distinguish the cases $K = 1$ (in which case we set $\Gamma^{-1} = \lambda$ so that $\Gamma \downarrow 0$ corresponds to $\lambda \to \infty$) in Propositions A.3 and A.4, and finite $K > 1$. We assume that the forward map has been dimensionally truncated to a finite number $J < \infty$ of parameters as in Section 2.5 and there exists a unique maximum of the Bayesian potential at $\boldsymbol{y}_0$. It can be easily shown that the assumption on the uniqueness of the maximizer implies $K \geq J$ (by considering the linear uncertainty-to-observation map $\mathcal{G}(y) = Ay$, $A \in \mathbb{R}^{K \times J}$). We point out that, for computational purposes, the integrals Z_Γ and Z'_Γ in (2.16) are expressed in terms of the Lebesgue measure $\mathrm{d}\boldsymbol{y}$ on the parameter domain $\mathcal{U}$. Then, for uniform prior π_0, we are interested in the maximizer of

$$\Theta(\boldsymbol{y}) = \exp\big(-\Phi_\Gamma(u; \delta)\big|_{u = \langle u \rangle + \sum_{j=1}^J y_j \psi_j}\big)$$

and for Gaussian prior π_0, (cf. [13], Eq. (4.6))

$$\Theta(\boldsymbol{y}) = \exp\left(-\Phi_\Gamma(u; \delta)\big|_{u = \langle u \rangle + \sum_{j=1}^J y_j \psi_j} - \frac{1}{2} \|y\|_2^2\right).$$

assuming parametrization (2.17) of the uncertain datum u with normally distributed parameters y_j, $j = 1, \ldots, J$. In order to unify notation for the ensuing asymptotic analysis, we introduce the parameter θ, where $\theta = 0$ corresponds to the uniform prior, $\theta = 1$ to the Gaussian prior.

3.1. $\Gamma \downarrow 0$, Case $K = 1$, $J = 1$

Since $K = 1$, the variance Γ of the Gaussian observation noise is a real-valued random variable and we use the Laplace asymptotics with large parameter $\lambda = \Gamma^{-1} > 0$.

Both integrals Z_Γ and Z'_Γ in (2.16) are of the same type; comparing (2.16) with integrals of the form $F(\lambda) = \int_U \phi(\boldsymbol{y}) \exp[\lambda S(\boldsymbol{y})] \mathrm{d}\boldsymbol{y}$ in order to use Laplace's method (cf. Appendix A in Sect. 6), we find

$$\lambda S(\boldsymbol{y}) = -\frac{1}{\Gamma}\left(\Phi_1(\boldsymbol{y}; \delta) + \frac{\theta}{2}\Gamma \|\boldsymbol{y}\|_2^2\right) \tag{3.1}$$

with $\Phi_1(\boldsymbol{y}; \delta) = \frac{1}{2}\left((\delta - \mathcal{G}(u))^\top (\delta - \mathcal{G}(u))\right)$, so that $-2\Gamma S(\boldsymbol{y}) = \|r(\boldsymbol{y})\|_2^2 + \theta\Gamma \|\boldsymbol{y}\|_2^2$ where $r(\boldsymbol{y}) := \mathcal{G}(\boldsymbol{y}) - \delta$ denotes the scalar residual of the data δ w.r. to the uncertainty-to-observation map $\mathcal{G}(\boldsymbol{y})$, and where $\theta = 0$ for uniform prior π_0 and $\theta = 1$ for Gaussian prior π_0. In (3.1), $\|\boldsymbol{y}\|_2$ denotes the Euclidean norm in $\mathbb{R}^J$. To verify the assumptions of Proposition A.3, we calculate

$$S'(\boldsymbol{y}) = -r(\boldsymbol{y})r'(\boldsymbol{y}) - \theta\Gamma\boldsymbol{y}, \quad S''(\boldsymbol{y}) = -\left\{r'(\boldsymbol{y})r'(\boldsymbol{y}) + r(\boldsymbol{y})r''(\boldsymbol{y}) + \theta\Gamma\right\}. \tag{3.2}$$

Based on the expressions (3.2), we see that, in the uniform case, i.e. $\theta = 0$, $\boldsymbol{y}_0 \in \mathrm{int}(U)$ is critical wr. to $S()$ if either $r(\boldsymbol{y}) = 0$ or if $r'(\boldsymbol{y}) = 0$. The former, compatible case corresponds to an exactly solvable inversion, where the observed data is reproduced exactly for some realization $u(\boldsymbol{y}_0) \in X$ of the uncertainty. Note that in the presence of observational noise, exact recovery of the observational data is undesirable in practice, since, in case of ill-posedness of the inverse problem, the exact inversion leads to fitting of high frequencies to the noise, the so-called "overfitting" effect. However, the analysis presented here will be used to develop Bayesian estimation algorithms with performance independent of the size of the noise covariance Γ, i.e. the maximizer of (3.1) will not be used as an estimate of the unknown parameters.

In the latter, incompatible case, we have

$$\exists \boldsymbol{y}_0 \in \mathrm{int}(U): \quad r(\boldsymbol{y}_0) \neq 0, \ r'(\boldsymbol{y}_0) = 0, \ \mathrm{sgn}(r(\boldsymbol{y}_0))r''(\boldsymbol{y}_0) \geq 0. \tag{3.3}$$

We remark that due to the definition $r(\boldsymbol{y}) = \mathcal{O}(q(\boldsymbol{y}))$ with $\mathcal{O}(\cdot) \in \mathcal{X}'$ and δ being independent of $\boldsymbol{y}$, we find

$$r'(\boldsymbol{y}) = \mathcal{O}(q'(\boldsymbol{y})), \ r''(\boldsymbol{y}) = \mathcal{O}(q''(\boldsymbol{y})).$$

The second differential $r''(\boldsymbol{y})$ requires knowledge of the Hessian $D_{\boldsymbol{y}}^2(G(u(\boldsymbol{y}); f))$ of the solution map for the forward problem. Under the assumption that there exists a unique, nondegenerate maximizer $\boldsymbol{y}_0 \in \mathrm{int}(U)$, sufficient conditions are

$$S(\boldsymbol{y}_0) \leq 0, \ \ S'(\boldsymbol{y}_0) = 0, \ \ S''(\boldsymbol{y}_0) < 0. \tag{3.4}$$

In particular, then, all assumptions of Proposition A.3 hold, and both integrals, Z_Γ and Z'_Γ, admit asymptotic expansions (A.5) as $\Gamma \downarrow 0$. The explicit form (A.6) of the principal term of the asymptotics allows to infer:

Theorem 3.1. *Assume that $\mathcal{G}(\cdot)$ and δ are such that the assumptions of Proposition A.3 hold; in particular, that the potential (3.1) satisfies conditions (3.3)−(3.4). Then, for uniform prior π_0, the Bayesian estimate in (2.16) admits an asymptotic expansion*

$$\mathbb{E}^{\pi^\delta}[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} \sim \tilde{a}_0 + \tilde{a}_1 \Gamma + \tilde{a}_2 \Gamma^2 + \dots \quad \Gamma \downarrow 0, \tag{3.5}$$

where $\tilde{a}_0 = \phi(\boldsymbol{y}_0)$ so that

$$\mathbb{E}^{\pi^\delta}[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} = \phi(\boldsymbol{y}_0)(1 + o(1)), \quad \Gamma \downarrow 0. \tag{3.6}$$

The latter relation (3.6) remains valid in the case of a Gaussian prior π_0.

Proof. In the nondegenerate case, with uniform prior π_0, the parameter range is bounded, $\theta = 0$ in (3.1)−(3.3) and (3.4) hold. Proposition A.3 allows to infer that as $\lambda = \Gamma^{-1} \to \infty$, both integrals Z_Γ and Z'_Γ in (2.16) admit asymptotic expansions (A.5). The explicit form (A.6) of the principal term of the asymptotics then implies that the quotient allows to infer that as $\lambda = \Gamma^{-1} \to \infty$, both integrals Z_Γ and Z'_Γ in (2.16) admit asymptotic expansions

$$\exp(-\Gamma^{-1} S(\boldsymbol{y}_0))Z_\Gamma \sim \Gamma^{J/2} \sum_{k \geq 0} a_k \Gamma^k, \qquad \exp(-\Gamma^{-1} S(\boldsymbol{y}_0))Z'_\Gamma \sim \Gamma^{J/2} \sum_{k \geq 0} a'_k \Gamma^k$$

as $\Gamma \downarrow 0$. In particular, these quantities depend continuously on $\Gamma \in [0, \Gamma_0]$ for some $\Gamma_0 > 0$. From the theorem on quotients of asymptotic expansions ([16], Thm. I.3.1 item 3), there holds the asymptotic expansion

$$\mathbb{E}^{\pi^\delta}[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} = \frac{\exp(-\Gamma^{-1} S(\boldsymbol{y}_0))Z'_\Gamma}{\exp(-\Gamma^{-1} S(\boldsymbol{y}_0))Z_\Gamma} \sim \sum_{k \geq 0} \tilde{a}_k \Gamma^k,$$

and from the explicit form (A.6) follows $\tilde{a}_0 = a'_0/a_0 = \phi(\boldsymbol{y}_0)$.

In the Gaussian case, the parameter domain is unbounded and $\theta = 1$ in (3.1), so that the function S depends (linearly) on $\lambda = \Gamma^{-1}$. Since $J = 1$, the function S is the regularized least-squares functional

$$S(\boldsymbol{y}) = -\frac{1}{2} \left[\|r(\boldsymbol{y})\|_2^2 + \Gamma \|\boldsymbol{y}\|_2^2 \right].$$

This function has, for $\Gamma > 0$, a nondegenerate maximum at $\boldsymbol{y}_0(\Gamma)$. This assumption allows to derive a result: let $\boldsymbol{y}_0, \Gamma_0$ denote the (nondegenerate) maximizer of the function S. Then, in a neighbourhood $U(\Gamma_0)$ of Γ_0, exists a continuously differentiable function $\boldsymbol{y}_0(\Gamma)$ with $\boldsymbol{y}_0(\Gamma_0) = \Gamma_0$ and $\boldsymbol{y}_0(\Gamma)$ is a strict local maximizer of $S(\boldsymbol{y})$ for all $\Gamma \in U_{\Gamma_0}$. Hence, as $\Gamma \downarrow 0$, the maximum of $S(\boldsymbol{y})$ remains in a fixed, compact subset of $\mathbb{R}$. We may then apply Proposition A.2 to conclude (3.6) in the Gaussian case. $\qquad \square$

The results (3.5), (3.7) show that *despite the generally exponential growth wr. to $\Gamma \downarrow 0$ in the asymptotics (A.5), (A.8) for either of the constants Z_Γ and Z'_Γ*, (which we remark in passing shows that the exponential dependence on $1/\Gamma$ in the Smolyak quadrature error bounds in [30] can, in general, not be improved), *the Bayesian estimate (2.16) given by Z'_Γ/Z_Γ has a finite one-sided limit as $\Gamma \downarrow 0$* which, in case that the Hessian $S''(\boldsymbol{y}_0)$ is nondegenerate and that $S(\boldsymbol{y})$ is unimodal at the critical point $\boldsymbol{y}_0 \in \mathrm{int}(U)$ of the residual $r(\boldsymbol{y}_0)$ in (3.1), equals the QoI ϕ evaluated at this (MAP) point.

For Bayesian potentials $\Phi_\Gamma(\boldsymbol{y}; \delta)$ which attain a unique, global minimum at $\boldsymbol{y}_0 \in \text{int}(U)$ which is degenerate, *i.e.* where the Hessian $S''(\boldsymbol{y}_0)$ has a nontrivial nullspace, then, for every finite truncation dimension J as in Section 2.5, there exist $N(J) \in \mathbb{N}$ and $r_k(J) \in \mathbb{Q}$ such that

$$\mathbb{E}^{\pi^\delta}[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} = \frac{a_{00}[\phi]}{a_{00}[1]}(1 + o(\Gamma)) \quad \text{as} \quad \Gamma \downarrow 0, \tag{3.7}$$

where $a_{00}[\phi]$, $a_{00}[1]$ denote the leading terms in the asymptotic expansions (A.8) of Z' and of Z, respectively, which are independent of Γ (but depend on δ and on J).

3.2. $\Gamma \downarrow 0$, Case $1 < K < \infty$

The asymptotic analysis in the case $K > 1$ many observables can be derived in an analogous way: for uniform prior π_0, the asymptotic expansion follows from Proposition A.3, and for Gaussian prior, the relation (3.6) is a consequence of the general result ([14], Thm. 3.5).

Consider now a finite number $1 < K < \infty$ of observables. Then $\Gamma \in \mathbb{R}^{K \times K}$ is the symmetric and (assumed) positive definite covariance matrix of the observation noise $\eta \in \mathbb{R}^K$. It can therefore be diagonalized:

$$\Gamma = PMP^\top, \quad M = \text{diag}\{\gamma_1, \dots \gamma_K\}, \ 0 < \gamma_1 \leq \dots \leq \gamma_K, \ P^\top P = 1. \tag{3.8}$$

In the following, we will assume that the covariance matrix of the noise is of the form $\Gamma = \gamma I$, *i.e.* $\gamma = \gamma_1 = \dots = \gamma_K$. The case when all eigenvalues of Γ tend to zero at the same rate, *i.e.* when $\gamma_K \downarrow 0$ while γ_K/γ_1 remains bounded, is analogous to the case $\Gamma = \gamma I$. Various intermediate cases (*e.g.* with the γ_k tending to zero at different rates) will not be elaborated here.

Since $K > 1$, the residual of the data δ wr. to the uncertainty-to-observation map $R(\boldsymbol{y}) = \mathcal{G}(\boldsymbol{y}) - \delta \in \mathbb{R}^K$ is a K-vector with component residuals $r_k(\boldsymbol{y})$, $k = 1, \dots, K$. For (3.1), we choose in Proposition A.3

$$S(\boldsymbol{y}) = -\Phi_\Gamma(\boldsymbol{y}; \delta) - \theta\frac{1}{2}\|\boldsymbol{y}\|_2^2 = -\frac{1}{2}R(\boldsymbol{y})^\top \Gamma^{-1} R(\boldsymbol{y}) - \theta\frac{1}{2}\|\boldsymbol{y}\|_2^2 = -\frac{1}{\gamma}\left(\frac{1}{2}R(\boldsymbol{y})^\top R(\boldsymbol{y}) - \theta\frac{1}{2}\gamma\|\boldsymbol{y}\|_2^2\right). \tag{3.9}$$

Under the assumption of a unique, nondegenerate maximizer $\boldsymbol{y}_0 \in \text{int}(U)$, $\boldsymbol{y}_0$ satisfies the sufficient conditions given by (3.4) and (3.5) of Theorem 3.1 holds also in this case. Note that this assumption implies that $K \geq J$.

The asymptotic expansions (3.5), (3.7) in Theorem 3.1 show that the Bayesian estimate (2.16) converges, in the zero observation noise limit and for nondegenerate critical points, to the QoI ϕ at critical points $\boldsymbol{y}_0$ of the parameter sequence $\boldsymbol{y}$. These parameters, in turn, can be determined numerically from the data by the solution of a nonlinear least-squares problem (for the potential Φ_Γ) rather than by numerical integration.

Moreover, Theorem 3.1 shows that in the limit $\Gamma = 0$ the Bayesian estimate (2.16) behaves numerically as quotient of infinite quantities which admits a finite limiting value at $\Gamma = 0$. This suggests that the deterministic quadrature approach of [29, 30] for the evaluation of Z_Γ, Z'_Γ becomes numerically unstable as $\Gamma \downarrow 0$. To deal with positive, but small observation noise (co)variance Γ, therefore, the quadrature algorithms must be modified in order to remain numerically stable. As suggested by the asymptotic expansions (3.5) and by the explicit form (3.6) of its leading term, the numerical treatment of the limit $\Gamma \downarrow 0$ will require addressing the deterministic, nonlinear least-squares problem

$$\min_{\boldsymbol{y} \in U} \Phi_\Gamma(\boldsymbol{y}; \delta) + \theta/2\|\boldsymbol{y}\|_2^2, \tag{3.10}$$

where $\theta = 0$ in the uniform case, and $\theta = 1$ in the Gaussian case. The minimization problem (3.10) is generally ill-conditioned for countably-parametric operator equations with uniform prior, as considered in Section 2.3.1. In the lognormal case with Gaussian prior discussed in Section 2.3.2, the minimizer of (3.10) depends on Γ, generally. It corresponds to a Tikhonov regularized solution of the nonlinear least-squares problem $\min_{\boldsymbol{y} \in U} \Phi_\Gamma(\boldsymbol{y}; \delta)$ and converges, as $\Gamma \downarrow 0$, to a MAP estimator (see, *e.g.*, [14, 23], Chap. 3.1.1 and [13], Sect. 2.2 for more details on the MAP estimator and its relation to regularized, deterministic least-squares minimization problems).

4. NUMERICAL ANALYSIS OF SMALL OBSERVATION NOISE COVARIANCE Γ

We analyze computational Bayesian inversion for positive, but possibly small observation noise (co)variance $\Gamma > 0$. We also assume throughout that the Hessian $S''(\boldsymbol{y}_0)$ of the Γ-scaled least-squares functional

$$S(\boldsymbol{y}) = \Gamma \left(-\Phi_\Gamma(\boldsymbol{y}; \delta) - \frac{\theta}{2} \|\boldsymbol{y}\|_2^2 \right) \tag{4.1}$$

is nondegenerate, negative definite, *uniformly with respect to* $0 < \Gamma < 1$. Then, the asymptotic analysis of Section 3 applies (see Sect. A.3 for remarks on the degenerate case (3.7)). Quasi-Newton iterations with symmetric updates are known to produce locally superlinearly convergent approximations of both, critical point $\boldsymbol{y}_0$ of S and of its Hessian $S''_{\boldsymbol{yy}}(\boldsymbol{y}_0)$ (we refer to [10,18] for details; optionally, also 2nd order adjoint techniques [19] could be used). The point $\boldsymbol{y}_0$ is known to be related to a MAP estimate, in the Gaussian case (see [13], Sect. 4.3). We therefore assume that (*e.g.* upon termination of the QN-SR1 algorithm from [10]), that the point $\boldsymbol{y}_0$ and the Hessian $S''_{\boldsymbol{yy}}(\boldsymbol{y}_0)$ of the Bayesian potential Φ are available.

Then, from (A.5), (A.6), *upon termination of the QN process the leading term of the asymptotic expansion* (A.5) *in the $\Gamma \downarrow 0$ limit is accessible for both Z_Γ and Z'_Γ in* (2.16). Also, since $S(\boldsymbol{y})$ is either independent of Γ (for uniform prior) or depends linearly on it (Gaussian prior), *the performance of the QN method is independent of Γ*.

Based on Theorem 3.1, we propose two computational strategies which we prove to be robust with respect to vanishing observation noise (co)variance: first, we assume available knowledge of $\boldsymbol{y}_0$ and $S''(\boldsymbol{y}_0)$ to "precondition" dimension-adaptive, deterministic Smolyak quadratures proposed in [29,30]: for uniform prior, they are based on the midpoint rule as lowest order quadrature rule and for Gaussian prior, on the lowest Gauss–Hermite quadrature formula. For small, but positive $\Gamma > 0$, shifting the coordinate origin to $\boldsymbol{y}_0$ will thus identify the dominant contribution to the posterior expectation in the first sweep of the dimension-adaptive Smolyak algorithm. For small values of Γ, however, the high curvatures in the shifted posterior density due to the concentration entails excessive refinements of the adaptive quadratures in all concentrating coordinates. To render the performance of the adaptive Smolyak algorithm independent of $0 < \Gamma < 1$, we propose reparametrization of the posterior density near MAP-points $\boldsymbol{y}_0$ of posterior concentration. We prove in Theorem 4.1 ahead that this reparametrization renders the integrand functions in the adaptive Smolyak quadrature scheme unimodal, with maximum at $\boldsymbol{y} = 0$, and holomorphic with curvatures which are bounded independently of Γ. We remark that inclusion of curvature information on the Bayesian potential Φ_Γ near concentration points of the posterior has been proposed to increase efficiency of MCMC methods. In [25], operator weighted proposals based on Hessian information of the likelihood are introduced and (numerically) shown to significantly speed-up MCMC. Details on the use of Hessian information within the so-called "stochastic Newton framework" can be found *e.g.* in [4,26]. A further MCMC variant relies on the exploration of the Riemann geometry of the parameter space to automatically adapt to the local structure of the posterior. We refer to [20] for more details. The low dimensionality of the data-informed subspace is investigated *e.g.* in [11,12]. The idea of "likelihood-informed" ("DILI") subspaces has been proposed to design MCMC in this low-dimensional subspace. It has been found improve computational efficiency, but can not, of course, increase convergence rates beyond the intrinsic limit $1/2$ of MC. We also remark that reparametrization the posterior to alleviate ill-conditioning due to, *e.g.*, concentration effects, has been suggested in [27], albeit with a different approach related to optimal transport.

The second strategy is based on (generalized) *Richardson extrapolation to the limit of zero observation noise* justified by the asymptotic expansions (3.5) and (3.7). In this case, Γ is treated as a algorithm-parameter as follows: for data δ with given, small observation noise variance $\Gamma_{\text{obs}} > 0$, compute (in parallel) estimates (2.16) for several, *synthetic*, large values Γ_k of Γ, with $0 \leq \Gamma_{\text{obs}} \ll \Gamma_k \leq 1$ and extrapolate to $0 \leq \Gamma_{\text{obs}} \ll 1$.

4.1. Curvature rescaling

We develop the approach in the (notationally less involved) case of $K = 1$ observation functional. We apply Proposition A.3 with S as in (3.1), (3.9). We assume in (3.8) that $\gamma = \Gamma > 0$ and initially for uniform

prior π_0. Then, $S(\boldsymbol{y})$ *in* (4.1) *is independent of* Γ. A Quasi-Newton method with symmetric rank 1 (SR1) update (*cf.* [10, 18]) converges superlinearly and yield, upon termination, the (unique by Assumption 2. of Prop. A.3) maximum point $\boldsymbol{y}_0$ of $S(\boldsymbol{y})$ in U and the (positive) definite Hessian approximation $H_S \sim -S''(\boldsymbol{y}_0)$ at $\boldsymbol{y}_0$ (we refer to [10, 18] for details). Due to the use of symmetric QN updates and assumption 2 of Prop. A.3), then, the Hessian H_S is symmetric positive definite, and the total work required by QN for $S(\boldsymbol{y})$ in (4.1) scales polynomially in J (updating the inverse of the SR1 approximation using the Sherman–Morrison formula leads to a performance similar to the BFGS algorithm, *e.g.* discussed in [17]) *uniformly with respect to* Γ (as $S(\boldsymbol{y})$ in (4.1) is either independent of Γ or depends linearly on it). Denote by μ_j with $0 < \mu_J \leq \mu_{J-1} \leq \ldots \leq \mu_1$ its eigenvalues (enumerated in decreasing order of magnitude and counting multiplicity), and by Q the $J \times J$ orthogonal matrix of its J eigenvectors, *i.e.*

$$H_S Q = QM, \quad M := \mathrm{diag}\{\mu_1, \ldots, \mu_J\}. \tag{4.2}$$

We next perform an *affine* change of variables $\boldsymbol{y} = \varphi(\check{\boldsymbol{x}})$ in $S(\boldsymbol{y})$ such that $\boldsymbol{y}_0 = \varphi(0)$ and such that the Hessian of $(S \circ \varphi)(\check{\boldsymbol{x}})$ is diagonal: $(D_{\boldsymbol{x}}^2(S \circ \varphi))(0) = M = \mathrm{diag}\{\mu_1, \ldots, \mu_J\}$. With $\check{V} := Q^\top(U - \{\boldsymbol{y}_0\})$, we find

$$\int_U \exp(\lambda S(\boldsymbol{y}))\phi(\boldsymbol{y})\mathrm{d}\pi_0(\boldsymbol{y}) = \int_{\check{V}} \check{g}(\check{\boldsymbol{x}})\mathrm{d}\pi_0(\check{\boldsymbol{x}}), \tag{4.3}$$

where $\check{g}(\check{\boldsymbol{x}}) := \exp(\lambda S(\boldsymbol{y}_0 + Q\check{\boldsymbol{x}}))\phi(\boldsymbol{y}_0 + Q\check{\boldsymbol{x}})$.

By the orthogonality of Q, the integration domain $\check{V}$ is a rotated and translated unit (wr. to the prior π_0) cube. The transformed function $\check{S}(\check{\boldsymbol{x}}) := S(\boldsymbol{y}_0 + Q\check{\boldsymbol{x}})$ is analytic in a vicinity of $\check{\boldsymbol{x}} = 0$; therefore

$$\check{S}(\check{\boldsymbol{x}}) = S(\boldsymbol{y}_0) + \frac{1}{2}\check{\boldsymbol{x}}^\top Q^\top S''_{\boldsymbol{yy}}(\boldsymbol{y}_0)Q\check{\boldsymbol{x}} + O(\|\check{\boldsymbol{x}}\|_2^3) = S(\boldsymbol{y}_0) + \frac{1}{2}\check{\boldsymbol{x}}^\top M\check{\boldsymbol{x}} + O(\|\check{\boldsymbol{x}}\|_2^3)$$

with $O(\cdot)$ being uniform w.r. to Γ.

The application of Proposition A.3 to the transformed integral (4.3) (for $\lambda = \Gamma^{-1} \gg 1$) gives that the transformed integrand function $g(\check{\boldsymbol{x}})$ in (4.3) will depend on the coordinate $\check{x}_j$ *to leading order* (as the affine coordinate change in (4.3) will reach the normal form (A.7) only up to higher order terms) as $\exp(-\Gamma^{-1}\mu_j\check{x}_j^2/2)$. The ordering (4.2) of the μ_j then implies that the strongest "concentration" of the integrand function $g(\check{\boldsymbol{x}})$ as $\Gamma \downarrow 0$ occurs in coordinate $\check{x}_1$ at $\check{x}_1 = 0$, on scale $\Gamma^{-1/2}\mu_1^{1/2}$. This observation suggests curvature-rescaling of the form

$$\check{\boldsymbol{x}} := M^{-1/2}\Gamma^{1/2}\boldsymbol{x}, \quad g(\boldsymbol{x}) := \check{g}(M^{-1/2}\Gamma^{1/2}\boldsymbol{x}). \tag{4.4}$$

Theorem 4.1. *Under Assumption* 2.1 *and for nondegenerate Hessian* $S''_{\boldsymbol{yy}}$ *and for analytic parametric forward maps, for uniform prior* π_0, *the curvature rescaling transformation*

$$\boldsymbol{y} = \varphi(\boldsymbol{x}) := \boldsymbol{y}_0 + QM^{-1/2}\Gamma^{1/2}\boldsymbol{x} \tag{4.5}$$

with M *as in* (4.2) *yields a transformed posterior density* $g(\boldsymbol{x}) := \exp(\Gamma^{-1}(S \circ \varphi)(\boldsymbol{x}))(\phi \circ \varphi)(\boldsymbol{x})$ *that is analytic in a neighborhood of* $\boldsymbol{x} = 0 \in \mathbb{R}^J$. *The size of the domain of analyticity of* $g(\boldsymbol{x})$ *is independent of* $0 < \Gamma < 1$. *In this domain, the derivatives of the rescaled posterior density admit analytic regularity estimates which are uniform with respect to* Γ.

Proof. As π_0 is uniform, both $\phi(\boldsymbol{y})$ and $S(\boldsymbol{y})$ are independent of Γ and analytic. Since the affine change of variables $\check{\varphi}$ is independent of Γ, also the functions $\check{S}(\check{\boldsymbol{x}}) = S(\boldsymbol{y}_0 + Q\check{\boldsymbol{x}})$ and $\phi(\boldsymbol{y}_0 + Q\check{\boldsymbol{x}})$ are analytic w.r. to $\check{\boldsymbol{x}}$ in a neighborhood $\check{V}$ of $\check{\boldsymbol{x}} = 0$ which is independent of Γ. There, $\check{S}$ admits the convergent power series representation

$$\check{S}(\check{\boldsymbol{x}}) = S(\boldsymbol{y}_0) + \frac{1}{2}\check{\boldsymbol{x}}^\top M\check{\boldsymbol{x}} + \sum_{k \geq 3} \sum_{|\alpha|=k} t_\alpha \check{\boldsymbol{x}}^\alpha, \tag{4.6}$$

where all coefficients $t_\alpha := \frac{1}{\alpha!}(D_{\check{x}}^\alpha \check{S})(0)$ are independent of Γ. Multiplying (4.6) with Γ^{-1} and inserting (4.4) results in the (nonasymptotic) identity

$$\frac{(S \circ \varphi)(\boldsymbol{x})}{\Gamma} = \frac{S(\boldsymbol{y}_0)}{\Gamma} + \frac{1}{2}\boldsymbol{x}^\top \boldsymbol{x} + \Gamma^{1/2}\left\{\sum_{k \geq 3} \Gamma^{(k-3)/2} \sum_{|\alpha|=k} t_\alpha (M^{-1/2}\boldsymbol{x})^\alpha\right\}. \tag{4.7}$$

Here, the power series in parentheses converges uniformly w.r. to $\Gamma \leq 1$ for $\boldsymbol{x} \in M^{1/2}\check{V}$. Hence, as $\Gamma \downarrow 0$, $\Gamma^{-1}(S \circ \varphi)(\boldsymbol{x})$ tends to a quadratic. All $\boldsymbol{x}$-derivatives of the rescaled integrand function $S \circ \varphi$ exist and are bounded uniformly with respect to Γ. Moreover, the convergence radius of the power series shown in parentheses in (4.7) *increases* with decreasing Γ. An analogous power series argument shows that the QoI $(\phi \circ \varphi)(\boldsymbol{x})$ is analytic with respect to the curvature-rescaled coordinates $\boldsymbol{x}$, with domain of analyticity that is even increasing as $\Gamma \downarrow 0$. $\qquad\square$

4.2. Curvature-rescaled adaptive Smolyak quadrature

The preceding observations motivate *curvature rescaled, adaptive Smolyak quadrature*. Given a tolerance parameter $\tau > 0$, and $\lambda = \Gamma^{-1} > 0$, (where we think of Γ as observation noise covariance Γ_{obs} associated with data δ in (2.2)), define the closed intervals $I_j(\tau, \lambda) := \{x_j \in \mathbb{R} : \exp(-\lambda\mu_j x_j^2/2) \geq \tau\}$, $j \in \mathbb{J}$. The I_j are bounded intervals centered at $x_j = 0$ which are nested by the ordering (4.2) of the μ_j: $I_1 \subseteq I_2 \subseteq \ldots \subseteq I_J \subseteq \ldots$. Define

$$J_*(\tau, \lambda) := \begin{cases} 0 & \text{if } \{j \in \mathbb{J}|I_j(\tau,\lambda) \subset \check{V}\} = \emptyset, \\ \max\{j \in \mathbb{J}|I_j(\tau,\lambda) \subset \check{V}\} & \text{otherwise.} \end{cases} \tag{4.8}$$

Notice that $J_*(\tau, \lambda)$ in (4.8) is monotonically increasing for decreasing τ at fixed $\lambda = \Gamma^{-1}$.

If $J_* = 0$, there is no coordinate in which the integrand function $f(\boldsymbol{x})$ in the transformed integral (4.3) over V is concentrating. If $J_* > 0$, the integrand function in the transformed integral (4.3) over V concentrates numerically (at threshold τ) near $x_j = 0$ in the coordinates $x_1, \ldots, x_{J_*}$. Accordingly, if $U = [-1, 1]^J$, prior to application of the adaptive Smolyak quadrature algorithm, we rescale the integration coordinates according to

$$\hat{x}_j := x_j\xi_j, \quad \xi_j := \left(\frac{\mu_j/\Gamma}{2|\ln\tau|}\right)^{1/2}, \quad j = 1, \ldots, J_* \tag{4.9}$$

and set $\hat{x}_j := x_j$ for $j > J_*$. Then, *truncate the integration domains to $I_j(\tau, \lambda)$ for dimensions $j = 1, \ldots, J_*$*:

$$\check{V}_\tau := \prod_{1 \leq j \leq J_*} I_j(\tau, \lambda) \times \prod_{j > J_*} \check{V}_j. \tag{4.10}$$

For the resulting truncation error holds the error bound

$$\left|\int_{\check{V}} f(\boldsymbol{x})\mathrm{d}\pi_0(\boldsymbol{x}) - \int_{\check{V}_\tau} f(\boldsymbol{x})\mathrm{d}\pi_0(\boldsymbol{x})\right| \leq \pi_0(\check{V}\backslash\check{V}_\tau) \sup_{\boldsymbol{y} \in U} \|\phi(\boldsymbol{y})\|_{\mathcal{S}} \leq \tau \sup_{\boldsymbol{y} \in U} \|\phi(\boldsymbol{y})\|_{\mathcal{S}}.$$

The *variable metric adaptive Smolyak quadrature algorithm* consists in applying the adaptive Smolyak algorithm from [30] to the rescaled integrand function where coordinates $(y_1, \ldots, y_{J_*})$ are transformed with (4.5).

Remark 4.2. The proof of Theorems 4.1 and (4.9) show that, *asymptotically as $\Gamma \downarrow 0$*, at any fixed, finite truncation dimension J, the posterior density eventually concentrates in *all* coordinates y_j, $j = 1, \ldots, J$.

The compactness of the Hessian $S''_{\boldsymbol{yy}}(\boldsymbol{y}_0)$ shown in Proposition B.1 implies that, generically, $\{\mu_j\}_{j \geq 1}$ accumulates at 0 for increasing J. Proposition B.1 implies that for every fixed obervation noise covariance $\Gamma_{\mathrm{obs}} > 0$ and for every prescribed truncation parameter $\tau > 0$ exists a finite "crossover dimension" $J_*(\Gamma_{\mathrm{obs}}, \tau)$ such that $\mu_j/\Gamma_{\mathrm{obs}} \leq 1$ for all $j \geq J_*$. The compactness and the spectrum of the Hessian $S''_{\boldsymbol{yy}}(\boldsymbol{y}_0)$ provide for every $\Gamma_{\mathrm{obs}} > 0$ a (finite!) bound on the dimension of the parameter space where the Bayesian posterior can concentrate. We also refer to the survey [1] for discussion of second order information at the MAP point and of "effective" dimension in importance sampling acceleration of MCMC methods.

Theorem 4.1 implies that the curvature-rescaling (4.5) in conjunction with the adaptive Smolyak quadratures from [29,30] *in rescaled coordinates* for Z_Γ and Z'_Γ will converge with rate and constants which are independent of Γ.

4.3. Extrapolation to observation noise covariance $0 \le \Gamma_{\mathbf{obs}} \ll 1$

The previous regularizations required nondegeneracy and explicit knowledge of the Hessian at the critical point $\boldsymbol{y}_0 \in U$. We therefore present an alternative approach which avoids explicitly accessing curvature information, which is mathematically justified for both, regular as well as degenerate Hessians $S''(\boldsymbol{y}_0)$. It is based on generalized Richardson extrapolation with respect to Γ of Bayesian estimates to observation noise variance $0 \le \Gamma_{\mathrm{obs}} \ll 1$.

This approach is justified by the asymptotic expansions (3.7) with respect to $\Gamma \downarrow 0$ which in turn are based on Propositions A.3 and A.4. We emphasize that *only the existence of an asymptotic expansion* such as (A.5) is used, and no explicit expressions of coefficients in these expansions is required in our computational strategy. We present the details, for simplicity only in the case $K = 1$, following the generalized Richardson extrapolation algorithm developed and analyzed in [34].

4.3.1. Generalized Richardson extrapolation

Let $B(\Gamma) : \mathbb{R}_> \mapsto \mathbb{R}$ be a scalar function of a continuous variable $\Gamma > 0$, defined on $0 < \Gamma \le \Gamma_0 < \infty$. Assume that there exist constants B and β_k, $k = 1, 2, \ldots$, which are independent of Γ and functions $\varrho_k(\Gamma)$ which form an asymptotic sequence in the sense that

$$\varrho_{k+1}(\Gamma) = o(\varrho_k(\Gamma)) \quad \text{as} \quad \Gamma \downarrow 0$$

and assume that $B(\Gamma)$ admits the asymptotic expansion

$$B(\Gamma) \sim B + \sum_{k=1}^{\infty} \beta_k \varrho_k(\Gamma) \quad \text{as} \quad \Gamma \downarrow 0. \tag{4.11}$$

For a strictly monotonically decreasing sequence $\Gamma_0 > \Gamma_1 > \Gamma_2 > \ldots$ in $(0, \Gamma_0]$ with $0 = \lim_{k \to \infty} \Gamma_k$ and for each pair (j, q) of positive integers, define the sequence $\bar{\beta}_0, \bar{\beta}_1, \ldots, \bar{\beta}_q$ as solution of the linear system of $q + 1$ equations for the $q + 1$ unknowns $\beta_0^{jq}, \ldots, \beta_q^{jq}$ which is given by

$$B(\Gamma_l) = \sum_{k=0}^{q} \beta_k^{jq} \varrho_k(\Gamma_l), \quad j \le l \le j + q. \tag{4.12}$$

Then the β_0^{jq} obtained in (4.12) are approximations of $\lim_{\Gamma \downarrow 0} B(\Gamma)$; specifically, there holds (*cf.* [34], Thm. 2.2):

Proposition 4.3. *Assume that $B(\Gamma)$ admits the asymptotic expansion* (4.11). *Then, for some monotonically decreasing sequence* $\{\Gamma_k\}_{k \ge 1}$, *accumulating at $\Gamma = 0$ such that*

$$\lim_{l \to \infty} \frac{\varrho_k(\Gamma_{l+1})}{\varrho_k(\Gamma_l)} = c_k \ne 1 \; \text{ and } \; c_j \ne c_k \; \text{ for } \; j \ne k, \tag{4.13}$$

for every $q \in \mathbb{N}$ fixed, the β_0^{jq} and $B = \lim_{\Gamma \downarrow 0} B(\Gamma)$ in (4.12) *satisfy*

$$\beta_0^{jq} - B \sim \beta_{q+1} \left[\prod_{i=1}^{q} \left(\frac{\beta_{q+1} - \beta_i}{1 - \beta_i} \right) \right] \varrho_{q+1}(\Gamma_j) \quad as \quad j \to \infty. \tag{4.14}$$

Under condition (4.13), $\{\beta_0^{jq}\}_{j \ge 1}$ *tends to B faster than* $\{\beta_0^{j,q-1}\}_{j \ge 1}$, *i.e.*

$$\frac{\beta_0^{jq} - B}{\beta_0^{j,q-1} - B} = O\left(\frac{\varrho_{q+1}(\Gamma_j)}{\varrho_q(\Gamma_j)} \right) = o(1) \quad \text{as} \quad j \to \infty.$$

4.3.2. Extrapolation to observation noise covariance $0 \leq \Gamma_{\mathrm{obs}} \ll 1$

Comparing the preceding result on generalized Richardson extrapolation with the asymptotic expansions (3.5) in the regular case (where, according to Theorem 3.1, we have $\varrho_k(\Gamma) = \Gamma^k$), we see immediately that in the regular case (3.5), Assumption (4.13) is satisfied.

Proposition 4.4. *Assume that $\mathcal{G}(\cdot)$ and δ are such that all Assumptions of Proposition* A.3 *hold. Then, for the monotonically decreasing sequence $(\Gamma_k)_{k \geq 1}$ with $\Gamma_k = h_0 \rho^k$, $0 < \rho < 1$, $h_0 > 0$, Assumption* (4.13) *is fulfilled.*

Proof. In the nondegenerate case, Theorem 3.1 and (3.5) imply $\varrho_k(\Gamma) = \Gamma^k$. Therefore, there holds

$$\lim_{l \to \infty} \frac{\varrho_k(\Gamma_{l+1})}{\varrho_k(\Gamma_l)} = \lim_{l \to \infty} \left(\frac{\Gamma_{l+1}}{\Gamma_l}\right)^k = \rho^k < 1.$$

Therefore, the assumptions of Proposition 4.3 are satisfied, and (4.14) holds. $\qquad\square$

Even without explicit knowledge of coefficients and exponents in the asymptotic expansions (3.5), we may therefore apply (4.12) in order to extrapolate to the limit $\Gamma \downarrow 0$ as follows.

Given *one set of data δ*, the Bayesian estimate (2.16) is approximated numerically for a sequence $\boldsymbol{\Gamma} = \{\Gamma_k\}_{k=1}^K$ of *synthetic (i.e. larger than Γ_{obs}) variances which decrease monotonically.* Owing to numerical instability in the extrapolation of Z_Γ and Z'_Γ in (2.16) for positive, but small observation noise covariance Γ_{obs}, according to [34] the sequence $\boldsymbol{\Gamma} = \{\Gamma_k\}_{k \geq 1}$ should be chosen as

$$\boldsymbol{\Gamma} : \Gamma_k = 2^{-k+1} \quad \text{(geometric sequence).} \tag{4.15}$$

We note in passing that other sequences with slower than geometric decrease (4.15) are infeasible even in the nondegenerate case, as they violate the stability conditions in [34].

To avoid ambiguity in the notation of the observational noise and of the artificial noise, we will denote in the following the variance in the additive Gaussian noise η in the measurement data δ in (2.2) by Γ_{obs}. Generalized extrapolation to small $\Gamma_{\mathrm{obs}} > 0$ or to the limit $\Gamma_{\mathrm{obs}} = 0$ then proceeds by interpolating the Bayesian predictions Z'_{Γ_k} computed for different values of k with a high order polynomial, and evaluating at the actual given variance $\Gamma_{\mathrm{obs}} \geq 0$.

The extrapolation approach is also viable for small, positive observation noise variance $\Gamma_{\mathrm{obs}} > 0$ for which a direct quadrature evaluation would be infeasible due to concentration phenomena.

Extrapolation to small observation noise covariance $\Gamma_{\mathrm{obs}} \geq 0$ is based on:

(a) for one given set of data δ, the Smolyak quadrature algorithm can be executed separately and in parallel for each synthetic observation variance Γ_k,

(b) the integrand functions which are to be evaluated in the adaptive Smolyak quadrature approximation of the integral(s) Z_{Γ_k} and Z'_{Γ_k} in (2.16) *depend on Γ_k only via the Bayesian potential Φ_{Γ_k} in* (2.4). *Therefore, given a set of quadrature points in U, one numerical evaluation of the uncertainty-to-observation map $\mathcal{G}(\cdot)$ per quadrature point is necessary to compute both quadrature approximations Z_{Γ_k} and Z'_{Γ_k} in* (2.16).

While the generalized Richardson extrapolation to the limit is mathematically justified by the asymptotic expansions (3.5), (3.7) and by the analysis in [34], we remind that the present setting will require their use for *positive, small but fixed observation noise covariance $\Gamma_{\mathrm{obs}} > 0$*; the next result is analogous to (4.14).

Proposition 4.5. *Assume that the quantity $B(\Gamma) \in C([0, \Gamma_0]) \cap C^{q+1}(0, \Gamma])$ for some integer $q \geq 1$ and that it admits an asymptotic expansion* (4.11) *with $\varrho_k(\Gamma) = \Gamma^k$ as in* (3.5) *for the Bayesian estimate in the nondegenerate case.*

Given an integer $j \in \mathbb{N}$ and a sequence of (synthetic) observation noise covariances $1 = \Gamma_0 > \Gamma_1 > \ldots > \Gamma_j > \ldots > \Gamma_{j+q} \gg \Gamma_{\mathrm{obs}} \geq 0$ (which are algorithmic parameters *unrelated to the given observation noise*

variance Γ_{obs}) with $q \geq 1$ fixed and with Γ_{j+q} (substantially) larger than the actual observation noise covariance $\Gamma_{\text{obs}} \geq 0$, compute the coefficients $(\beta_k^{jq})_{k=0}^{q}$ in (4.12). Then, there holds the asymptotic error bound

$$|B(\Gamma_{\text{obs}}) - \sum_{k=0}^{q} \beta_k^{jq} \varrho_k(\Gamma_{\text{obs}})| = O(|\Gamma_j - \Gamma_{\text{obs}}|^{q+1}).$$

Here, the constant implied in $O()$ depends on q and on $B(\Gamma)$, but is independent of j.

Proof. For $\Gamma_{\text{obs}} = 0$, the assertion is Proposition 4.3. We may therefore assume that $\Gamma_{\text{obs}} > 0$. The continuity $B \in C([0, \Gamma_0])$ and the monotonicity of the sequence $\{\Gamma_k\}_{k \geq 0}$ imply that for $q \geq 1$, and for every $q + 1$ tuple $\{\Gamma_k\}_{k=j}^{j+q}$, there is a unique interpolation polynomial $\pi_q(\Gamma; \{\Gamma_k\}_{k=j}^{j+q})$ which interpolates $B(\Gamma)$ in the tuple $\{\Gamma_k\}_{k=j}^{j+q}$. For the interpolation error at the point Γ_{obs} holds

$$B(\Gamma_{\text{obs}}) - \pi_q(\Gamma_{\text{obs}}; \{\Gamma_k\}_{k=j}^{j+q}) = \frac{B^{(q+1)}(\xi)}{(q+1)!}\omega_{q+1}(\Gamma_{\text{obs}}), \quad \xi \in \text{conv}\{\Gamma_{\text{obs}}, \Gamma_j, \Gamma_{j+1}, \ldots, \Gamma_{j+q}\}$$

with $\omega_{q+1}(x) = \prod_{0 \leq k \leq q}(x - \Gamma_{j+k})$. The error representation remains valid in the case (of main interest to us) that $\Gamma_{\text{obs}} \notin \text{conv}\{\Gamma_k\}_{k=j}^{j+q}$. Since, for $0 \leq \Gamma_{\text{obs}} < \Gamma_{j+q} < \ldots < \Gamma_j \leq \Gamma_0$, this polynomial satisfies the error representation, we estimate

$$|B(\Gamma_{\text{obs}}) - \pi_q(\Gamma_{\text{obs}}; \{\Gamma_k\}_{k=j}^{j+q})| \leq \frac{1}{(q+1)!}\|B^{(q+1)}\|_{L^\infty(\Gamma_{\text{obs}}, \Gamma_j)}|\Gamma_j - \Gamma_{\text{obs}}|^{q+1}$$

and the assertion follows for $\Gamma_{\text{obs}} > 0$. The preceding argument remains valid even for $\Gamma_{\text{obs}} = 0$ which concludes the proof. For the limiting case $\Gamma_{\text{obs}} = 0$, (4.14) provides information about the asymptotic behaviour. $\square$

5. NUMERICAL EXPERIMENTS

We illustrate the foregoing results for the model parametric elliptic boundary value problem

$$-\text{div}(u\nabla p) = f \quad \text{in } D := [0, 1], \ p = 0 \quad \text{in } \partial D, \tag{5.1}$$

with $f(x) = 100 \cdot x$. The diffusion coefficient is assumed to be affine-parametric, *i.e.*

$$u(x, y) = 0.15 + y_1\psi_1(x) + y_2\psi_2(x),$$

with $J = 2$, $\mathbb{J} = \{1, 2\}$, $\langle u \rangle = 0.15$, $\psi_1(x) = 0.1\sin(\pi x)$, $\psi_2(x) = 0.025\cos(2\pi x)$ and with $y_j \sim \mathcal{U}[-1, 1]$, $j \in \mathbb{J}$.

The forward problem is numerically solved by a finite element method using continuous, piecewise linear shape functions on a uniform mesh with meshwidth $h = 2^{-8}$. The quantity of interest ϕ is assumed to be the solution of the forward problem at the midpoint of the domain and we assume that the solution of the forward problem can be observed at $x = 0.25$ and $x = 0.75$, *i.e.* the observation operator $\mathcal{O}$ consists of K_{obs} system responses at $K_{\text{obs}} = 2$ observation points at $x = 0.25$ and $x = 0.75$. The goal of computation is, for given (noisy) data δ,

$$\delta = \mathcal{G}(u) + \eta,$$

with $\eta \sim \mathcal{N}(0, \Gamma)$ and $\mathcal{G} : X \to \mathbb{R}^{K_{\text{obs}}}$, with $K_{\text{obs}} = 2^{N_K} - 1$, $N_K = 2$, the expectation of the observed solution of the forward model, *i.e.* our aim is to approximate

$$Z' = \int_U \exp(-\Phi(u; \delta))\phi(u)\Big|_{u = \langle u \rangle + \sum_{j=1}^{2} y_j\psi_j} \pi_0(\text{d}y),$$
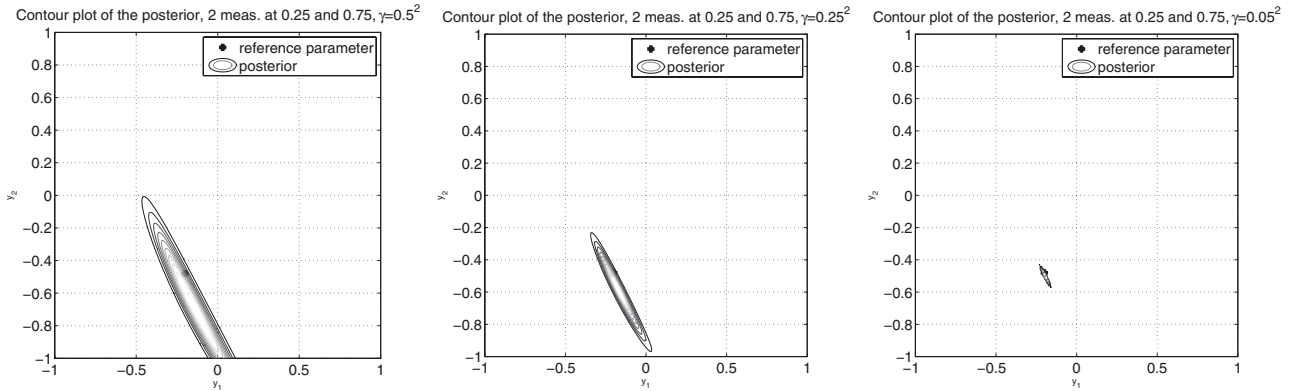
FIGURE 1. Contour plot of the posterior density $\Phi_{\Gamma_{\mathrm{obs}}}$ with isotropic Gaussian observational noise with mean zero and with covariances $\Gamma_{\mathrm{obs}} = 0.5^2 I$ (*left*), $\Gamma_{\mathrm{obs}} = 0.25^2 I$ (*middle*) and $\Gamma_{\mathrm{obs}} = 0.05^2 I$ (*right*). The concentration of the posterior as well as its anisotropy are clearly visible.

with the QoI ϕ given by the solution p at the midpoint $x = 0.5$ and with the normalization constant

$$Z = \int_U \exp\big(-\Phi(u;\delta)\big)\Big|_{u=\langle u\rangle + \sum_{j=1}^2 y_j\psi_j} \pi_0(\mathrm{d}y)\,,$$

so that the expected QoI is Z'/Z. The noise $\eta = (\eta_j)_{j=1,\ldots,K_{\mathrm{obs}}}$ is assumed independent and identically distributed. The concentration effect of the posterior and numerical instability of the adaptive Smolyak quadrature schemes is due to the covariance $\Gamma_{\mathrm{obs}} = \gamma I$ with $\gamma = 0.25^2$.

We observe a strong concentration effect of the posterior density and identify a unique maximizer of the posterior in the interior of the parameter domain, *cf.* Figure 1. Further decrease of the parameter γ controlling the synthetic covariance of the observational noise clearly exhibits the asymptotic behavior according to Theorem 3.1. Figure 1 shows the consistent concentration of the posterior around the reference parameter value used to generate the (synthetic) measurement data.

The first strategy which we will apply to overcome the difficulties arising from the concentration effect due to small observation noise covariance will be the curvature rescaling regularization. *I.e.*, the Smolyak quadrature will be "preconditioned" by shifting the origin to the maximizer of the posterior and by rescaling the integrand using second order information at the extremal point. As Γ_{obs} is isotropic, the maximizer y_0 of the posterior density $\Theta_{\Gamma_{\mathrm{obs}}}(\boldsymbol{y})$ in (2.13) can be computed by minimizing $\frac{1}{2}(\delta - \mathcal{G}(u))^\top(\delta - \mathcal{G}(u))\Big|_{u=\sum_{j=1}^2 y_j\psi_j}$ using a trust-region Quasi-Newton approach with SR1 updates as described for example in [9,10] (see also [31], Appendix C). The trust-region subproblems are iteratively solved by a CG-Steihaug iteration (*e.g.* [28]). The results presented in ([31], Appendix C) ensure (locally superlinear) convergence of symmetric rank-1 (SR1) Hessian updates to the exact Hessian of the Bayesian posterior at MAP. The approximated Hessian H_{SR1} is diagonalized $H_{SR1} = QMQ^\top$ and the integrand is regularized by the *curvature-rescaling* transformation (4.5), *i.e.*

$$\boldsymbol{y} \to \boldsymbol{y}_0 + QM^{-1/2}\Gamma^{1/2}\boldsymbol{z}, \quad \boldsymbol{z} \in \mathbb{R}^{\mathbb{J}}. \tag{5.2}$$

The transformed posterior shown in Figure 2 suggests that the most significant contributions of the posterior can be captured by a quadratic approximation, consistent to the analysis presented in Section 3. By Theorem 4.1, the domain of analyticity after the curvature-rescaling transformation (4.5) is independent of Γ_{obs}.
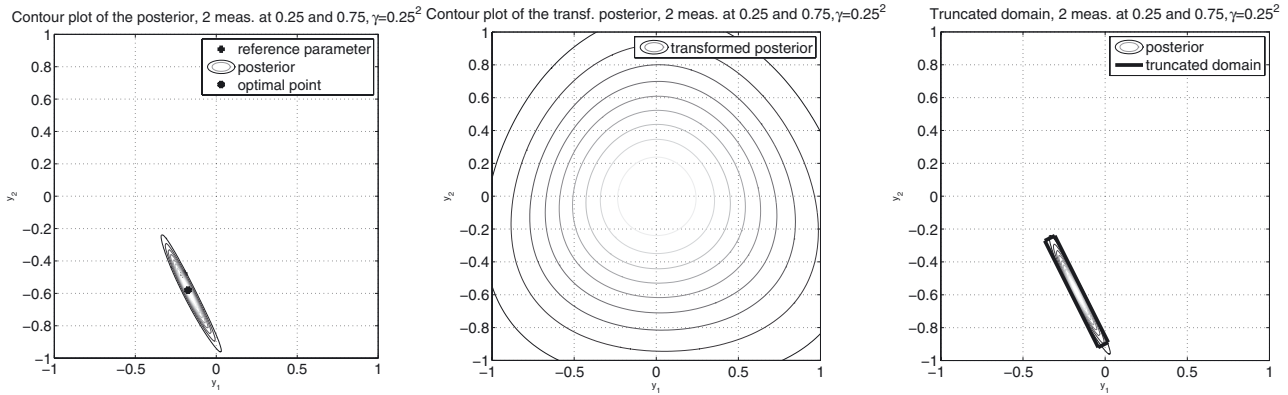
FIGURE 2. Contour plot of the posterior density with observation noise $\Gamma_{\mathrm{obs}} = 0.25^2 I$ (*left*), contour plot of the curvature rescaled posterior (*middle*) and truncated domain of integration (4.10) in the original coordinate system (*right*) with $\tau = 0.16$.

Thus, the preconditioned Smolyak quadrature is expected to converge with rates independent of the observational noise covariance Γ_{obs}.

Comparing the estimated, absolute error curves using the Smolyak approach for the original integrand and the transformed integrand, shown in Figure 3, we observe that the error indicator of the Smolyak algorithm fails. This leads to premature termination of the adaptive Smolyak algorithm, a well known problem of greedy adaptation strategies for integrands concentrated on a small region of the integration domain. The adaptive Smolyak algorithm applied after the reparametrization converges as expected by Theorem 4.1. Comparison to a reference solution computed by a full tensor grid based on Clenshaw−Curtis points of order 14 gives a relative error of $8.28036\mathrm{e}{-4}$.

Next, we discuss the extrapolation method presented in Section 4.3. We consider the synthetic variances

$$\Gamma_k = \gamma_k I \ \text{ with } \ \gamma_k = 2^{-(k-1)} \ \ 1 \le k \le 11. \tag{5.3}$$

The conditional expectation $\mathbb{E}^{\pi^\delta}[\phi]$ is approximated by extrapolation to the observation variance Γ_{obs} of the Bayesian estimates for the synthetic observation noise variances $\Gamma_k = \gamma_k I > \Gamma_{\mathrm{obs}}$. Specifically, we construct a sequence of synthetic noise variances Γ_k and apply the Smolyak quadrature to approximate the sequence $\mathbb{E}_{\Gamma_k}^{\pi^\delta}[\phi]$; we then extrapolate the resulting sequence of Bayesian estimates to the observation noise variance $0 < \Gamma_{\mathrm{obs}} \ll 1$ by the generalized Richardson extrapolation described in Proposition 4.3.

Choosing large, synthetic variances Γ_k regularizes posterior densities, but to estimate the Bayesian quantity by extrapolation, the computation of Bayesian estimates also for small synthetic variances is needed. To avoid the issues of the error estimator caused by the concentration effects, the greedy strategy is enforced to iterate for a prescribed number of iterations. Figure 4 shows the (absolute) error curves of the adaptive Smolyak algorithm with same data δ, for synthetic observational noise variance $\Gamma_k = 2^{-(k-1)}I$, where the error in each iteration is computed by a tensor grid reference solution. Considering the asymptotic expansion (3.5) of the Bayesian estimate

$$\mathbb{E}^{\pi^\delta}[\phi]\frac{Z_\Gamma'}{Z_\Gamma} \sim a_0 + a_1 \Gamma^1 + a_2 \Gamma^2 + \dots$$

with $a_0 = \phi(\boldsymbol{y}_0)$, (generalized) Richardson extrapolation with (5.3) yields Table 1.

The next example is the model parametric elliptic boundary value problem (5.1) with Gaussian prior and with lognormal diffusion coefficient (2.12) defined as $\ln(u(x,y)) = y_1 \psi_1(x) + y_2 \psi_2(x)$, where with $J = 2$, $\mathbb{J} = \{1, 2\}$, $\psi_1(x) = 0.1 \sin(\pi x)$, $\psi_2(x) = 0.025 \cos(2\pi x)$ and with $y_j \sim \mathcal{N}(0, 1)$, $j \in \mathbb{J}$. Under the assumption that the
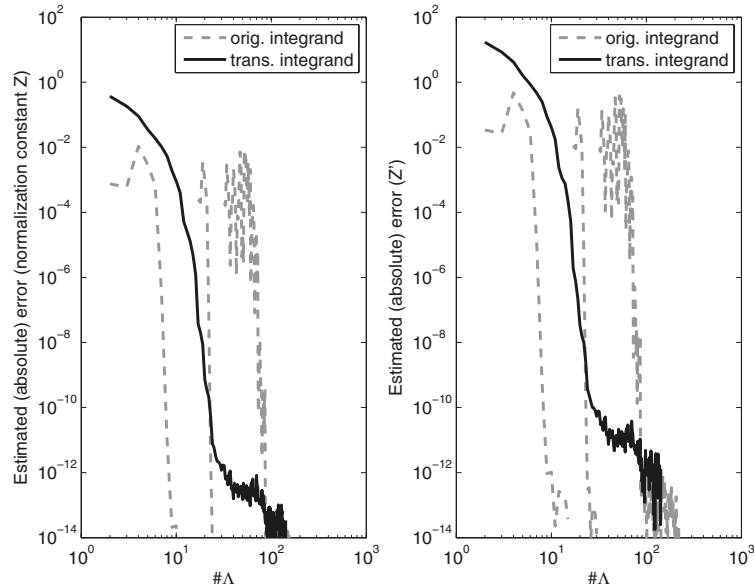
FIGURE 3. Comparison of the estimated (absolute) error curves using the dimension-adaptive Smolyak approach (based on univariate Clenshaw−Curtis points) for the original integrand (gray) and the curvature rescaled integrand (black) for the computation of $Z_{\Gamma_{\mathrm{obs}}}$ (*left*) and $Z'_{\Gamma_{\mathrm{obs}}}$ (*right*) with observational noise $\Gamma_{\mathrm{obs}} = 0.25^2 I$. Posterior reparametrization (5.2) clearly improves convergence behaviour of the adaptive Smolyak quadrature. Nonmonotonic error behaviour due to dimension and order adaptation.
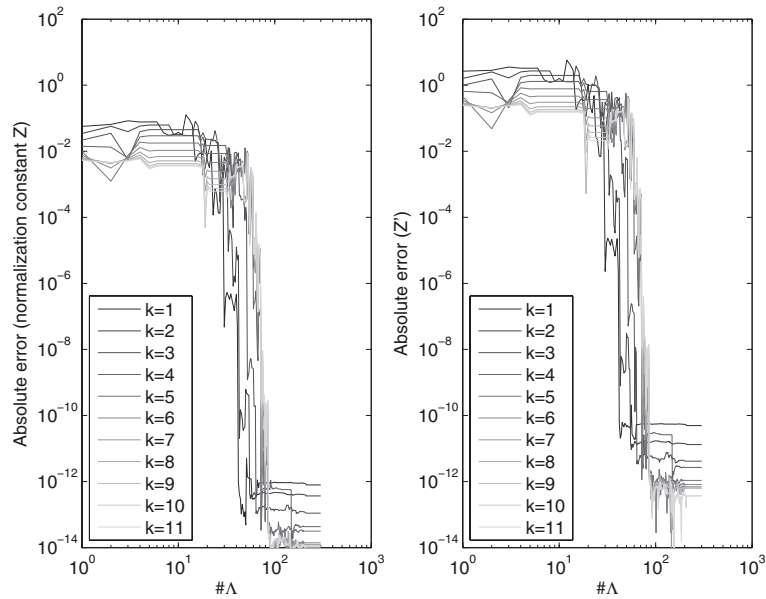

FIGURE 4. Comparison of the (absolute) error curves using dimension adaptive Smolyak quadrature (based on univariate Clenshaw−Curtis points) for the computation of the quantities Z_{Γ_k} (*left*) and Z'_{Γ_k} (*right*) with synthetic noise variances (5.3).

TABLE 1. Estimated relative errors for extrapolation to observation noise variance Γ_{obs}. Reference value from direct overintegration using tensor product quadrature.

# Synthetic Variances $\Gamma_1, \ldots, \Gamma_k$										
1	0.02070011									
2	0.01193666	0.00317320								
3	0.00612788	0.00031911	0.00063226							
4	0.00277697	0.00057394	0.00087162	0.00090582						
5	0.00112845	0.00052006	0.00050210	0.00044932	0.00041888					
6	0.00043617	0.00025612	0.00016814	0.00012043	0.00009850	0.00008817				
7	0.00017238	0.00009141	0.00003651	0.00001770	0.00001086	0.00000803	0.00000676			
8	0.00007256	0.00002726	0.00000588	0.00000150	0.00000042	0.00000008	0.00000004	0.00000010		
9	0.00003256	0.00000743	0.00000082	0.00000010	0.00000001	0.00000001	0.00000001	0.00000001	0.00000001	
10	0.00001531	0.00000195	0.00000013	0.00000003	0.00000002	0.00000002	0.00000002	0.00000002	0.00000002	0.00000002
11	0.00000739	0.00000052	0.00000037	0.00000002	0.00000002	0.00000002	0.00000002	0.00000002	0.00000002	0.00000002	0.00000002

random variables y_j are independent, the prior π_0 is given by the two-dimensional Gaussian measure $\mathcal{N}(0, I)$ on $\mathbb{R}^2$, *i.e.* the parameter domain is $U = \mathbb{R}^2$. Before discussing further details, we remark that the unbounded parameter domain $U = \mathbb{R}^J$ violates the requirements for the Laplace asymptotics in Section 6. The exponential decay of the Gaussian density as $|\boldsymbol{y}| \to \infty$ and nondegeneracy of the potential Φ in (3.1) allow the "localization" of the argument in the proof of Proposition A.3 to a compact subset $\tilde{U} \subset U = \mathbb{R}^2$.

We adopt the setting of the uniform test case, *i.e.* the solution is computed by a finite element method using continuous, piecewise linear ansatz functions on a uniform mesh in the spatial domain D with meshwidth $h = 2^{-8}$ and we assume that the observation operator $\mathcal{O}$ consists of $K_{\text{obs}} = 2$ system responses $o_k(\cdot)$ (being point evaluations at $x = 0.25$ and $x = 0.75$, respectively) and the quantity of interest ϕ is defined as the solution of the forward problem at $x = 0.5$.

For given (noisy) data δ as in (2.3) with $\eta \sim \mathcal{N}(0, \Gamma_{\text{obs}})$ and $\mathcal{G} : X \to \mathbb{R}^{K_{\text{obs}}}$, with $K_{\text{obs}} = 2$, we are interested in the behaviour of the posterior density

$$\Theta(\boldsymbol{y}) = \exp\left(-\Phi_{\Gamma_{\text{obs}}}(u; \delta)\Big|_{u = \sum_{j=1}^2 y_j \psi_j} - \frac{1}{2}\|y\|_2^2 \right), \tag{5.4}$$

where the extra term $\frac{1}{2}\|y\|_2^2$ is due to the Gaussian prior density wr. to $\mathrm{d}\boldsymbol{y}$. The observation noise covariance Γ_{obs} is assumed to be of the form $\Gamma_{\text{obs}} = \gamma I$ with $\gamma = 0.01^2$, *cf.* Figure 5.

To apply the curvature rescaling regularization, we solve the following minimization problem

$$\min_{\boldsymbol{y} \in \mathbb{R}^2} \frac{1}{2}\left((\delta - \mathcal{G}(u))^\top \Gamma_{\text{obs}}^{-1} \left(\delta - \mathcal{G}(u)\Big|_{u = \sum_{j=1}^2 y_j \psi_j} \right) + \|y\|_2^2 \right)$$

by the trust-region Quasi-Newton approach with SR1 updates. The nondegeneracy of S at $\boldsymbol{y}_0$ implies locally superlinear convergence of both, function values and Hessian, of the potential at the extremal point y_0. The Hessian is used to regularize the integrand by the *curvature-rescaling* (5.2). Due to the unbounded parameter domain, for the Gaussian prior, there is no additional approximation error introduced by truncation of parameter domains while rescaling the integrand.

As in the uniform case, curvature rescaling (4.5) results in near quadratic behavior of the posterior density, *cf.* Figure 5, which suggests convergence of the Smolyak algorithm independent of the observational noise covariance Γ_{obs}. Figure 6 confirms that curvature rescaling (4.5) of the integrand functions prior to adaptive Smolyak quadrature restores robust w.r.t. Γ_{obs} convergence, in accordance with Theorem 4.1.

To further reduce the number of quadrature points, we exploit the knowledge of $\boldsymbol{y}_0$ and Theorem 4.1, (4.7), which implies asymptotically, as $\Gamma \to 0$, quadratic behaviour of the rescaled posterior densities. We propose a quadrature scheme in spherical polar coordinates: discretizing the radial coordinate by the Gauss-Laguerre
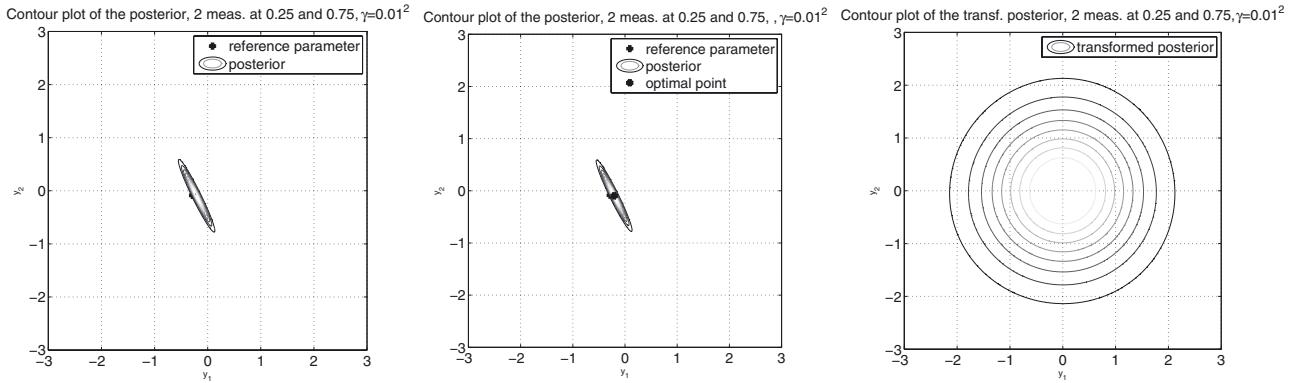
FIGURE 5. Contour plot of the posterior density with observational noise $\Gamma_{\mathrm{obs}} = 0.01^2 I$ (*left*), the minimizer of the potential (*middle*) and contour plot of the transformed posterior (*right*). Asymptotically spherical level-lines of the rescaled posterior density shown in right subfigure are in agreement with Theorem 4.1 and (4.6), (4.7).
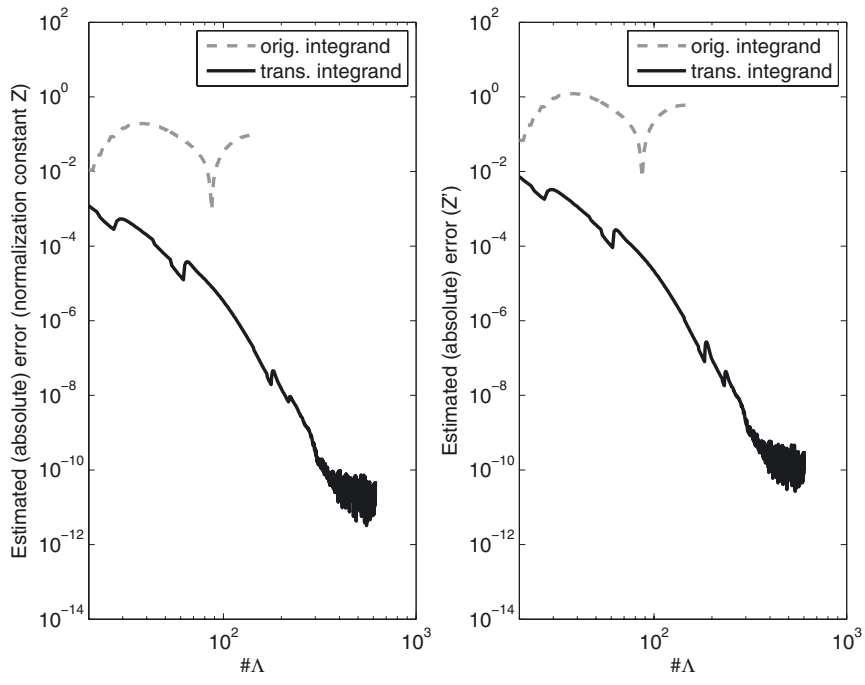

FIGURE 6. Comparison of the estimated (absolute) error using the dimension adaptive Smolyak quadrature (based on univariate Gauss–Hermite points) for the original integrand (gray) and the transformed integrand (black) for the computation of $Z_{\Gamma_{\mathrm{obs}}}$ (*left*) and $Z'_{\Gamma_{\mathrm{obs}}}$ (*right*) with observational noise $\Gamma_{\mathrm{obs}} = 0.01^2 I$. Posterior reparametrization (5.2) clearly improves convergence behaviour of the adaptive Smolyak quadrature. Nonmonotonic error behaviour due to dimension and order adaptation, and the nonnested univariate quadrature points.
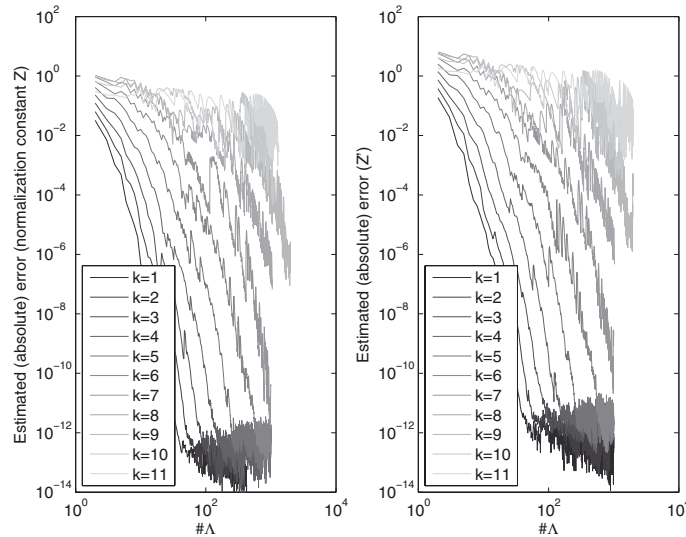
FIGURE 7. Comparison of the estimated (absolute) error using dimension and order adaptive Smolyak quadrature (based on univariate Gauss–Hermite points) for the computation of the quantities Z_{Γ_k} (*left*) and Z'_{Γ_k} (*right*) with synthetic noise variances $\Gamma_k = 2^{-(k-1)}$. Nonmonotonic error behaviour due to dimension and order adaptation, and the nonnested univariate quadrature points.

abscissas of order 2 and the angular coordinate by a trapezoidal rule with 3 grid points, the relative error of the numerical approximation of $\mathbb{E}^{\pi^\delta}[\phi]$ is below $3e-5$ and the quadrature effort is reduced to 6 forward simulations, *cf.* ([31], Fig. 7). We remark that the use of the trapezoidal rule for the angular integral is specific to the case $J = 2$ parameters. For $J \geq 3$, we propose approximation of the angular integral over $\mathbb{S}^{J-1}$ by adaptive Smolyak quadrature based on tensorized Gauss-Jacobi rules in hyperspherical coordinates as detailed in ([31], Appendix D).

Finally, we discuss the extrapolation method based on the sequence of synthetic variances defined in (5.3). Figure 7 indicates the convergence rate of the adaptive Smolyak algorithm with same data δ, for synthetic observational noise variance $\Gamma_k = 2^{-(k-1)}$. As the Smolyak algorithm fails to converge for small synthetic observational noise variances ($\Gamma_7, \ldots, \Gamma_{11}$), only the first six approximated values of the normalization constant Z (corresponding to large observation noise covariance) and the quantity Z' are used for the extrapolation based approach. Figure 7 plots the corresponding values for the quantities Z_k, Z'_k.

Using as reference value the approximation of $\mathbb{E}^{\pi^\delta}[\phi]$ computed by curvature-rescaled, adapative Smolyak quadrature (with absolute error tolerances $1e-10$), the results of the extrapolation based on (3.5) are summarized in Table 2.

Due to the convergence problems of the adaptive Smolyak quadrature caused by the concentration in the posterior densities in the presence of small observation noise covariance Γ_k, we additionally use the curvature-rescaling of the densities to enhance the convergence of the adaptive Smolyak quadrature for the sequence of synthetic noise variances Γ_k. We verify the performance of the extrapolation strategy. Specifically, for each value of the variance, we compute the optimal parameters of the least-squares problem and rescale accordingly. Figure 8 shows the improvement in convergence gained by the "curvature-rescaling" preconditioning of the adaptive Smolyak quadrature. The approximated values of the normalization constant Z_{Γ_k}, $k > 7$ and the quantity Z'_{Γ_k}, $k > 7$ are used to construct the sequence of conditional expectations $\mathbb{E}^{\pi^\delta}_{\Gamma_k}[\phi]$. The numerical results are presented in Table 3.

TABLE 2. Relative errors for generalized Romberg extrapolation to small observation noise variance $\Gamma_{\mathrm{obs}} > 0$.

#	Synthetic Variances $\Gamma_1, \ldots, \Gamma_k$						
1	0.00864275						
2	0.00828207	0.00792139					
3	0.00763215	0.00698223	0.00666917				
4	0.00657289	0.00551363	0.00502409	0.00478908			
5	0.00511237	0.00365186	0.00303127	0.00274658	0.00261041		
6	0.00350889	0.00190542	0.00132327	0.00107927	0.00096812	0.00091514	
7	0.00212986	0.0007508	0.00036607	0.00022921	0.00017254	0.00014688	0.00013468

TABLE 3. Relative errors for generalized Romberg extrapolation to small observation noise variance $\Gamma_{\mathrm{obs}} > 0$.

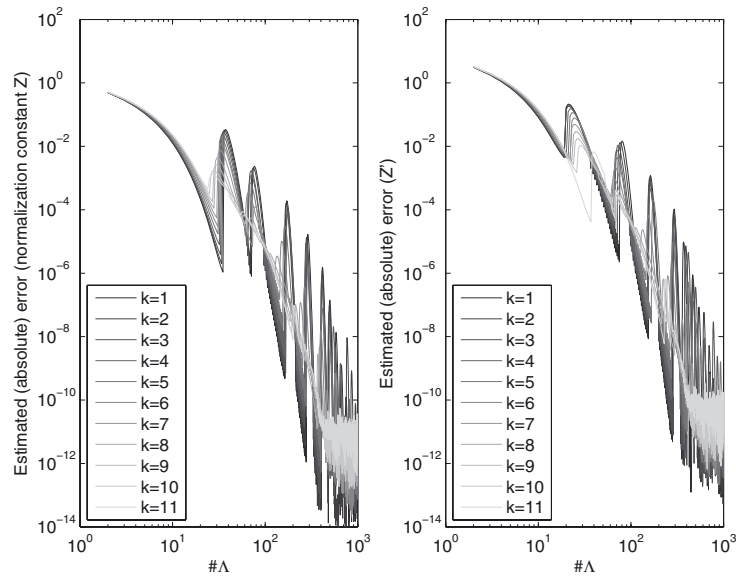| #	Synthetic Variances $\Gamma_1, \ldots, \Gamma_k$									
1	0.00864275									
2	0.00828207	0.00792139								
3	0.00763215	0.00698223	0.00666917							
4	0.00657289	0.00551363	0.00502409	0.00478908						
5	0.00511237	0.00365186	0.00303127	0.00274658	0.00261041					
6	0.00350889	0.00190542	0.00132327	0.00107927	0.00096812	0.00091514				
7	0.00212986	0.00075083	0.00036597	0.00022921	0.0001725	0.00014688	0.00013468			
8	0.00116504	0.00020022	0.00001669	0.00003321	0.00005071	0.00005791	0.00006116	0.00006270		
9	0.00057981	0.00000542	0.00007397	0.00008692	0.00009050	0.00009179	0.00009233	0.00009257	0.00009269	
10	0.00025547	0.00006885	0.00090000	0.00009227	0.00009265	0.00009271	0.00009273	0.00009273	0.00009273	0.00009273
11	0.00009002	0.00007544	0.00007763	0.00007586	0.00007477	0.00007419	0.00007390	0.00007375	0.00007368	0.00007364	0.00007362

FIGURE 8. Comparison of the estimated (absolute) error using the (curvature-preconditioned) adaptive Smolyak approach (Gauss–Hermite points) for the computation of Z_{Γ_k} (*left*) and Z'_{Γ_k} (*right*) for synthetic observation noise variances Γ_k in Table 2.
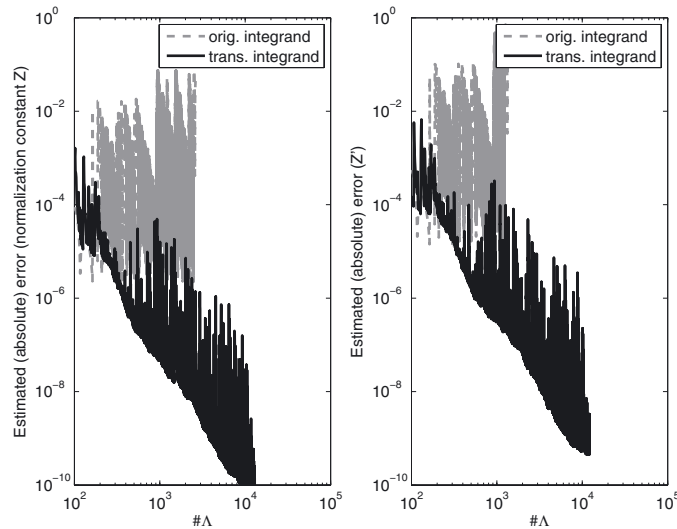
FIGURE 9. Comparison of the estimated (absolute) error using the dimension and order adaptive Smolyak quadrature (based on univariate Gauss–Hermite points) for the original integrand (gray) and the transformed integrand (black) for the computation of $Z_{\Gamma_{\mathrm{obs}}}$ (*left*) and $Z'_{\Gamma_{\mathrm{obs}}}$ (*right*) with observation noise variance $\Gamma_{\mathrm{obs}} = 0.01^2 I$, $J = 5$. Posterior reparametrization (5.2) clearly improves convergence behaviour of the adaptive Smolyak quadrature. Nonmonotonic error behaviour due to dimension and order adaptation, and the nonnested univariate quadrature points.

In summary, rescaling of the posterior densities (2.13), (2.15) in the Bayesian estimate (2.16) based on second order information at the extremum of these (assumed unimodal) posterior densities to "precondition" the adaptive Smolyak quadrature allows for dimension-, and Γ-independent convergence rates of adaptive Smolyak quadrature. In particular, under a dimension truncation and a nondegeneracy assumption, this "curvature-rescaling" of the posterior renders the parametric integrands analytic with Γ-independent domains of analyticity. The adaptive Smolyak quadrature approach is robust in the limit of vanishing observation noise covariance. The extrapolation approach per se does not require the solution of an additional minimization problem and shows satisfactory approximation results in the considered test example, *cf.* Table 3. However, the extrapolation method requires the approximation of the normalization constant Z_Γ and of the quantity Z'_Γ for several large, synthetic variances Γ_k monotonically decreasing to zero. Computing these estimates could be done in parallel. Without curvature-based rescaling, extrapolation in general will encounter numerical stability problems for small observation noise covariances Γ_k, as observed in Figure 7.

To investigate the performance of the preconditioned Smolyak approach when the posterior concentrates only in a (lower-dimensional) subspace (determined, for example, by dominant eigenspaces of the Hessian of $S(\boldsymbol{y}_0)$), we consider Bayesian inversion of the model parametric elliptic boundary value problem (2.9), as defined above, in the 5-parameter setting $\ln(u(x,y)) = \sum_{j=1}^5 \psi_j(x) y_j$ with $\psi_j(x) = 0.1/2^{j-1} \sin(2^{j-1}\pi x)$, $j = 1, 3, 5$, $\psi_j(x) = 0.1/2^{j-1} \cos(2^{j-1}\pi x)$, $j = 2, 4$ and with $y_j \sim \mathcal{N}(0,1)$, $j \in \mathbb{J} = \{1, \ldots, 5\}$.

The eigenvalues of the approximated Hessian of the Bayesian potential by QN with SR1 updates indicate that the posterior concentrates in two coordinates. According to the strategy discussed in Section 4.1, we shift the origin to the maximizer of the posterior density (the minimizer of the potential Φ), rotate the coordinate system and rescale the two integration coordinates to remove the degeneracy of the integrand. The superior performance of the preconditioned, adaptive Smolyak algorithm, compared to its performance applied to the original integrand is shown in Figure 9.

6. Discussion and conclusions

We proposed and analyzed an adaptive, deterministic quadrature approach to Bayesian inversion for high-dimensional, holomorphic-parametric operator equations, extending [29, 30]. It is based on a deterministic representation of the uncertain input data as well as of the Bayesian posterior densities; specifically, as a holomorphic function of possibly countably many parameters. We considered two particular prior measures π_0: uniform prior π_0^{unif} and Gaussian prior π_0^{gauss}. Being product measures of probability measures on the coordinates $y_j \in \boldsymbol{y}$ this implies statistical independence of these coordinates. All methods and results of the present work extend to prior measures π_0 which are absolutely continuous w.r. to π_0^{unif}, with density $\varrho(\boldsymbol{y})$, $i.e.$ $\mathrm{d}\pi_0(\boldsymbol{y}) = \varrho(\boldsymbol{y})\mathrm{d}\pi_0^{\mathrm{unif}}(\boldsymbol{y})$, provided that $\varrho(\boldsymbol{y})$ is independent of Γ and holomorphic w.r. to the coordinates $y_j \in \boldsymbol{y}$. This allows to encode prior knowledge and correlation structure among the coordinates y_j.

We propose the use of QN methods with symmetric low-rank updates to identify computationally MAP points as well as second order information on the Bayesian posterior density at these points. We proposed, based on this computed second order information and a given observation noise covariance $\Gamma_{\mathrm{obs}} > 0$ a notion of *numerical concentration dimension* of the Bayesian posterior. This information is used to effect a change of coordinates which we proved to resolve the concentration of the posterior: the curvature-based coordinate change renders all derivatives of the rescaled Bayesian posterior density bounded independently of the observation noise covariance Γ, so that dimension-adaptive Smolyak quadratures applied to the rescaled posterior converge independently of the observation noise covariance Γ.

We also established, under certain nondegeneracy assumptions of the covariance weighted least-squares potential, an asymptotic expansion of the Bayesian estimate with respect to vanishing observation noise (co)variance Γ_{obs}. We use this asymptotic expansion as basis for an extrapolation of the Bayesian estimate to the limit which reduces estimates for small noise covariances to a few, possibly parallel, computations for larger, synthetic covariances for one given set of data δ. In the case of vanishing observational noise, an important aspect in Bayesian statistics is the study of posterior consistency, *i.e.* the recovery of the truth in the limit. The analysis presented here focuses solely on the numerical behavior of the posterior for fixed observational data and varying noise covariance, with the goal to design methods computing a Bayesian estimate with computational costs independent on the size of the noise covariance. However, the insight gained by the analysis presented here can be useful to study the posterior consistency in the small noise and large observational data limit. The uniform with respect to Γ analyticity of the curvature-rescaled posterior density implies Γ-independent convergence rates also for higher-order, Quasi Monte-Carlo integration methods analyzed in [15].

The present analysis did not account for *discretization errors* in the forward problems which can, in general, not be solved in closed form. Only *discretizations* of the forward problems are computationally accessible. This introduces an additional discretization error into the estimates of Z_Γ, Z'_Γ defined in (2.16). We expect (for well-posed forward equations and stable Petrov-Galerkin discretizations) the conclusions of the present analysis remain valid for sufficiently fine discretizations. We refer to [6] for an analysis of such PG discretizations and of model order reduction on the accuracy of Bayesian estimates, also in the nonlinear case.

Our main results, Theorems 3.1 and 4.1, did not use the affine-parametric structure (2.6) in an essential way; all results will remain valid for general, holomorphic-parametric forward models as considered, for example, in [7].

Generalized Richardson extrapolation in the case of *degenerate asymptotics* (3.7) is not applicable as soon as $\log(\Gamma)$-terms are present, as then even geometric sequences $\{\Gamma_j\}_{j\geq 0}$ of synthetic variances violate the stability condition (4.13) in case that (A.8) holds with $N > 0$. For $N = 0$ in (A.8), however, geometric synthetic observation noise covariance sequences $\{\Gamma_j\}_{j\geq 0}$ will result in a stable extrapolation scheme for the leading term in (3.7), even without explicit knowledge of the exponents $r_k \in \mathbb{Q}$.

Appendix A: Laplace's method

The asymptotic structure of integrals of the form

$$F(\lambda) = \int_U \phi(\boldsymbol{y}) \exp[\lambda S(\boldsymbol{y})]\mathrm{d}\boldsymbol{y} \tag{A.1}$$

as the parameter $\lambda \to \infty$ is obtained by Laplace's method. In (A.1), U is a bounded domain in $\mathbb{R}^J$, with parameter dimension $J < \infty$, and $\boldsymbol{y} = (y_1, \ldots, y_J) \in U$, $\lambda \in \mathbb{R}$ is a parameter and $S(\boldsymbol{y})$ is a sufficiently smooth, real-valued function.

We first consider the nondegenerate case, and address briefly the degenerate case in Section A.3. The proof proceeds *recursive by dimension* with the recursion based on the univariate result from ([16], Chap. II.2.1) for parametric integrand functions; we present this case first.

A.1. Parametric one-dimensional case

For $J = 1$, we consider the asymptotics of integrals of the form

$$F(\lambda, \alpha) = \int_a^b f(x, \alpha) \exp(\lambda S(x, \alpha)) \mathrm{d}x, \quad \lambda \to \infty, \tag{A.2}$$

where $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{R}^k$ is a parameter vector. If the function $S(x, \alpha)$ admits for every fixed parameter $\alpha \in G \subset \mathbb{R}^k$ from some bounded parameter domain G a unique, nondegenerate maximum $x_0(\alpha) \in I := [a, b]$ and if, for every $\alpha \in G$ the point $x_0(\alpha)$ does not approach ∂I, the *Laplace asymptotics hold uniformly (w.r. to $\alpha \in G$)*. Specifically, assume

(A1) in (A.2), $f(x, \alpha)$ and $S(x, \alpha) \in C(\overline{I \times G}) \cap C^\infty(I \times G)$ and S is real-valued for $(x, \alpha) \in I \times G$,
(A2) for every fixed $\alpha \in G$ the function $S(x, \alpha)$ admits a unique maximum $x_0(\alpha) \in I$.
(A3) the maximum $x_0(\alpha)$ is nondegenerate uniformly w.r. to $\alpha \in G$: for every $\alpha \in G$ holds $-S_{xx}(x_0(\alpha), \alpha) \geq \delta_0 > 0$ and for all $\alpha \in G$ holds $x_0(\alpha) \in [a', b'] \subset [a, b]$ for some fixed $a < a' < b' < b$.

Proposition A.1 ([16], Thm. II.2.1). *Assume* (A1)$-$(A3). *Then $F(\lambda, \alpha)$ in* (A.2) *admits the asymptotic expansion*

$$F(\lambda, \alpha) \sim \exp(\lambda S(x_0, \alpha)) \sum_{j=0}^\infty c_j(\alpha) \lambda^{-j-1/2}, \quad \lambda \to \infty, \tag{A.3}$$

where the coefficients c_j, $j = 0, 1, 2, \ldots$ are given by

$$c_j(\alpha) = \frac{\Gamma(j + 1/2)}{(2j)!} \left(\frac{\mathrm{d}}{\mathrm{d}x} \right)^j \left[f(x, \alpha) \left(\frac{2(S(x_0, \alpha) - S(x, \alpha))}{(x - x_0(\alpha))^2} \right)^{-j-1/2} \right]_{x = x_0(\alpha)}.$$

The first term in the asymptotic expansion (A.3) *reads ($O(\cdot)$ uniform w.r. to $\alpha \in G$)*

$$F(\lambda, \alpha) = \left(\frac{2\pi}{-\lambda S_{xx}(x_0(\alpha), \alpha)} \right)^{1/2} \times \exp(\lambda S(x_0(\alpha), \alpha)) \left[f(x_0(\alpha), \alpha) + O(\lambda^{-1}) \right].$$

In the context of lognormal Gaussian models, as considered in Section 2.3.2, there arise improper integrals where the function $S(\boldsymbol{y})$ in (3.1) depends generally (analytically) on Γ at $\Gamma = 0$ (*cf.* (3.10)). In place of (A.2), we consider therefore the asymptotics of

$$F(\lambda) = \int_{-a(\lambda)}^{a(\lambda)} f(x, \lambda) \exp[S(x, \lambda)] \mathrm{d}x, \quad \lambda \to \infty \tag{A.4}$$

where λ, a, f and S are real-valued, smooth functions. We assume that for all λ, $S(\cdot, \lambda)$ has a unique, global maximum at $x_0(\lambda)$ which is nondegenerate, *i.e.*

$$S'_x(x_0(\lambda), \lambda) = 0, \qquad S''_{xx}(x_0(\lambda), \lambda) < 0$$

Proposition A.2 ([16], Thm. II.2.2). *Assume that there exists a real-valued function $\mu(\lambda) > 0$ with $\mu(\lambda) \to \infty$ as $\lambda \to \infty$ such that, as $\lambda \to \infty$, in (A.4) holds*

$$S''_{xx}(x,\lambda) = S''_{xx}(x_0(\lambda),\lambda)[1 + o(1)], \quad f(x,\lambda) = f(x_0(\lambda),\lambda)[1 + o(1)]$$

uniformly w.r. to $x \in U(x_0(\lambda))$ where $U(x_0(\lambda)) := \{x : |x - x_0(\lambda)| \leq \mu(\lambda)|S''_{xx}(x_0(\lambda),\lambda)|^{-1/2}\} \subset \mathbb{R}$. Then,

$$\int_{U(x_0(\lambda))} f(x,\lambda)\exp[S(x,\lambda)]\mathrm{d}x = \left(\sqrt{-\frac{2\pi}{S''_{xx}}}f\exp[S]\right)|_{x=x_0(\lambda)}[1 + o(1)], \quad as \quad \lambda \to \infty.$$

A.2. Nondegenerate case in finite dimension $J > 1$

We assume that ϕ is real-valued (*i.e.* $\mathcal{S} = \mathbb{R}$; all assertions hold verbatim when the QoI $\phi : U \mapsto \mathcal{S}$ takes values in a Banach space $\mathcal{S}$) and uniform prior π_0.

Proposition A.3 (*cf.* [16], Thm. II.4.1).

Consider $F(\lambda)$ as in (A.1) under the assumptions

1. $\phi, S \in C(\overline{U}; \mathbb{R})$ *with U a bounded domain in $\mathbb{R}^J$, $J < \infty$,*
2. $\max\{S(\boldsymbol{y}) : \boldsymbol{y} \in \overline{U}\}$ *is attained only at one isolated point $\boldsymbol{y}_0 \in \mathrm{int}(U)$,*
3. $S \in C^3(B_\varepsilon(\boldsymbol{y}_0))$ *for a closed ball at $\boldsymbol{y}_0$ of radius $\varepsilon > 0$ so small that $B_\varepsilon(\boldsymbol{y}_0) \subset U$,*
4. *the maximum at $\boldsymbol{y}_0$ is nondegenerate.*

Then there holds, as $\lambda \to \infty$, the asymptotic expansion

$$\exp[-\lambda S(\boldsymbol{y}_0)]F(\lambda) \sim \lambda^{-J/2}\sum_{k=0}^{\infty} a_k \lambda^{-k}. \tag{A.5}$$

In particular, as $\lambda \to \infty$ the leading term a_0 in (A.5) has the exact representation

$$F(\lambda) = \exp[\lambda S(\boldsymbol{y}_0)](2\pi/\lambda)^{J/2}\frac{\phi(\boldsymbol{y}_0)}{\sqrt{|\det(S''(\boldsymbol{y}_0))|}}(1 + O(\lambda^{-1})). \tag{A.6}$$

Here, $S''(\boldsymbol{y}_0) \in \mathbb{R}^{J \times J}_{sym}$ denotes the (negative definite) Hessian of $S(\boldsymbol{y})$ at $\boldsymbol{y}_0$.

Proof. The proof is recursively by dimension, using the asymptotic expansion in Proposition A.1 of parametric, univariate integrals. The argument also elucidates the curvature rescaling preconditioning in Section 4.1, so that we indicate here the derivation of (A.6), from ([16], Chap. II.) Choosing $\varepsilon > 0$ sufficiently small, we partition the domain U of integration in (A.1) into $U = B_\varepsilon(\boldsymbol{y}_0) + (U \backslash B_\varepsilon(\boldsymbol{y}_0)) =: B_\varepsilon(\boldsymbol{y}_0) + V_\varepsilon$ and we may write the integral (A.1) accordingly as $F(\lambda) = F_\varepsilon(\lambda) + G_\varepsilon(\lambda)$ with F_ε defined as in (A.1) but with the domain of integration being $B_\varepsilon(\boldsymbol{y}_0)$ in place of U. Using that $\boldsymbol{y}_0$ is the global maximum of the integrand function in $\overline{U}$, there exists $\delta' > 0$ (depending on $\varepsilon > 0$ and on S but independent of $\lambda > 0$) such that $|G_\varepsilon(\lambda)| = O(\exp(\lambda(S(\boldsymbol{y}_0) - \delta')))$ for all $\lambda > 0$. This integal, being of (exponentially) lower asymptotic order as $\lambda \to \infty$, will not contribute to (A.6). Thus the asmptotics of $F(\lambda)$ as $\lambda \to \infty$ in turn is completely determined by $F_\varepsilon(\lambda)$. To prove (A.6), we perform a C^3-diffeomorphic change of variables $\boldsymbol{x} = \varphi(\boldsymbol{y})$ for all $\boldsymbol{y} \in B_\varepsilon(\boldsymbol{y}_0)$ such that a) $\varphi(\boldsymbol{y}_0) = \boldsymbol{y}_0$, b) $\varphi'(\boldsymbol{y}_0) = \mathbf{1}$, c) $\varphi(B_\varepsilon(\boldsymbol{y}_0)) = \widetilde{B_\varepsilon(\boldsymbol{y}_0)} \subset U$ and such that

$$(S \circ \varphi)(\boldsymbol{x}) = S(\boldsymbol{y}_0) + \frac{1}{2}\sum_{j=1}^{J}\mu_j x_j^2. \tag{A.7}$$

Here, $\mu_j < 0$ are the eigenvalues of the Hessian $S''_{\boldsymbol{yy}}$ at $\boldsymbol{y}_0$ (which are negative according to assumptions 2 and 4). The existence of a C^3-diffeomorphism φ with (A.7) (which, like the function $S(\boldsymbol{y})$, is independent of λ) follows from the Morse Lemma by assumption 3. Changing variables in the integral $F_\varepsilon(\lambda)$, gives

$$F_\varepsilon(\lambda) = \exp(\lambda S(\boldsymbol{y}_0)) \int_{\widetilde{B_\varepsilon(\boldsymbol{y}_0)}} \exp\left(\frac{\lambda}{2} \sum_{j=1}^J \mu_j (x_j)^2\right) (\phi \circ \varphi)(\boldsymbol{x}) \left(D_{\boldsymbol{y}}\varphi\right)(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}.$$

Next, we choose $0 < \varepsilon' < \varepsilon$ such that the axiparallel (wr. to the coordinates $\boldsymbol{x}$) cube $C(\boldsymbol{y}_0; \varepsilon') \subset \widetilde{B_\varepsilon(\boldsymbol{y}_0)}$, and we split as before $F_\varepsilon(\lambda) = F_\varepsilon^1(\lambda) + G_\varepsilon^1(\lambda)$. The asymptotics as $\lambda \to \infty$ are dominated by $F_\varepsilon^1(\lambda)$ which takes the form

$$F_\varepsilon^1(\lambda) = \exp(\lambda S(\boldsymbol{y}_0)) \int_{|x_1| < \varepsilon'/2} \exp\left(\lambda \mu_1 \frac{x_1^2}{2}\right) \Upsilon(x_1) \mathrm{d}x_1$$

where, with the notation $\boldsymbol{x}' = (x_2, x_3, \dots, x_J)$,

$$\Upsilon(x_1) := \int_{|\boldsymbol{x}'|_\infty < \varepsilon'/2} \exp\left(\frac{\lambda}{2} \sum_{j=2}^J \mu_j (x_j)^2\right) (\varrho \circ \varphi)(\boldsymbol{x})(D_{\boldsymbol{y}}\varphi)(x_1; \boldsymbol{x}') \mathrm{d}\boldsymbol{x}'.$$

As $\lambda \to \infty$, the Laplace asymptotics Proposition A.1 for the univariate integral apply to $F_\varepsilon^1(\lambda)$. This yields

$$F_\varepsilon^1(\lambda) = \exp(\lambda S(\boldsymbol{y}_0)) \lambda^{-1/2} \Upsilon(y_{0,1})(1 + O(\lambda^{-1})).$$

Iterating this reasoning for $\Upsilon(y_{0,1})$ which is a $J-1$-dimensional integral of the same type $J-1$ times implies (A.5). An analysis of the spherical integral in polar-coordinates finally yields (A.6) (see [16], Prop. 4.1). □

A.3. Degenerate case

The nondegeneracy of $S''(\boldsymbol{y}_0)$ is essential for the validity of the asymptotic expansion (A.5) and for obtaining the explicit formula (A.6). Based on the (generic) compactness of the Hessian in the countably-parametric setting (*cf.* Sect. A.3), for large, finite truncation dimension J, the parametric Hessian will generically degenerate. In the case when assumption 4. in Proposition A.3 does not hold, (A.6) becomes invalid, but a generalization of (A.5) can be established. The result is less explicit than (A.5).

Proposition A.4. *Under assumptions* $1-3$ *of Proposition A.3, there exists* $N \in \mathbb{N}$ *and constants* $a_{kl} \in \mathbb{R}$ *such that for* $\lambda \to \infty$, *for* $F(\lambda)$ *as in* (A.1) *there holds the power-logarithmic asymptotics*

$$\exp(-\lambda S(\boldsymbol{y}_0)) F(\lambda) \sim \sum_{k=0}^\infty \left(\sum_{l=0}^N a_{kl} \lambda^{-r_k} (\log \lambda)^l\right). \tag{A.8}$$

Here, $r_k \in \mathbb{Q}$ *with* $J/2 \le r_0 \le r_1 < \dots < r_s$ *with* $r_s \to \infty$ *as* $s \to \infty$.

For the proof, we refer to ([16], Sect. II.3.4). The derivation and justification of (A.8) is considerably more involved than the classical asymptotics (A.5) as its proof is based on "resolution of singularities" [2] (see also [16], Thm. II.4.3). We also remark that, contrary to (A.5), to date the leading term a_{00} in (A.8) is *not* known in closed form, in general, and the exponents $r_k \in \mathbb{Q}$ in (A.8) depend on J.

APPENDIX B: COMPACTNESS OF THE HESSIAN FOR UNIFORM PRIOR π_0

For uniform prior π_0 on the parameter sequences $\boldsymbol{y}$ with bounded ranges, *i.e.* $U = [-1, 1]^{\mathbb{N}}$, we investigate the Hessian $S''(\boldsymbol{y})$ in (3.1), (3.2). At a critical point $\boldsymbol{y}_0 \in \mathrm{int}(U)$ we have $r'(\boldsymbol{y}_0) = 0$ so that we find, in the case (3.3) of incompatible data, *i.e.* when $\Phi_\Gamma(\boldsymbol{y}_0, \delta) > 0$, since G depends linearly on f

$$S''(\boldsymbol{y}_0) = -r(\boldsymbol{y}_0) r''(\boldsymbol{y}_0) = -r(\boldsymbol{y}_0) \mathcal{O}(G''(\boldsymbol{y}_0)) = -r(\boldsymbol{y}_0) \mathcal{O}(q''(\boldsymbol{y}_0)).$$

For the following result, we assume $\mathbb{J} = \mathbb{N}$, X is a separable Hilbert space and $(\varphi)_{j\in\mathbb{N}}$ is a countable ONB. Then, X is isomorphic to the sequence space $\ell^2(\mathbb{N})$ and we may identify, *via* the ONB $(\varphi)_{j\in\mathbb{N}}$, any bounded, linear operator $B \in \mathcal{L}(\ell^2(\mathbb{N}), \ell^2(\mathbb{N}))$ with the associated bi-infinite matrix $(B_{ij})_{i,j\in\mathbb{N}}$ given by $B_{ij} = \langle \varphi_i, A\varphi_j \rangle_X$. In particular, the Hessian $(D_u^2 q)(u) \in X \times X \mapsto \mathcal{X}$ is associated with the bilinear mapping $\ell^2(\mathbb{N}) \times \ell^2(\mathbb{N}) \mapsto \mathcal{X}$ induced by the bi-infinite matrix $(\partial_{y_i y_j}^2 q(\boldsymbol{y}))_{i,j\in\mathbb{N}} \in \mathcal{X}^{\mathbb{N}\times\mathbb{N}}$.

Proposition B.1. *Assume that the set of uncertainties is contained in a separable Hilbert space X with X-ONB $(\varphi_j)_{j\in\mathbb{N}}$. Assume further that the distributed parameters $u \in X$ admit the representation* (2.5) *with $|y_j| \leq 1$ and $\psi_j = c_j\varphi_j$ where, for some $C, \eta > 0$, there holds the bound (which implies* (2.18)*)*

$$\forall j \in \mathbb{N}: \qquad c_j := \|\psi_j\|_X \leq Cj^{-1-\eta}. \tag{B.1}$$

Then, for every $\boldsymbol{y} \in U$, and for every observation functional $\mathcal{O}(\cdot) \in \mathcal{X}'$, the bi-infinite matrix $(\mathcal{O}((\partial_{y_i y_j}^2 q)(\boldsymbol{y})))_{i,j\in\mathbb{N}}$ corresponding to the second differential $(D_u^2 \mathcal{G})(u)$ of the uncertainty-to-observation map with u as in (2.5) *induces a compact operator on $\ell^2(\mathbb{N})$.*

Proof. Fix $\boldsymbol{y} \in U$ arbitrary. The assumed X orthonormality of the basis $(\varphi_j)_{j\in\mathbb{N}}$ of X and the (isometric) identification of X with $\ell^2(\mathbb{N})$ implies that it remains to verify that the bi-infinite Hessian of the parametric uncertainty-to-observation map $H(\boldsymbol{y}) := (\mathcal{O}((\partial_{y_i y_j}^2 q)(\boldsymbol{y})))_{i,j\in\mathbb{N}}$ can be approximated, in norm, by a sequence $(H^J(\boldsymbol{y}))_{J\in\mathbb{N}}$ of matrices of finite rank J.

Since there holds, for every $B \in \mathcal{L}(\ell^2(\mathbb{N}), \ell^2(\mathbb{N}))$ and for every $x \in \ell^2(\mathbb{N})$, $\|Bx\|_{\ell^2(\mathbb{N})} \leq \|B\|_F \|x\|_{\ell^2(\mathbb{N})}$ with the "Frobenius-norm" $\| \circ \|_F$ given by $\|B\|_F^2 = \sum_{i,j} |B_{ij}|^2$ (which majorizes the induced spectral-norm), it is sufficient to approximate $H(\boldsymbol{y})$ in $\| \circ \|_F$ by a sequence $\{H^J\}_{J\in\mathbb{N}} \subset \mathcal{L}(\ell^2(\mathbb{N}), \ell^2(\mathbb{N}))$ of operators whose ranks are bounded by J. To this end, we choose H^J as finite sections of the bi-infinite matrix H, *i.e.*

$$H_{ij}^J(\boldsymbol{y}) := \begin{cases} H_{ij}(\boldsymbol{y}) = (\partial_{y_i y_j}^2 \mathcal{G})(\boldsymbol{y}) = \mathcal{O}((\partial_{y_i y_j}^2 q)(\boldsymbol{y})), & 1 \leq i,j \leq J \\ 0 & \text{else} \end{cases}$$

and estimate, for $\boldsymbol{y} \in U$,

$$\|H(\boldsymbol{y}) - H^J(\boldsymbol{y})\|_F^2 = \sum_{i,j>J} |(H_{ij} - H_{ij}^J)(\boldsymbol{y})|^2 \leq \|\mathcal{O}\|_{\mathcal{X}'}^2 \sum_{i,j>J} \|\partial_{y_i y_j}^2 q(\boldsymbol{y})\|_{\mathcal{X}}^2.$$

Using that there exists $C > 0$ such that with c_j as in (B.1) holds

$$\sup_{\boldsymbol{y}\in U} \|\partial_{y_i y_j}^2 q(\boldsymbol{y})\|_{\mathcal{X}} \leq C c_i c_j \quad \forall i,j \in \mathbb{N}$$

we find with the assumption (B.1) on the size of c_j that

$$\|H(\boldsymbol{y}) - H^J(\boldsymbol{y})\|_F^2 \leq C^2 \|\mathcal{O}\|_{\mathcal{X}'}^2 \sum_{i,j>J} i^{-2(1+\eta)} j^{-2(1+\eta)}.$$

We set $i = r\cos\theta$, $j = r\sin\theta$ and majorize the double sum by a Riemann integral,

$$\|H(\boldsymbol{y}) - H^J(\boldsymbol{y})\|_F^2 \lesssim \|\mathcal{O}\|_{\mathcal{X}'}^2 \int_{r=J}^{\infty} \int_{\theta=0}^{\pi/2} r^{-2-2\eta} \left(\frac{1}{(\sin\theta)^{2+2\eta}} + \frac{1}{(\cos\theta)^{2+2\eta}} \right) r\,dr\,d\theta \lesssim \|\mathcal{O}\|_{\mathcal{X}'}^2 \int_{r=J}^{\infty} r^{-1-2\eta}\,dr$$

for every $\eta > 0$, which implies the bound

$$\sup_{\boldsymbol{y}\in U} \|H(\boldsymbol{y}) - H^J(\boldsymbol{y})\|_F \lesssim \|\mathcal{O}\|_{\mathcal{X}'} J^{-\eta} \quad J \to \infty. \tag{B.2}$$

$\square$

Remark B.2. Condition (B.1) implies that the sequence $c = (c_j)_{j\geq 1} \in \ell^p(\mathbb{N})$ for $1/(1 + \eta) < p < 1$, *i.e.* for $\eta < 1/p - 1$ which, for small $\eta > 0$, is close to the minimal condition $p = 1$ for the parametrization (2.5) to remain meaningful. Compare to condition (2.8) in (iii) of the affine perturbation Assumption 2.1. We also note that the truncation $H^J(\boldsymbol{y})$ of the Hessian corresponds to J-term truncation of the uncertainty parametrization (2.5). When (2.5) is obtained by a Karhunen−Loève expansion, therefore, H^J contains the curvature information of the forward mapping restricted to its J principal components with (B.2) providing a quantitative bound on the truncation error, resp. on the curvature information contained in the omitted part $H - H^J$. Also note the analogy of the bound (B.2) with (2.17), (2.18).

Remark B.3. Compactness of the Hessian has been observed to play a crucial role for the efficiency of computational Bayesian inversion in several applications recently; we refer to ([26], Sects. 4.3 and 4.4) for numerical evidence and to [3] for an analogous compactness result in Bayesian shape inversion problems in two space dimensions, for acoustic scattering; both these applications are covered by Proposition B.1.

References

[1] S. Agapiou, O. Papaspiliopoulos, D. Sanz-Alonso and A.M. Stuart, Importance Sampling: Computational Complexity and Intrinsic Dimension. Preprint `arXiv:1511.06196v1` (2015).

[2] M. Atyiah, Resolution of Singularities and Division of Distributions. *Comm. Pure Appl. Math.* **23** (1970) 145–150.

[3] T. Bui-Thanh and O. Ghattas, Analysis of the Hessian for Inverse Scattering Problems. Part II: Inverse Medium Scattering of Acoustic Waves. *Inverse Problems* **28** (2012) 055002.

[4] T. Bui-Thanh, C. Burstedde, O. Ghattas, J. Martin, G. Stadler and L.C. Wilcox, Extreme-scale UQ for Bayesian Inverse Problems Governed by PDE. ACM/IEEE Supercomputing SC12, Gordon Bell Prize Finalist. Salt Lake City, Utah (2012).

[5] J. Charrier, Strong and Weak Error Estimates for the Solutions of Elliptic PDEs with random coefficients. *SIAM J. Numer. Anal.* **50** (2013) 216–246.

[6] P. Chen and Ch. Schwab, Sparse-Grid, Reduced-Basis Bayesian Inversion: Nonaffine-Parametric Nonlinear Equations. *J. Comput. Phys.* **316** (2016) 470–503.

[7] A. Chkifa, A. Cohen and C. Schwab, Breaking the Curse of Dimensionality in Sparse Polynomial Approximation of Parametric PDEs. *J. Math. Pures Appliquées* **103** (2015) 400–428.

[8] A. Cohen, R. DeVore and C. Schwab, Analytic Regularity and Polynomial Approximation of Parametric and Stochastic Elliptic PDEs. *Anal. Appl.* **9** (2011) 1–37.

[9] A. Conn, N. Gould and P. Toint, Global Convergence of a Class of Trust Region Algorithms for Optimization with Simple Bounds. *SIAM J. Numer. Anal.* **25** (1988) 433–460.

[10] A. Conn, N. Gould and P. Toint, Convergence of Quasi-Newton Matrices Generated by the Symmetric Rank-One Update. *Math. Program.* **50** (1991) 177–195.

[11] T. Cui, J. Martin, Y.M. Marzouk, A. Solonen and A. Spantini, Likelihood-Informed Dimension Reduction for Nonlinear Inverse Problems. *Inverse Problems* **30** (2014) 114015.

[12] T. Cui, K.J.H. Law and Y.M. Marzouk, Dimension-Independent Likelihood-Informed MCMC. *J. Comput. Phys.* **304** (2016) 109–137.

[13] M. Dashti and A.M. Stuart, The Bayesian Approach to Inverse Problems. *Handbook of Uncertainty Quantification*, edited by R. Ghanem, D. Higdon and H. Owhadi. Springer (2015).

[14] M. Dashti, K. Law, A.M. Stuart and J. Voss, MAP Estimators and their Consistency in Bayesian Nonparametric Inverse Problems. *Inverse Problems* **29** (2013) 095017.

[15] J. Dick, T.Q. LeGia and Ch. Schwab, Higher Order Quasi Monte Carlo Integration for Holomorphic, Parametric Operator Equations. *J. Uncertainty. Quantification* **4** (2016) 48–79.

[16] M.V. Fedoryuk, Asimptotika: integraly i ryady. Spravochnaya Matematicheskaya Biblioteka. [Mathematical Reference Library] (in russian). Nauka Publ., Moscow, Russia (1987).

[17] R. Fletcher, *An Overview of Unconstrained Optimization in Algorithms for Continuous Optimization*, edited by E. Spedicato. Springer, The Netherlands (1994) 109–14.

[18] R.P. Ge and M.J.D. Powell, The Convergence of Variable Metric Matrices in Unconstrained Optimization. *Math. Program.* **27** (1983) 123–143.

[19] M.B. Giles and N.A. Pierce, An Introduction to the Adjoint Approach to Design. *Flow, Turbulence and Combustion* **65** (2000) 393–415.

[20] M. Girolami and B. Calderhead, Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *J. R. Stat. Soc.: Series B* **73** (2011) 123–214.

[21] M. Hansen and C. Schwab, Analytic Regularity and Best N-Term Approximation of High-Dimensional Parametric Initial Value Problems. *Vietnam J. Math.* **41/2** (2012) 181–215,.

[22] V.H. Hoang and C. Schwab, Analytic Regularity and Polynomial Approximation of Stochastic, Parametric Elliptic Multiscale PDEs. *Anal. Appl.* **11** (2013) 1350001.

[23] J. Kaipio and E. Somersalo, Statistical and Computational Inverse Problems. Vol. 160 of *Appl. Math. Sci.* Springer (2005).

[24] B.T. Knapik, A. W. van der Vaart and J.H. van Zanten, Bayesian Inverse Problems with Gaussian Priors. *Ann. Stat.* **39** (2011) 2626–2657.

[25] K.J.H. Law, Proposals which speed up Function-Space MCMC. *J. Comput. Appl. Math.* **262** (2014) 127–138.

[26] J. Martin, L.C. Wilcox, C. Burstedde and O. Ghattas and A Stochastic, Newton MCMC Method for Large-Scale Statistical Inverse Problems with Application to Seismic Inversion. *SIAM J. Sci. Comput.* **34** (2012) A1460–A1487.

[27] T.A. Moselhy and Y.M. Marzouk, Bayesian Inference with Optimal Maps. *J. Comput. Phys.* **231** (2012) 7815–7850.

[28] J. Nocedal and S.J. Wright, *Numerical Optimization*. Springer, 2nd Edition (1999).

[29] C. Schillings and Ch. Schwab, Sparse, adaptive Smolyak Quadrature for Bayesian inverse problems. *Inverse Problems* **29** (2013) 065011.

[30] C. Schillings and Ch. Schwab, Sparsity in Bayesian inversion of parametric operator equations. *Inverse Problems* **30** (2014) 065007.

[31] C. Schillings and Ch. Schwab, Scaling Limits in Computational Bayesian Inversion. Report 2014-26, Seminar for Applied Mathematics. ETH Zürich, http://www.sam.math.ethz.ch (2014).

[32] C. Schillings, M. Sunnåker, J. Stelling and Ch. Schwab, Efficient Characterization of Parametric Uncertainty of Complex (Bio)chemical Networks. *PLOS Comp. Biol.* (2015), http://dx.doi.org/10.1371/journal.pcbi.1004457.

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

[34] A. Sidi, On a generalization of the Richardson extrapolation process. *Numer. Math.* **57** (1990) 365–377.

[35] A.M. Stuart, Inverse problems: a Bayesian approach. *Acta Numerica* **19** (2010) 451–559.

[36] L. Tierney and J.B. Kadane, Accurate Approximations for Posterior Moments and Marginal Densities. *J. Am. Stat. Assoc.* **81** (1986) 82–86.