

Scaling Limits in Computational Bayesian Inversion

C. Schillings and Ch. Schwab

Research Report No. 2014-26
October 2014

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

SCALING LIMITS IN COMPUTATIONAL BAYESIAN INVERSION

CLAUDIA SCHILLINGS¹ AND CHRISTOPH SCHWAB²

Abstract. Computational Bayesian inversion of operator equations with distributed uncertain input parameters is based on an infinite-dimensional version of Bayes' formula established in [31] and its numerical realization in [27,28]. Based on the sparsity of the posterior density shown in [29], dimension-adaptive Smolyak quadratures afford higher convergence rates than MCMC in terms of the number M of solutions of the forward (parametric operator) equation [27,28]. The error bounds and convergence rates obtained in [27,28] 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 related to MAP estimators obtained from Tikhonov regularized least-squares functionals. Non-intrusive identification of concentration points and curvature information of the posterior density at these points by Quasi-Newton (QN) minimization of the Bayesian potential with SR1 updates from [7,14] is proposed. Two Bayesian estimation algorithms with robust in Γ performance are developed: first, dimension-adaptive Smolyak quadrature from [27,28] combined with a novel, curvature-based reparametrization of the parametric Bayesian posterior density near the (assumed unique) global maximum of the posterior density and, second, extrapolation to the limit of vanishing observation noise variance. For either approach, we prove convergence with rates independent of the number of parameters as well as of the observation noise variance Γ . The generalized Richardson extrapolation to the limit $\Gamma \downarrow 0$ due to A. Sidi [30] is justified by establishing asymptotic expansions wr. to $\Gamma \downarrow 0$ of the Bayesian estimates. Numerical experiments are presented which indicate a performance independent of Γ on the curvature-rescaled, adaptive Smolyak algorithm.

1991 Mathematics Subject Classification. 65M32, 65N35, 65K10.

Oct 12, 2014.

1. INTRODUCTION

The problem of efficient computational methods of Bayesian inversion of partial differential equations has attracted considerable attention in statistics in the context of "large data", and in computational science in recent years; we refer to [21,23,31] and the references there. 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", ie. with random field inputs. Here, upon choosing an (unconditional) base of the space X of admissible uncertain input data, the Bayesian

Keywords and phrases: Bayesian inverse problems, Parametric operator equations, Smolyak quadrature, Sparsity, Non-Gaussian prior, Quasi-Newton methods, SR1 update, Posterior reparametrization, Richardson extrapolation

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

inversion and estimation formally becomes an infinite-dimensional, parametric deterministic *quadrature* problem, with quadrature understood against 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 [27,28] and the references there. 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 is key to efficient computational Bayesian inversion. For example, in MCMC methods it appears in connection with the so-called “burn-in” period of the samplers.

To develop efficient numerical treatment of concentrating posterior densities in the sparse, dimension-adaptive deterministic quadrature methods from [27,28] is the purpose of the present paper. We propose to numerically identify loci of posterior concentration in parameter spaces by Quasi-Newton methods with symmetric rank-1 (SR1) update, [7] applied to the (co)variance-weighted Bayesian misfit functional. These methods afford locally *superlinear convergence of iterates and Hessians*. We prove that the second order information on the Bayesian posterior density at the MAP point obtained from these methods upon regular termination allows for posterior “desingularization” via *curvature-based reparametrization* of the posterior which, in conjunction with dimension-adaptive Smolyak quadratures as in [27,28], affords convergence of the numerical Bayesian estimates which is robust with respect to the observation noise variance Γ . We also establish novel asymptotic expansions of the Bayesian estimate with respect to observation noise variance Γ . The asymptotic expansions developed in this paper generalize the results in [22] for the linear Gaussian case to nonlinear forward problems.

The outline of this paper is as follows: in §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 base. To this end, we recapitulate the formalism of [27,28] whereby the Bayesian estimate can be expressed as a formally infinite-dimensional integral wr. to the Bayesian posterior density. §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, cp. [9, Sec.3]) limit which equals the QoI evaluated at a so-called MAP point \mathbf{y}_0 in the (possibly infinite-dimensional) parameter domain. §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 . §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 §3 also justify a new (generalized) “extrapolation to the limit” approach developed in §4.3, which can be combined with curvature rescaling. The required second order information at a MAP point \mathbf{y}_0 can be obtained “for free” by Quasi-Newton methods with symmetric updates (reviewed, for the reader’s convenience, in Appendix C in §10) which exhibit superlinear convergence also of 2nd derivative information in work that is polynomially-scaling w.r. to the dimension J of the parameter space (under appropriate assumptions). 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 [30]. In §5 numerical experiments are presented which confirm the theoretical results. Appendix A in §8 collects known results on Laplace’s method for the asymptotic analysis of integrals depending on a parameter, and Appendix B in §9 establishes the compactness of the Hessian, for bounded parameter ranges and uniform prior π_0 . Appendix C in §10 collects known information on QN methods with symmetric rank-1 (SR1 for short) updates, and Appendix D in §11 lists hyper-spherical polar coordinates.

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 [10,31]. We admit forward models of possibly countably-parametric operator equations depending on an uncertain datum u taking values in a separable Banach space X . Assuming an unconditional base $\{\psi_j\}_{j \in \mathbb{J}}$ of X , we parametrize the uncertain datum $u \in X$ via a (possibly countably infinite) sequence $\mathbf{y} = (y_j)_{j \in \mathbb{J}}$ of parameters.

2.1. Bayesian Inversion in Infinite Dimension

By $G : X \rightarrow R$ we denote a “forward” response map from the separable Banach space X of uncertain, distributed parameters u into some state (Banach) space \mathcal{X} of responses. We assume given observation data δ taking values in a second (Banach) space Y (which will be assumed, throughout the present paper, to be isomorphic to \mathbb{R}^K). We equip X , Y and \mathcal{X} with norms $\|\cdot\|_X$, $\|\cdot\|_Y$ and with $\|\cdot\|_{\mathcal{X}}$, respectively. Consider the operator equation with uncertain datum $u \in X$: given $f \in \mathcal{Y}'$, find $q \in \mathcal{X}$ such that

$$\text{Given } u \in X, \text{ find } q \in \mathcal{X} : \quad A(u; q) = f \quad (1)$$

where the uncertain operator $A(u; \cdot) \in \mathcal{L}(\mathcal{X}, \mathcal{Y}')$ is assumed to be boundedly invertible, at least locally for uncertain input data u sufficiently close to a nominal input $u_0 \in X$, ie. 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 , ie. the solution $q \in \mathcal{X}$ of the forward problem (1) is the image of a map $G : X \times \mathcal{Y}'$ to \mathcal{X} ,

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

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 a bounded, linear *observation functional* $\mathcal{O}(\cdot) : \mathcal{X} \rightarrow Y = \mathbb{R}^K$. Specifically, $\mathcal{O}(\cdot)$ is a *bounded linear observation operator* on the space \mathcal{X} of system responses, ie. $\mathcal{O} \in (\mathcal{X}')^K$, the dual space of the space \mathcal{X} of system responses. Note that we assume that there are a finite number K of observables, constituting (for each realization of u) a measurement vector taking values in $Y = \mathbb{R}^K$, where we equip \mathbb{R}^K with the Euclidean norm, denoted by $|\cdot|$.

In this setting, we wish to predict *computationally* an expected (under the Bayesian posterior) system response of the QoI, conditional on given, noisy measurement data δ . Specifically, we assume the data δ to consist of observations of QoI system responses corrupted by additive, centered Gaussian noise, ie.

$$\delta = \mathcal{O}(G(u)) + \eta \in Y \quad (2)$$

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, ie. a random vector $\eta \sim \mathcal{N}(0, \Gamma)$, for a positive definite covariance operator Γ on \mathbb{R}^K (ie., a symmetric, positive definite $K \times K$ covariance matrix Γ) which we assume to be known. We define the *uncertainty-to-observation map* $\mathcal{G} = \mathcal{O} \circ G : X \rightarrow Y = \mathbb{R}^K$ so that

$$\delta = \mathcal{G}(u) + \eta = (\mathcal{O} \circ G)(u) + \eta : X \mapsto L_{\Gamma}^2(Y; g_{\Gamma}) \quad (3)$$

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 (eg. [10, Thm. 3.3]), we define the least-squares functional (also referred to as “potential” in [10, 31], or as “mismatch” resp. “misfit” functional in the literature, eg. [2]) $\Phi_{\Gamma} : X \times Y \rightarrow \mathbb{R}$ by $\Phi_{\Gamma}(u; \delta) = \frac{1}{2}R(u)^{\top} \Gamma^{-1}R(u)$, where the residual R at data δ and at an instance $u \in X$ of the uncertain parameter is

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

For observation noise (co)variance $\Gamma > 0$, the Bayesian potential $\Phi_\Gamma(u; \delta) : X \times Y \rightarrow \mathbb{R}$ takes the form

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

In [31], [10, Section 3.2], 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 the prior measure π_0 on the space X of uncertain parameters u , the posterior distribution μ^δ of the random variable $\delta|u$ is absolutely continuous with respect to the prior π_0 . In particular, then, the Radon-Nikodym derivative of the Bayesian posterior wr. to the prior measure admits a bounded density Θ_Γ wr. to the prior π_0 on X .

2.2. Uncertainty Parametrization

We parametrize the uncertain datum u in the forward equation (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 $\mathbb{J} = \{1, 2, \dots, J\}$ 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 $\mathbb{J} = \mathbb{N}$. Then, we assume that there exists a Schauder base $\{\psi_j\}_{j \in \mathbb{J}}$ of X such that, for some ‘‘nominal’’ value $\langle u \rangle \in X$ of the uncertain datum u , and for some coefficient sequence $\mathbf{y} = (y_j)_{j \in \mathbb{J}}$ (uniquely associated with $u - \langle u \rangle \in X$) the uncertainty u is parametrized by the sequence \mathbf{y} in the sense that there holds

$$u = u(\mathbf{y}) := \langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j \in X \quad (5)$$

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 (5) is deterministic. In the case of a uniform prior, in order to place (2), (5) into the (probabilistic) Bayesian setting of [31], 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 countable cartesian product of sets with the product sigma-algebra $\mathcal{B} = \otimes_{j \in \mathbb{J}} \mathcal{B}^1$, 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 $\pi_0 = \otimes_{j \in \mathbb{J}} \frac{1}{2} \lambda^1$ with λ^1 denoting the Lebesgue measure on $[-1, 1]$. Then (U, \mathcal{B}, π_0) becomes (as countable product of probability spaces) a probability space on the set U of all sequences of coefficient vectors \mathbf{y} in the uncertainty parametrization (5). The uncertain datum u in (5) becomes a random field, with π_0 charging the possible realizations of u . As indicated in [5, 27, 29], *analyticity of uncertainty parametrization (5) with respect to the parameter sequence \mathbf{y}* can be used to derive sparsity results for this posterior. In the case of a Gaussian prior, a parametrization of the form (5) of the Gaussian random field can be obtained via a Karhunen-Lo'ève expansion, where $(y_j)_{j \in \mathbb{J}}$ is an iid sequence of $\mathcal{N}(0, 1)$ random variables. We refer to [31, Chapter 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 (1) is parametrized as in (5). Accordingly, and with slight abuse of notation we write $R(\mathbf{y})$ and $\mathcal{G}(\mathbf{y})$ in place of $R(u(\mathbf{y}))$ and of $\mathcal{G}(u(\mathbf{y}))$, respectively.

2.3. Forward Models

We recapitulate classes of abstract, countably-parametric operator equations considered in [27, 28]. 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 $A : \mathcal{X} \rightarrow \mathcal{Y}'$. 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 $\mathbf{y} = (y_j)_{j \in \mathbb{J}} \in U$ of (possibly countably many) parameters results in a parametric operator equation of the form

$$A(\mathbf{y}) = A_0 + \sum_{j \in \mathbb{J}} y_j A_j \in \mathcal{L}(\mathcal{X}, \mathcal{Y}'). \quad (6)$$

Here, either $\mathbb{J} = \{1, 2, \dots, J\}$ 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 §2.4 ahead.

In (6), $\mathbf{y} = (y_j)_{j \in \mathbb{J}}$ can be, for example, an i.i.d 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” $A_0 = A(0)$. Affine parameter dependences (6) result when the unknown u in (5) is modelled as random field via its Karhunen-Loève expansion in X (or in a closed subspace $X' \subset X$).

In order for the sum in (6) to converge, 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 (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

$$\sum_{j \in \mathbb{J}} b_j \leq \kappa < 1, \quad \text{where } b_j := \|A_0^{-1} A_j\|_{\mathcal{L}(\mathcal{X}, \mathcal{X})}, \quad (7)$$

- (3) (*p* summability) For some $0 < p < 1$, the operators B_j are *p*-summable, in the sense that with the sequence $b = (b_j)_{j \in \mathbb{J}}$ as in (7) holds

$$\|b\|_{\ell^p(\mathbb{J})}^p = \sum_{j \in \mathbb{J}} b_j^p < \infty. \quad (8)$$

Condition (7) (and, hence, Assumption 2.1) is sufficient for the bounded invertibility of $A(\mathbf{y})$, uniformly with respect to the parameter sequence $\mathbf{y} \in U = [-1, 1]^{\mathbb{J}}$. The next result from [28] makes this precise.

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

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

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

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

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

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

The forward maps $q : U \rightarrow \mathcal{X}$ and $\mathcal{G} : U \rightarrow \mathbb{R}^K$ are globally Lipschitz and admit analytic continuations wr. to the parameters y_j into the complex domain. Specifically (see [29, Lemma 3.3, Theorem 3.4]) if q and \bar{q} are solutions

of (1) with the same right hand side f with operators $A(\mathbf{y})$ and $A(\mathbf{y}')$, respectively, then the forward solution map $\mathbf{y} \rightarrow q(\mathbf{y}) = (A(\mathbf{y}))^{-1}f$ is Lipschitz as a mapping from U into \mathcal{X} , ie. there exists a constant $C > 0$ (depending only on κ and π_0 in Assumption 2.1) such that for every $\mathbf{y}, \tilde{\mathbf{y}} \in U$ holds

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

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

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

The prototypical example for (6) is the linear, elliptic diffusion problem

$$-\nabla \cdot (u(x, \mathbf{y}) \nabla q) = f \quad \text{in } D, \quad q(\cdot, \mathbf{y})|_{\partial D} = 0. \quad (9)$$

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

$$u(x, \mathbf{y}) = a(x, \mathbf{y}) := \bar{a}(x) + \sum_{j \geq 1} y_j \psi_j(x) \quad (10)$$

where $\bar{a}, \psi_j \in C^{0,\alpha}(\bar{D})$ for some $0 < \alpha < 1$, and $|y_j| \leq 1$ so that $\mathbf{y} = (y_j)_{j \geq 1} \in U$. Convergence of (10) for $\mathbf{y} \in U$ is ensured by assuming that for some $0 < p \leq 1$

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

ie. we have (5) with $X = C^{0,\alpha}(\bar{D})$. Uniform (with respect to the parameter vector \mathbf{y}) inf-sup conditions are implied by uniform ellipticity of (9). This, in turn, is ensured by imposing that there exist constants $\mu_0 > 0$ and $0 < \kappa < 1$ such that

$$0 < \mu_0 \leq \operatorname{ess\,inf}_{x \in D} \bar{a}(x), \quad \kappa := \|\mathbf{b}\|_{\ell^1(\mathbb{N})} < 1. \quad (11)$$

Further instances of the possibly nonlinear operator equations admissible in the present theory can be found in [4, 5, 18–20].

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 $U = \mathbb{R}^{\mathbb{J}}$, cp. eg. [3], [10, Section 3.4]. In a bounded Lipschitz domain $D \subset \mathbb{R}^d$, we consider once more the diffusion problem (9), with

$$u(x, \mathbf{y}) = \exp(a), \quad (12)$$

with $a(x, \mathbf{y})$ an isotropic, Gaussian random field in D as in (10). Condition (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 $\mathbb{J} = \{1, \dots, J\}$ and $U = \mathbb{R}^J$, ie. 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). In [9], Bayesian posterior consistency for the MAP estimator, which can be defined as the minimiser of an Onsager-Machlup functional on the Cameron-Martin space of the prior in the infinite dimensional case, is established. In both the small noise limit and large sample size limit, the Map estimator for Gaussian prior π_0 is shown in [9] to concentrate on the truth.

2.4. Dimension Truncation

The uncertainty parametrization (5) renders the forward map $A(u; q)$ parametric, which we express by writing $A(\mathbf{y}; q) := (A(u; q))|_{u=u(\mathbf{y})}$. We consider now the case $J = \infty$, ie., that $\mathbb{J} = \mathbb{N}$, so that $A(\mathbf{y}; q)$ depends on the sequence $\mathbf{y} = (y_j)_{j \in \mathbb{N}}$ 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, ie. we replace $u \in X$ in (5) by its J -term truncation

$$u^{(J)} = \langle u \rangle + \sum_{j=1}^J y_j \psi_j \in X. \quad (13)$$

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 $u \in X_0 \subset X$ there holds

$$\|u - u^{(J)}\|_X \leq C J^{-s}. \quad (14)$$

Under the assumption that the dependence of the forward operator $A(u; q)$ on the uncertain parameter $u \in X$ is Lipschitz, ie. 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 (5) and $u_2 = u^{(J)}$ as in (13), 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

$$\|q - q^{(J)}\|_{\mathcal{X}} \leq C J^{-s} \quad (15)$$

where $C > 0$ is possibly different from the constant in (14). We therefore note that in the ensuing developments, it suffices to assume that the uncertain input $u \in X$ is parametrized according to (13) with a sufficiently large, but finite number J of parameters.

2.5. Parametric Bayesian Posterior Density

Motivated by [27, 29], 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 §2.3.1 being uniform, for this prior we admit in (5) sequences \mathbf{y} which take values in the parameter domain $U = [-1, 1]^J$. In the lognormal case in §2.3.2, we admit as priors $\pi_0 \sim \mathcal{N}(0, C)$ Gaussian measures on $U = \mathbb{R}^N$ with trace class covariance operator C . As explained in §2.2, this leads to parametric, deterministic forward problems in the probability space

$$(U, \mathcal{B}, \pi_0).$$

With the parameter domain U as in (2.5) the parametric forward map $\Xi : U \rightarrow \mathbb{R}^K$ is given by

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

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 [10, 31] and, in its parametric version,

from [29]. To do so, we view U as unit ball in $\ell^\infty(\mathbb{J})$, the Banach space of bounded sequences taking values in U .

Proposition 2.2. *Assume that $\Xi : \bar{U} \rightarrow \mathbb{R}^K$ is bounded and continuous. Then $\pi^\delta(d\mathbf{y})$, the distribution of $\mathbf{y} \in U$ given δ , is absolutely continuous with respect to $\pi_0(d\mathbf{y})$, ie.*

$$\frac{d\pi^\delta}{d\pi_0}(\mathbf{y}) = \frac{1}{Z_\Gamma} \Theta_\Gamma(\mathbf{y})$$

with the parametric Bayesian posterior Θ_Γ given by

$$\Theta_\Gamma(\mathbf{y}) := \exp(-\Phi_\Gamma(u; \delta)) \Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} \quad (16)$$

where the Bayesian potential Φ_Γ is as in (4) and the normalization constant Z is given by

$$Z_\Gamma = \mathbb{E}^{\pi_0} [1] = \int_U \Theta_\Gamma(\mathbf{y}) \pi_0(d\mathbf{y}) . \quad (17)$$

In computational Bayesian inversion we are concerned with computational approximation of a “most likely” *system response* $\phi : X \rightarrow \mathcal{S}$ for given (noisy) observation data δ . The QoI ϕ 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

$$\begin{aligned} \Psi_\Gamma(\mathbf{y}) &= \Theta_\Gamma(\mathbf{y}) \phi(u) \Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} \\ &= \exp(-\Phi_\Gamma(u; \delta)) \phi(u) \Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} : U \rightarrow \mathcal{S} . \end{aligned} \quad (18)$$

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

$$\begin{aligned} \mathbb{E}^{\mu^\delta}[\phi] &= \frac{Z'_\Gamma}{Z_\Gamma} = \frac{1}{Z_\Gamma} \int_{\mathbf{y} \in U} \Psi_\Gamma(\mathbf{y}) \pi_0(d\mathbf{y}) \\ &= \frac{1}{Z_\Gamma} \int_{\mathbf{y} \in U} \exp(-\Phi_\Gamma(u; \delta)) \phi(u) \Big|_{u=\langle u \rangle + \sum_{j \in \mathbb{J}} y_j \psi_j} \pi_0(d\mathbf{y}) \end{aligned} \quad (19)$$

where we introduced the integral

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

We observe that for the examples (with either uniform or Gaussian prior π_0 on U), it has been shown in [10, Thms. 4.6, 4.7] that the estimate (15) of the impact of dimension-truncation (13) is inherited by the Bayesian estimate (19).

Based on (19), in [27,28] we approximated Z'_Γ and Z_Γ which, in the parametrization with respect to $\mathbf{y} \in U$, take the form of infinite-dimensional integrals with respect to the uniform prior $\pi_0(d\mathbf{y})$. In [27,28], we proposed the use of dimension adaptive Smolyak quadrature to the numerical evaluation of the integrals Z_Γ and Z'_Γ . We emphasize that the very definitions (17) and (19) imply that $0 < \Gamma \ll 1$ entails concentration of the posterior density Θ_Γ in (16).

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

We are interested in Bayesian prediction (19) in the case that the (co)variance Γ of the (assumed Gaussian) noise η in the observation data δ in (2) concentrates, ie. when $\Gamma \downarrow 0$. This will induce concentration of

the density (18) in u . Based on the integral representation (19), (17), and bearing in mind the definition (4) of the Bayesian potential, the asymptotic behaviour as $\Gamma \downarrow 0$ of the Bayesian estimate (19) follows from an asymptotic analysis of the integrals Z_Γ and Z'_Γ in (19) by Laplace's method which we present next; necessary results and references are collected for convenience in §8.

Throughout, we assume that (possibly after dimension-truncating the parameter space as in §2.4) that $J < \infty$. We distinguish the cases $K = 1$ (in which case we set $\Gamma^{-1} = \lambda$ so that $\Gamma \downarrow 0$ corresponds to $\lambda \rightarrow \infty$) in Propositions 8.3 - 8.4, and finite $K > 1$, observing that the so-called large sample size limit $K \rightarrow \infty$ can be related, at least in the Gaussian case, to the limit $\Gamma \downarrow 0$ (cp. [9, Sec.4.1]). We assume that the forward map has been dimensionally truncated to a finite number $J < \infty$ of parameters as in §2.4 and there exists a unique maximum of the Bayesian potential at \mathbf{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 $\mathcal{G}(\mathbf{y}) = A\mathbf{y}$, $A \in \mathbb{R}^{K \times J}$). We point out that, for computational purposes, the integrals Z_Γ and Z'_Γ in (19) are expressed in terms of the Lebesgue measure $d\mathbf{y}$ on the parameter domain \mathcal{U} . Then, the integral in the lognormal case is over the Bayesian density

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

assuming a truncated parametrization of the uncertain datum u of the form (13) with normally distributed parameters y_j , $j = 1, \dots, J$, ie. the prior $\pi_0 = \mathcal{N}(0, I)$. In order to unify notation for the ensuing asymptotic analysis, we introduce the parameter θ , where $\theta = 0$ corresponds to the uniform case, $\theta = 1$ to the normally distributed case.

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

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 (19) are of the same type; comparing (19) with integrals of the form $F(\lambda) = \int_{\mathcal{U}} \phi(\mathbf{y}) \exp[\lambda S(\mathbf{y})] d\mathbf{y}$ in order to use Laplace's method (cf. Appendix A in §8, we find

$$\lambda S(\mathbf{y}) = -\frac{1}{\Gamma} \left(\Phi_1(\mathbf{y}; \delta) + \frac{\theta}{2} \Gamma \|\mathbf{y}\|_2^2 \right) \quad (20)$$

with $\Phi_1(\mathbf{y}; \delta) = \frac{1}{2} ((\delta - (\mathcal{O} \circ G)(u))^\top (\delta - (\mathcal{O} \circ G)(u)))$, so that $-2\Gamma S(\mathbf{y}) = \|r(\mathbf{y})\|_2^2 + \theta \Gamma \|\mathbf{y}\|_2^2$ where $r(\mathbf{y}) := \mathcal{G}(\mathbf{y}) - \delta$ denotes the scalar residual of the data δ wr. to the uncertainty-to-observation map $\mathcal{G}(\mathbf{y})$, and where $\theta = 0$ for uniform prior π_0 and $\theta = 1$ for $\pi_0 = N(0, I)$. In (20), $\|\mathbf{y}\|_2^2$ denotes the euclidean norm in \mathbb{R}^J (we remark that in the infinite-dimensional case, $\|\mathbf{y}\|_2^2$ is to be replaced by the Cameron-Martin norm of u , cp. [10, Section 4.3]). To verify the assumptions of Proposition 8.3, we calculate

$$S'(\mathbf{y}) = -r(\mathbf{y})r'(\mathbf{y}) - \theta\Gamma\mathbf{y}, S''(\mathbf{y}) = -\{r'(\mathbf{y})r'(\mathbf{y}) + r(\mathbf{y})r''(\mathbf{y}) + \theta\Gamma\}. \quad (21)$$

Based on the expressions (21), we see that, in the uniform case, ie. $\theta = 0$, $\mathbf{y}_0 \in \text{int}(U)$ is critical wr. to $S(\cdot)$ if either $r(\mathbf{y}) = 0$ or if $r'(\mathbf{y}) = 0$. The former, *compatible case* corresponds to an exactly solvable inversion, where the observed data is reproduced exactly for some realization $u(\mathbf{y}_0) \in X$ of the uncertainty. In the latter, *incompatible case*, we have

$$\exists \mathbf{y}_0 \in \text{int}(U) : r(\mathbf{y}_0) \neq 0, r'(\mathbf{y}_0) = 0, \text{sgn}(r(\mathbf{y}_0))r''(\mathbf{y}_0) > 0. \quad (22)$$

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

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

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

$$S(\mathbf{y}_0) \leq 0, \quad S'(\mathbf{y}_0) = 0, \quad S''(\mathbf{y}_0) < 0. \quad (23)$$

In particular, then, all assumptions of Proposition 8.3 hold, and both integrals, Z_Γ and Z'_Γ , admit asymptotic expansions (50) as $\Gamma \downarrow 0$. The explicit form (51) 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 8.3 hold; in particular, that the potential (20) satisfies conditions (22) - (23). Then, for uniform prior π_0 , the Bayesian estimate in (19) admits an asymptotic expansion*

$$\mathbb{E}^{\pi^\delta}[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} \sim \tilde{a}_0 + \tilde{a}_1\Gamma + \tilde{a}_2\Gamma^2 + \dots \quad \Gamma \downarrow 0, \quad (24)$$

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

$$\mathbb{E}^{\pi^\delta}[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} = \phi(\mathbf{y}_0)(1 + o(1)), \quad \Gamma \downarrow 0. \quad (25)$$

The latter relation (25) 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 (20) and (22) - (23) hold. Proposition 8.3 allows to infer that as $\lambda = \Gamma^{-1} \rightarrow \infty$, both integrals Z_Γ and Z'_Γ in (19) admit asymptotic expansions (50). The explicit form (51) of the principal term of the asymptotics then implies that the quotient allows to infer that as $\lambda = \Gamma^{-1} \rightarrow \infty$, both integrals Z_Γ and Z'_Γ in (19) admit asymptotic expansions

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

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 ([12, Thm. I.3.1 item 3.]), there holds the asymptotic expansion

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

and from the explicit form (51) follows $\tilde{a}_0 = a'_0/a_0 = \phi(\mathbf{y}_0)$.

In the Gaussian case, the parameter domain is unbounded and $\theta = 1$ in (20), so that the function S depends (linearly) on $\lambda = \Gamma^{-1}$. Since $J = 1$, the function S is the regularized least-squares functional

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

This function has, for $\Gamma > 0$, a nondegenerate maximum at $\mathbf{y}_0(\Gamma)$. This assumption allows to derive a result: let \mathbf{y}_0, Γ_0 denote the (nondegenerate) maximiser of the function S . Then, in a neighbourhood $U(\Gamma_0)$ of Γ_0 , exists a continuously differentiable function $\mathbf{y}_0(\Gamma)$ with $\mathbf{y}_0(\Gamma_0) = \mathbf{y}_0$ and $\mathbf{y}_0(\Gamma)$ is a strict local maximizer of $S(\mathbf{y})$ for all $\Gamma \in U_{\Gamma_0}$. Hence, as $\Gamma \downarrow 0$, the maximum of $S(\mathbf{y})$ remains in a fixed, compact subset of \mathbb{R} . We may then apply Proposition 8.2 to conclude (25) in the Gaussian case. \square

The results (24), (26) show that *despite the generally exponential growth wr. to $\Gamma \downarrow 0$ in the asymptotics (50), (53) 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 [28] can, in general, not be improved), the Bayesian estimate (19) given by Z'_Γ/Z_Γ has a finite one-sided limit as $\Gamma \downarrow 0$ which, in case that the Hessian $S''(\mathbf{y}_0)$ is nondegenerate and unimodal at the critical point $\mathbf{y}_0 \in \text{int}(U)$ of the residual $r(\mathbf{y}_0)$ in (20), equals the QoI ϕ evaluated at this point.*

For Bayesian potentials $\Phi_\Gamma(\mathbf{y}; \delta)$ which attain a unique, global minimum at $\mathbf{y}_0 \in \text{int}(U)$ which is degenerate, ie. where the Hessian $S''(\mathbf{y}_0)$ has a nontrivial nullspace, then, for every finite truncation dimension J as in §2.4, there exist $N(J) \in \mathbb{N}$ and $r_k(J) \in \mathbb{Q}$ such that

$$\mathbb{E}\pi^\delta[\phi] = \frac{Z'_\Gamma}{Z_\Gamma} = \frac{a_{00}[\phi]}{a_{00}[1]}(1 + o(\Gamma)) \quad \text{as } \Gamma \downarrow 0, \quad (26)$$

where $a_{00}[\phi], a_{00}[1]$ denote the leading terms in the asymptotic expansions (53) of Z' and of Z , respectively, which are independent of Γ (but depend on δ and on J).

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

The asymptotic analysis in the case $K > 1$ many observables are available is analogous: for uniform prior π_0 , the asymptotic expansion follows from Proposition 8.3, and for Gaussian prior, the relation (25) is a consequence of the general result [9, Thm. 3.5] which covers also the case $K = \infty$.

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:

$$\Gamma = PMP^\top, \quad M = \text{diag}\{\gamma_1, \dots, \gamma_K\}, \quad 0 < \gamma_1 \leq \dots \leq \gamma_K, \quad P^\top P = 1. \quad (27)$$

In the following, we will assume that the covariance matrix of the noise is of the form $\Gamma = \gamma I$, ie. $\gamma = \gamma_1 = \dots = \gamma_K$. The case when all eigenvalues of Γ tend to zero at the same rate, ie. when $\gamma_K \downarrow 0$ while γ_K/γ_1 remains bounded, is analogous to the case $\Gamma = \gamma I$. Various intermediate cases (eg. 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 $R(\mathbf{y}) = \mathcal{G}(\mathbf{y}) - \delta \in \mathbb{R}^K$ is a K -vector with component residuals $r_k(\mathbf{y}), k = 1, \dots, K$. In place of (20) we now choose in Proposition 8.3

$$\begin{aligned} S(\mathbf{y}) &= -\Phi_\Gamma(\mathbf{y}; \delta) - \theta \frac{1}{2} \|\mathbf{y}\|_2^2 = -\frac{1}{2} R(\mathbf{y})^\top \Gamma^{-1} R(\mathbf{y}) - \theta \frac{1}{2} \|\mathbf{y}\|_2^2 \\ &= -\frac{1}{\gamma} \left(\frac{1}{2} R(\mathbf{y})^\top R(\mathbf{y}) - \theta \frac{1}{2} \gamma \|\mathbf{y}\|_2^2 \right). \end{aligned}$$

Under the assumption of a unique, nondegenerate maximizer $\mathbf{y}_0 \in \text{int}(U)$, \mathbf{y}_0 satisfies the sufficient conditions given by (23) and (24) of Theorem 3.1 holds also in this case. Note that this assumption implies that $J \geq K$.

The asymptotic expansions (24), (26) in Theorem 3.1 show that the Bayesian estimate (19) converges, in the zero observation noise limit and for nondegenerate critical points, to the QoI ϕ at critical points \mathbf{y}_0 of the parameter sequence \mathbf{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 (19) behaves numerically as quotient of infinite quantities which admits a finite limiting value at $\Gamma = 0$. This suggests that the deterministic quadrature approach of [27, 28] for the evaluation of Z_Γ, Z'_Γ becomes numerically unstable as $\Gamma \downarrow 0$. To deal with positive, but small observation noise variance Γ , therefore, the quadrature algorithms must be modified in order to remain numerically stable. As suggested by the asymptotic expansions (24) and by the explicit form (25) of its leading term, the numerical treatment of the limit $\Gamma \downarrow 0$ will require addressing the deterministic, nonlinear least-squares problem

$$\min_{\mathbf{y} \in U} \Phi_\Gamma(\mathbf{y}; \delta) + \theta/2 \|\mathbf{y}\|_2^2, \quad (29)$$

where $\theta = 0$ in the uniform case, and $\theta = 1$ in the Gaussian case. The minimization problem (29) is generally ill-conditioned for countably-parametric operator equations with uniform prior, as considered in §2.3.1. In

the lognormal case discussed in §2.3.2, the minimizer of (29) depends on Γ , generally. It corresponds to a Tikhonov regularized solution of the nonlinear least-squares problem $\min_{\mathbf{y} \in U} \Phi_{\Gamma}(\mathbf{y}; \delta)$ and converges, as $\Gamma \downarrow 0$, to a MAP estimator (see, eg., [9], [21, Chap. 3.1.1] and [10, Section 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 Bayesian inversion for positive, but possibly small observation noise (co)variance $\Gamma > 0$. We also assume throughout that the Hessian $S''(\mathbf{y}_0)$ of the Γ -scaled least-squares functional

$$S(\mathbf{y}) = -\Gamma \Phi_{\Gamma}(\mathbf{y}; \delta) \quad (30)$$

defined in (20), (28) is nondegenerate, negative definite, *uniformly with respect to* $0 < \Gamma < 1$. Then, the asymptotic analysis of §3 applies (see §8.3 for remarks on the degenerate case (26)) and Quasi-Newton Iterations with symmetric updates as in [7,14] produce locally superlinearly convergent approximations of both, critical point \mathbf{y}_0 of S and of its Hessian $S''_{yy}(\mathbf{y}_0)$ (see §10 and Proposition 10.1 for details). The point \mathbf{y}_0 is known to be related to a MAP estimate, in the Gaussian case (see [10, Section 4.3]). We therefore assume that *upon termination of the QN-SR1 algorithm 10.1, (59), the point \mathbf{y}_0 and the Hessian $S''_{yy}(\mathbf{y}_0)$ of the Bayesian potential Φ are available.* Then, from (50), (51), *upon termination of the QN process the leading term of the asymptotic expansion (50) in the $\Gamma \downarrow 0$ limit is accessible for both Z_{Γ} and Z'_{Γ} in (19).* Also, since $S(\mathbf{y})$ 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 .*

We propose two computational strategies which are robust with respect to vanishing observation noise (co)variance: first, we exploit knowledge of \mathbf{y}_0 and $S''(\mathbf{y}_0)$ to “precondition” sparse, adaptive and deterministic Smolyak quadratures proposed in [27,28]: these are, in the uniform case, based on the midpoint rule as lowest order quadrature rule and in the Gaussian case on the lowest Gauss-Hermite quadrature formula. Thus, shifting the coordinate origin to \mathbf{y}_0 will 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 achieve uniform performance of Smolyak algorithm with respect to observation noise $\Gamma > 0$, we propose reparametrization of the posterior density near MAP-points \mathbf{y}_0 of posterior concentration based on its converged QN Hessian prior to running the adaptive Smolyak algorithm. In this way, the integrand functions in the adaptive Smolyak quadrature scheme are unimodal, with maximum at $\mathbf{y} = 0$, and 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 also increases efficiency of MCMC methods; we do not detail this aspect here, and refer to, eg., [16,23] for details.

Second, the asymptotic expansions (24) and (26) justify (generalized) *Richardson extrapolation to the limit of zero observation noise*. In this case, Γ is treated as a algorithm-parameter as follows: for data δ with given, small observation noise variance $\Gamma \geq 0$, compute (in parallel) estimates (19) for several, *synthetic*, large values of Γ , and extrapolate to $0 \leq \Gamma \ll 1$. We add that this extrapolation process is also relevant for the large observation limit $K \rightarrow \infty$ (cp. [9, Sec.4]).

4.1. Curvature Rescaling

We develop the approach in the (notationally less involved) case of $K = 1$ observation functional. We apply Proposition 8.3 with S as in (20), (28). We assume in (27) that $\gamma = \Gamma > 0$ and initially for uniform prior π_0 . Then, $S(\mathbf{y})$ in (30) is independent of Γ . A QN method with symmetric update (cp. §10) will converge superlinearly and will, upon termination, yield the (unique by Assumption 2. of Proposition 8.3) maximum point \mathbf{y}_0 of $S(\mathbf{y})$ in U and the (positive) definite Hessian approximation $H_S \sim -S''(\mathbf{y}_0)$ (cp. Prop. 10.1) at \mathbf{y}_0 . Due to the use of symmetric QN updates (cp. eg. (59) in Algorithm 10.1 and Assumption 2. of Proposition 8.3), then, the Hessian H_S is symmetric positive definite, and the total work required by QN for

$S(\mathbf{y})$ in (30) 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, eg. discussed in [13]) *uniformly with respect to* Γ (as $S(\mathbf{y})$ in (30) is either independent of Γ or depends linearly on it). Denote by $-\mu_j$ with $0 < \mu_1 \leq \mu_2 \leq \dots \leq \mu_{J-1} \leq \dots \leq \mu_J < 0$ its eigenvalues (enumerated in decreasing order of absolute magnitude and counting multiplicity), and by Q the $J \times J$ orthogonal matrix of its J eigenvectors, ie.

$$H_S Q = -Q M_J, \quad M_J := \text{diag}\{|\mu_1|, \dots, |\mu_J|\}. \quad (31)$$

We next perform an *affine* change of variables $\mathbf{y} = \varphi(\check{\mathbf{x}})$ in $S(\mathbf{y})$ such that $\mathbf{y}_0 = \varphi(0)$ and such that the Hessian of $(S \circ \varphi)(\check{\mathbf{x}})$ is diagonal:

$$(D_{\check{\mathbf{x}}}^2(S \circ \varphi))(0) = M = \text{diag}\{|\mu_1|, \dots, |\mu_J|\}.$$

With $\check{V} := Q^\top(U - \{\mathbf{y}_0\})$ we change variables to find

$$\int_{\check{V}} \exp(\lambda S(\mathbf{y})) \phi(\mathbf{y}) d\pi_0(\mathbf{y}) = \int_{\check{V}} \check{g}(\check{\mathbf{x}}) d\pi_0(\check{\mathbf{x}}), \quad (32)$$

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

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

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

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

By Proposition 8.3 applied to the transformed integral (32), therefore, for $\lambda = \Gamma^{-1} \gg 1$ the transformed integrand function $g(\check{\mathbf{x}})$ in (32) will depend on the coordinate \check{x}_j to *leading order* (as the affine coordinate change in (32) will reach the normal form (52) only up to higher order terms) as $\exp(-\Gamma^{-1} |\mu_j| \check{x}_j^2 / 2)$. The ordering (31) of the μ_j then implies that the strongest ‘‘concentration’’ of the integrand function $g(\check{\mathbf{x}})$ 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 suggests curvature-rescaling

$$\check{\mathbf{x}} := \Gamma^{1/2} M^{-1/2} \mathbf{x}, \quad g(\mathbf{x}) := \check{g}(\Gamma^{1/2} M^{-1/2} \mathbf{x}). \quad (33)$$

Theorem 4.1. *Under the above assumptions, in particular the nondegeneracy of the Hessian $S''_{\mathbf{y}\mathbf{y}}$ and the analyticity of the parametric forward maps for uniform prior π_0 , the curvature rescaling transformation*

$$\mathbf{y} = \varphi(\mathbf{x}) := \mathbf{y}_0 + \Gamma^{1/2} Q M^{-1/2} \mathbf{x} \quad (34)$$

yields a transformed posterior density $g(\mathbf{x}) := \exp(\Gamma^{-1}(S \circ \varphi)(\mathbf{x}))(\phi \circ \varphi)(\mathbf{x})$ that is analytic in a neighborhood of $\mathbf{x} = 0 \in \mathbb{R}^J$. Upon transformation, the size of its domain of analyticity is independent of $0 < \Gamma \leq 1$. 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 $\phi(\mathbf{y})$ and $S(\mathbf{y})$ are independent of Γ and analytic. Since the affine change of variables $\check{\varphi}$ is independent of Γ , also the functions $\check{S}(\check{\mathbf{x}}) = S(\mathbf{y}_0 + Q\check{\mathbf{x}})$ and $\phi(\mathbf{y}_0 + Q\check{\mathbf{x}})$ are analytic w.r. to $\check{\mathbf{x}}$ in a neighborhood \check{V} of $\check{\mathbf{x}} = 0$ which is independent of Γ . There, \check{S} admits the convergent power series representation

$$\check{S}(\check{\mathbf{x}}) = S(\mathbf{y}_0) - \frac{1}{2} \check{\mathbf{x}}^\top M \check{\mathbf{x}} + \sum_{k \geq 3} \sum_{|\alpha|=k} t_\alpha \check{\mathbf{x}}^\alpha, \quad (35)$$

where all coefficients $t_\alpha := \frac{1}{\alpha!}(D_{\check{x}}^\alpha \check{S})(0)$ are independent of Γ . Multiplying (35) with Γ^{-1} and inserting (33) results in

$$\frac{(S \circ \varphi)(\mathbf{x})}{\Gamma} = \frac{S(\mathbf{y}_0)}{\Gamma} - \frac{1}{2}\mathbf{x}^\top \mathbf{x} + \Gamma^{1/2} \left\{ \sum_{k \geq 3} \Gamma^{(k-3)/2} \sum_{|\alpha|=k} t_\alpha (M^{-1/2}\mathbf{x})^\alpha \right\}. \quad (36)$$

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

4.2. Curvature-rescaled adaptive Smolyak quadrature

The preceding observations motivate *curvature rescaled, adaptive Smolyak quadrature*. Given a tolerance parameter $\tau > 0$, and $\lambda = \Gamma^{-1} > 0$, 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 (31) of the μ_j : $I_1 \subseteq I_2 \subseteq \dots \subseteq I_J \subseteq \dots$. Define

$$J_*(\tau, \lambda) := \begin{cases} 0 & \text{if } \{j \in \mathbb{J} | I_j(\tau, \lambda) \subset V\} = \emptyset, \\ \max\{j \in \mathbb{J} | I_j(\tau, \lambda) \subset V\} & \text{otherwise.} \end{cases} \quad (37)$$

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

If $J_* = 0$, there is no coordinate in which the integrand function $f(\mathbf{x})$ in the transformed integral (32) over V is concentrating. If $J_* > 0$, the integrand function in the transformed integral (32) over V concentrates numerically (at threshold τ) near $x_j = 0$ in the coordinates x_1, \dots, x_{J_*} . Accordingly, if $U = [-1, 1]^J$, prior to application of the adaptive Smolyak quadrature algorithm, we rescale the integration coordinates according to

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

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

$$V_\tau := \prod_{1 \leq j \leq J_*} I_j(\tau, \lambda) \times \prod_{j > J_*} V_j.$$

For the resulting truncation error holds the error bound

$$\left| \int_V f(\mathbf{x}) d\pi_0(\mathbf{x}) - \int_{V_\tau} f(\mathbf{x}) d\pi_0(\mathbf{x}) \right| \leq \pi_0(V \setminus V_\tau) \sup_{\mathbf{y} \in U} \|\phi(\mathbf{y})\|_S \leq \tau \sup_{\mathbf{y} \in U} \|\phi(\mathbf{y})\|_S.$$

The *variable metric adaptive Smolyak quadrature algorithm* consists in applying the adaptive Smolyak algorithm from [28] to the rescaled integrand function where coordinates (y_1, \dots, y_{J_*}) are transformed with (34).

Remark 4.2. The proof of Theorem 4.1 shows that, *asymptotically as $\Gamma \downarrow 0$* , at any fixed, finite truncation dimension J , the posterior density will concentrate in *all* coordinates y_j , $j = 1, \dots, J$.

The (cp. §9) compactness of the Hessian $S''_{\mathbf{y}\mathbf{y}}(\mathbf{y}_0)$ implies that, generically, $\{|\mu_j|\}_{j \geq 1}$ accumulates at 0 for increasing J . Hence, for every fixed $\Gamma > 0$ there exists a crossover dimension $J_*(\Gamma)$ such that $\mu_j / \Gamma \leq 1$ for all $j \geq J_*$. The spectrum of the Hessian $S''_{\mathbf{y}\mathbf{y}}(\mathbf{y}_0)$ thus quantifies the (finite!) dimension of the parameter space where the Bayesian posterior can concentrate.

Theorem 4.1 implies that the curvature-rescaling (34) in conjunction with the adaptive Smolyak quadratures from [27, 28] *in rescaled coordinates* for Z_Γ and Z'_Γ will converge independently of Γ .

4.3. Extrapolation to vanishing observation noise variance $\Gamma = 0$

The previous regularizations required nondegeneracy of the Hessian at the critical point $y_0 \in U$. We therefore present an alternative approach which avoids explicitly accessing curvature information, which is mathematically justified for both, regular as well as degenerate Hessians $S''(y_0)$ and which is based on Richardson extrapolation to the limit of vanishing observation noise variance Γ .

The corresponding asymptotic expansions (26) with respect to $\Gamma \downarrow 0$ which are based on using Propositions 8.3 and 8.4. Precisely, only the *existence of an asymptotic expansion* such as (50) is used, and Richardson extrapolation to (possibly vanishing) observation variance $\Gamma \geq 0$ is performed. We present the details, for simplicity only in the case $K = 1$, following [30].

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 $0 < \Gamma \leq \Gamma_0 < \infty$. Assume that there exist constants B and β_k , $k = 1, 2, \dots$, which are independent of Γ and functions $q_k(\Gamma)$ which form an asymptotic sequence in the sense that

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

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

$$B(\Gamma) \sim B + \sum_{k=1}^{\infty} \beta_k q_k(\Gamma) \quad \text{as } \Gamma \downarrow 0. \quad (38)$$

For a strictly monotonically decreasing sequence $\Gamma_0 > \Gamma_1 > \Gamma_2 > \dots$ in $(0, \Gamma_0]$ with $0 = \lim_{k \rightarrow \infty} \Gamma_k$ and for each pair (j, q) of positive integers, define the sequence $\beta_0, \beta_1, \dots, \beta_q$ as solution of the linear system of $q + 1$ equations for the $q + 1$ unknowns $\beta_0^{jq}, \dots, \beta_q^{jq}$ which is given by

$$B(\Gamma_l) = \sum_{k=0}^q \beta_k^{jq} q_k(\Gamma_l), \quad j \leq l \leq j + q. \quad (39)$$

Then the β_0^{jq} obtained in (39) are approximations of $\lim_{\Gamma \downarrow 0} B(\Gamma)$; specifically, there holds (cp [30, Thm. 2.2]):

Proposition 4.3. *Assume that $B(\Gamma)$ admits the expansion (38). Then, for some monotonically decreasing sequence $\{\Gamma_k\}_{k \geq 1}$, accumulating at $\Gamma = 0$ such that*

$$\lim_{l \rightarrow \infty} \frac{q_k(\Gamma_{l+1})}{q_k(\Gamma_l)} = c_k \neq 1 \quad \text{and } c_j \neq c_k \quad \text{for } j \neq k, \quad (40)$$

for every $q \in \mathbb{N}$ fixed, the β_0^{jq} and $B = \lim_{\Gamma \downarrow 0} B(\Gamma)$ in (39) satisfy

$$\beta_0^{jq} - B \sim \beta_{q+1} \left[\prod_{i=1}^q \left(\frac{\beta_{q+1} - \beta_i}{1 - \beta_i} \right) \right] q_{q+1}(\Gamma_j) \quad \text{as } j \rightarrow \infty. \quad (41)$$

Under condition (40), $\{\beta_0^{jq}\}_{j \geq 1}$ tends to B faster than $\{\beta_0^{j,q-1}\}_{j \geq 1}$, ie.

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

4.3.2. Numerical extrapolation to zero observation noise covariance

Comparing the preceding result on generalized Richardson extrapolation with the asymptotic expansions (24) in the regular case (where, according to Theorem 3.1, we have $q_k(\Gamma) = \Gamma^k$), we see immediately that in the regular case (24), Assumption (40) is satisfied.

Proposition 4.4. *Assume that $\mathcal{G}(\cdot)$ and δ are such that all Assumptions of Proposition 8.3 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 (40) is fulfilled.*

Proof. In the nondegenerate case, Theorem 3.1 and (24) imply $q_k(\Gamma) = \Gamma^k$. Therefore, there holds

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

Therefore, the assumptions of Proposition 4.3 are satisfied, and (41) holds. \square

Even without explicit knowledge of coefficients and exponents in the asymptotic expansions (24), we may therefore apply (39) in order to extrapolate to the limit $\Gamma \downarrow 0$ as follows.

Given one set of data δ , the Bayesian estimate (19) is approximated numerically for a sequence $\Gamma = \{\Gamma_k\}_{k=1}^K$ of synthetic (ie. not based on the experimental data) variances in observation noise which decrease monotonically. Owing to numerical instability in the quadrature evaluation of Z_Γ and Z'_Γ in (19) for positive, but small observation noise covariance Γ , the sequence Γ should be chosen as $\{\Gamma_k\}_{k \geq 1}$:

$$\Gamma : \Gamma_k = 2^{-k+1} \text{ (geometric sequence)}. \quad (42)$$

We note in passing that other sequences with slower than geometric decrease (42) are infeasible even in the nondegenerate case, as they violate the stability conditions in [30].

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) by Γ_{obs} . Extrapolation to the limit $\Gamma = 0$ then proceeds in the usual fashion, ie. 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 $\Gamma_{obs} \geq 0$.

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

Extrapolation to small observation noise $\Gamma_{obs} \geq 0$ 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 (19) depend on Γ_k only via the Bayesian potential Φ_{Γ_k} in (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 all quadrature approximations Z_{Γ_k} and Z'_{Γ_k} in (19).

While the generalized Richardson extrapolation to the limit is mathematically justified by the asymptotic expansions (24), (26) and by the analysis in [30], we remind that the present setting will require their use for positive, small but fixed observation noise covariance $\Gamma_{obs} > 0$; the next result is analogous to (41).

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 (38) with $q_k(\Gamma) = \Gamma^j$ as in (24) for the Bayesian estimate in the nondegenerate case.*

For a sequence of (synthetic) observation noise covariances $1 = \Gamma_0 > \Gamma_1 > \dots > \Gamma_j > \dots > \Gamma_{j+q} \gg \Gamma_{obs} \geq 0$ (which Γ_k 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 $\Gamma_{obs} \geq 0$, compute the coefficients $(\beta_k^{jq})_{k=0}^q$

in (39). Then, there holds the asymptotic error bound

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

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

Proof. For $\Gamma_{obs} = 0$, the assertion is Proposition 4.3. We may therefore assume that $\Gamma_{obs} > 0$. 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

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

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

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

and the assertion follows for $\Gamma_{obs} > 0$. The preceding argument remains valid even for $\Gamma_{obs} = 0$ which concludes the proof. For the limiting case $\Gamma_{obs} = 0$, (41) gives a more detailed information about the asymptotic behaviour. \square

5. NUMERICAL EXPERIMENTS

We consider the model parametric elliptic boundary value problem

$$-\text{div}(u \nabla p) = f \quad \text{in } D := [0, 1], \quad p = 0 \quad \text{in } \partial D, \quad (43)$$

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

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

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 $x = 0.75$, ie. the observation operator \mathcal{O} consists of K_{obs} system responses at $K_{obs} = 2$ 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 $\mathcal{G} : X \rightarrow \mathbb{R}^{K_{obs}}$, with $K_{obs} = 2^{N_K} - 1$, $N_K = 2$, the expectation of the observed solution of the forward model, ie. our aim is to approximate

$$Z' = \int_U \exp(-\Phi(u; \delta)) \phi(u) \Big|_{u = \langle u \rangle + \sum_{j=1}^2 y_j \psi_j} \mu_0(dy),$$

with the normalization constant Z given by

$$Z = \int_U \exp(-\Phi(u; \delta)) \Big|_{u=\langle u \rangle + \sum_{j=1}^2 y_j \psi_j} \mu_0(dy),$$

so that the expectation of interest is given by Z'/Z . The noise $\eta = (\eta_j)_{j=1, \dots, K_{obs}}$ 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 $\Gamma_{obs} = \gamma I$ with $\gamma = 0.25^2$.

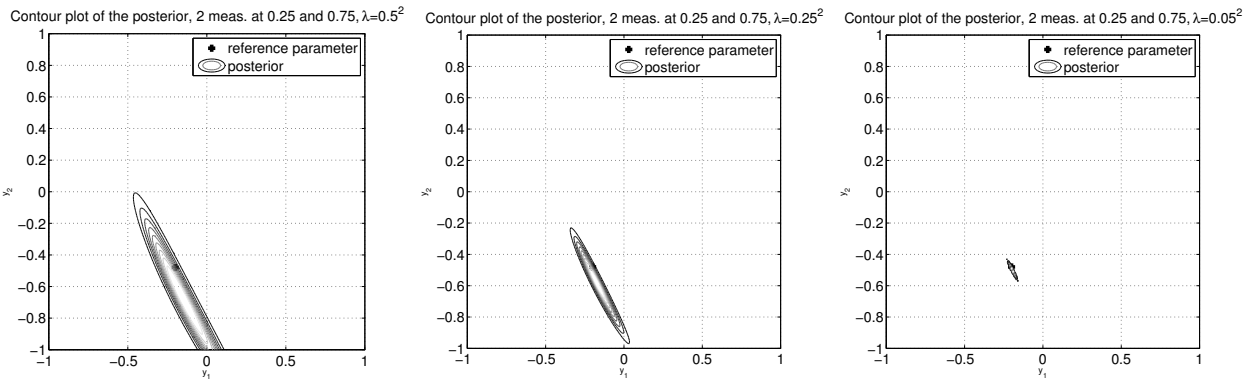


FIGURE 1. Contour plot of the posterior density $\Phi_{\Gamma_{obs}}$ with observational noise $\Gamma_{obs} = 0.5^2 I$ (left), $\Gamma_{obs} = 0.25^2 I$ (middle) and $\Gamma_{obs} = 0.05^2 I$ (right). The concentration of the posterior as well as its anisotropy are clearly visible.

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, cp. 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 in the small noise setting will be the curvature rescaling regularization, ie. 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. The maximizer y_0 of the posterior density $\Theta_{\Gamma_{obs}}(y)$ in (16) is computed by minimizing

$$\frac{1}{2}(\delta - \mathcal{G}(u))^T (\delta - \mathcal{G}(u)) \Big|_{u=\sum_{j=1}^2 y_j \psi_j}$$

using a trust-region Quasi-Newton approach with SR1 updates as described in §10. The trust-region sub-problems are iteratively solved by a CG-Steihaug method [26]. The results presented in Proposition 10.1 ensure (locally superlinear) convergence of SR1 updates to the Hessian of the potential. The approximated Hessian H_{SR1} is diagonalized $H_{SR1} = QMQ^T$ and the integrand is regularized by the *curvature-rescaling* transformation (34), ie.

$$y \rightarrow y_0 + \Gamma_{obs}^{1/2} QM^{-1/2} z, \quad z \in \mathbb{R}^J. \quad (44)$$

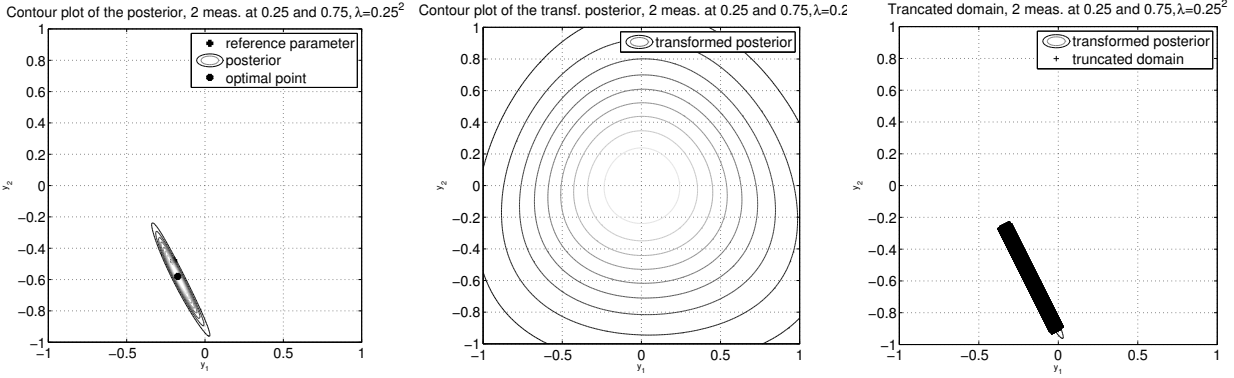


FIGURE 2. Contour plot of the posterior density with observational noise $\Gamma_{obs} = 0.25^2 I$ (left), contour plot of the transformed posterior (middle) and truncated domain of integration of the rescaled Smolyak approach in the original coordinate system (right) with $\tau = 0.16$.

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

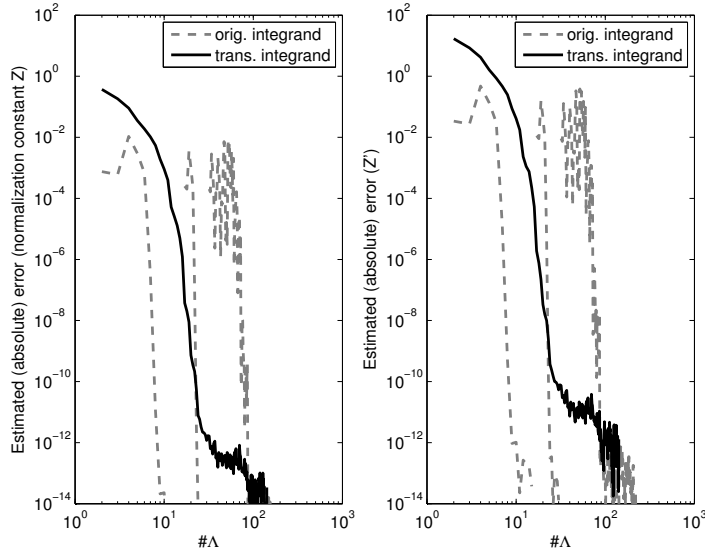


FIGURE 3. Comparison of the estimated (absolute) error curves using the Smolyak approach (Clenshaw-Curtis points) for the original integrand (gray) and the transformed integrand (black) for the computation of $Z_{\Gamma_{obs}}$ (left) and $Z'_{\Gamma_{obs}}$ (right) with observational noise $\Gamma_{obs} = 0.25^2 I$.

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 a too early stopping of the Smolyak algorithm, which is a well known problem of the greedy-type strategy for integrands concentrated on a small region of the integration domain. The Smolyak

algorithm applied to the transformed integrand shows a fast convergence, as expected, and comparison to a reference solution computed by a full tensor grid based on Clenshaw-Curtis points of order 14 gives a relative error of $8.28036e - 4$.

Next, we discuss the extrapolation method presented in §4.3. We therefore consider the following synthetic variance sequence

$$\gamma : \gamma_k = 2^{-(k-1)} \quad 1 \leq k \leq K = 11, \quad (45)$$

so that $\Gamma_k = \gamma_k I$ in the observation noise. The conditional expectation $\mathbb{E}^{\mu^\delta}[\phi]$ is approximated by extrapolation to the zero observation variance limit of the Bayesian estimates for the synthetic observation noise variances $\Gamma_k = \gamma_k I$. I.e., we construct a sequence of synthetic noise variances Γ_k and apply the Smolyak quadrature to approximate the sequence $\mathbb{E}_{\Gamma_k}^{\mu^\delta}[\phi]$; we then extrapolation the resulting sequence of Bayesian estimates to the observation noise variance $0 \leq \Gamma_{obs} \ll 1$ by Richardson extrapolation.

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 $\Gamma_k = 2^{-(k-1)}$, where the error in each iteration is computed by a tensor grid reference solution.

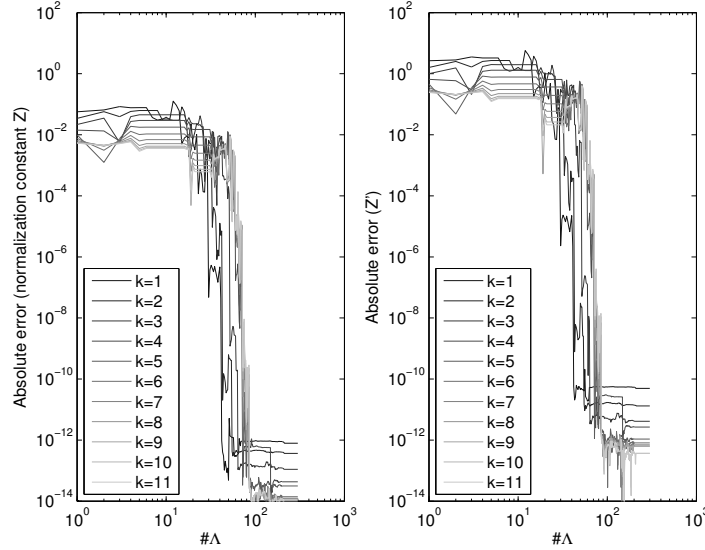


FIGURE 4. Comparison of the (absolute) error curves using adaptive Smolyak quadrature (Clenshaw-Curtis points) for the computation of the quantities Z_{Γ_k} (left) and Z'_{Γ_k} (right) with synthetic noise variances $\Gamma_k = 2^{-(k-1)}$.

Considering the asymptotic expansion (24) of the Bayesian estimate

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

with $a_0 = \phi(y_0)$, (generalized) Richardson extrapolation yields the following result, cp. Table 1. Extrapolation is based the geometric sequence $g_k := 2^{-(k-1)}$.

#	Synthetic Variances $\Gamma_1, \dots, \Gamma_k$									
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

TABLE 1. Relative errors for extrapolation to the zero observation variance Γ_{obs} . Reference value from direct overintegration using tensor product quadrature.

The next example is the model parametric elliptic boundary value problem (43) with lognormal diffusion coefficient (12) defined as

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

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 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 , ie. 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 8. The exponential decay of the Gaussian density as $|y| \rightarrow \infty$ and nondegeneracy of the potential Φ in (20) allow the ‘‘localization’’ of the argument in the proof of Proposition 8.3 to a compact subset $\tilde{U} \subset U = \mathbb{R}^2$.

We adopt the setting of the uniform test case, ie. 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 $K_{obs} = 2$ 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 (3) with $\eta \sim \mathcal{N}(0, \Gamma_{obs})$ and $\mathcal{G} : X \rightarrow \mathbb{R}^{K_{obs}}$, with $K_{obs} = 2$, we are interested in the behaviour of the posterior density

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

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

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

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

by the trust-region Quasi-Newton approach with SR1 updates. The nondegeneracy of S at 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* (44). 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.

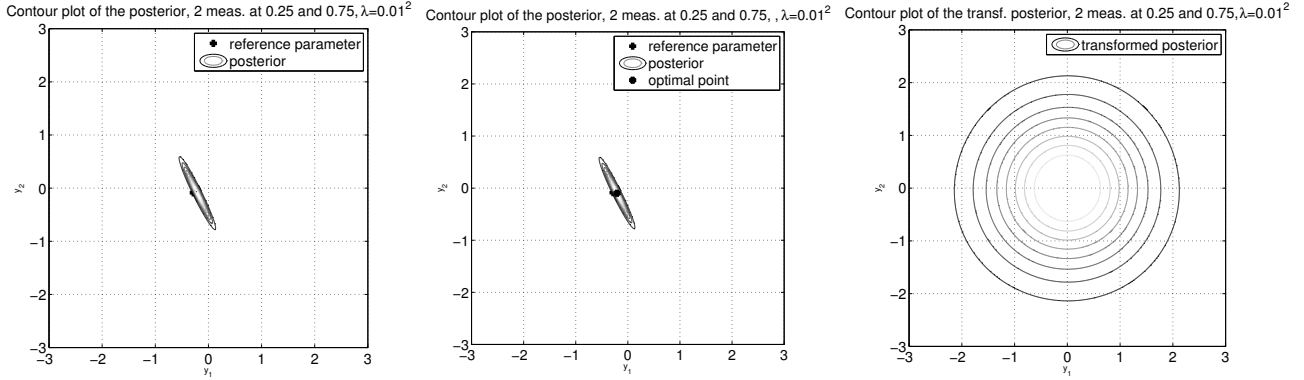


FIGURE 5. Contour plot of the posterior density with observational noise $\Gamma_{obs} = 0.01^2 I$ (left), the minimizer of the potential (middle) and contour plot of the transformed posterior (right). Asymptotically spherical level-lines of the rescaled posterior density are in agreement with Theorem 4.1 and (35).

Similar to the uniform case, we observe an almost quadratic behavior of the posterior, cp. Figure 5, which suggests, due to the curvature preconditioning, convergence of the Smolyak algorithm independent of the observational noise covariance Γ_{obs} .

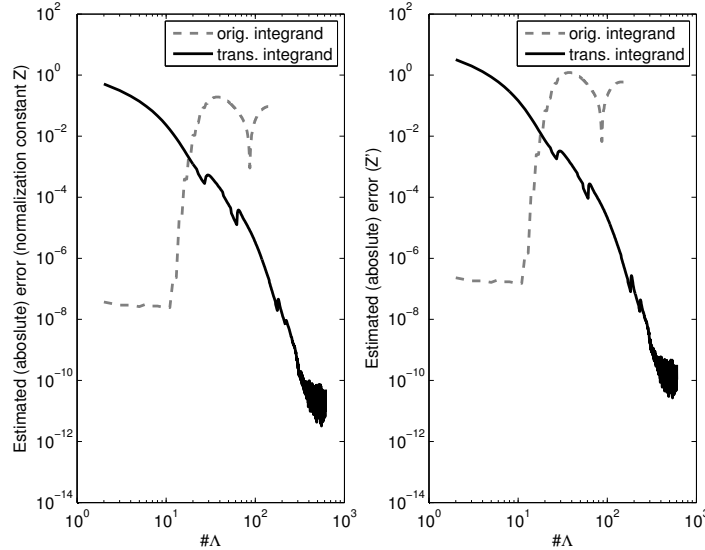


FIGURE 6. Comparison of the estimated (absolute) error using the Smolyak approach (Gauss-Hermite points) for the original integrand (gray) and the transformed integrand (black) for the computation of $Z_{\Gamma_{obs}}$ (left) and $Z'_{\Gamma_{obs}}$ (right) with observational noise $\Gamma_{obs} = 0.01^2 I$.

Figure 6 confirms that curvature rescaling (34) of the integrand functions prior to adaptive Smolyak quadrature restores robust w.r. to Γ_{obs} convergence, in accordance with Theorem 4.1.

To further reduce the number of quadrature points, we exploit the knowledge of \mathbf{y}_0 and Theorem 4.1, (36), which implies asymptotically, as $\Gamma \rightarrow 0$, quadratic behaviour of the rescaled posterior densities. We propose a quadrature scheme in spherical polar coordinates: discretizing the radial coordinate by the Gauss-Laguerre

abscissas of order 2 and the angular coordinate by a trapezoidal rule with 3 grid points, the relative error of the quantity $\mathbb{E}^{\mu^\delta}[\phi]$ is below $3e - 5$ and the quadrature effort is reduced to 6 forward simulations, cp. Figure 7.

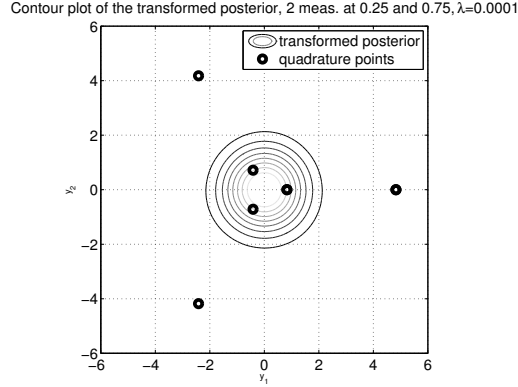


FIGURE 7. Contour plot of the transformed posterior with observational noise $\Gamma_{obs} = 0.01^2 I$ and the quadrature points resulting from integration in polar coordinates (trapezoidal rule, Gauss-Laguerre).

Note 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 S^{J-1} by adaptive Smolyak quadrature based on tensorized Gauss-Jacobi rules in hyperspherical coordinates as indicated in Appendix D, §11.

Finally, we discuss the extrapolation method based on the sequence of synthetic variances defined in (45). Figure 8 indicates the convergence rate of the adaptive Smolyak algorithm with same data δ , for synthetic observational noise variance $\Gamma_k = 2^{-(k-1)}$.

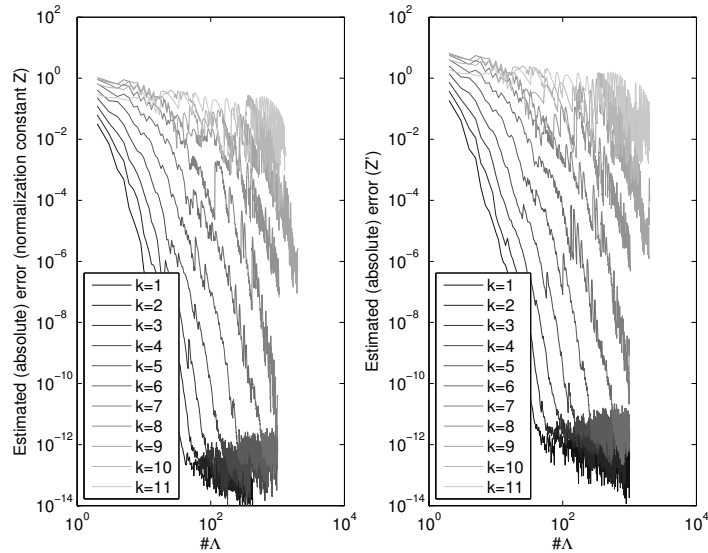


FIGURE 8. Comparison of the estimated (absolute) error using adaptive Smolyak quadrature (Gauss-Hermite points) for the computation of the quantities Z_{Γ_k} (left) and $Z_{\Gamma_k}^1$ (right) with synthetic noise variances $\Gamma_k = 2^{-(k-1)}$.

As the Smolyak algorithm fails to converge for small synthetic observational noise variances ($\Gamma_7, \dots, \Gamma_{11}$), only the first six approximated values of the normalization constant Z and the quantity Z' are used for the extrapolation based approach. Figure 8 plots the corresponding values for the quantities Z_k , Z'_k and the resulting approximated values of the conditional expectations $\mathbb{E}_{\Gamma_k}^{\mu^\delta}[\phi]$ are shown in Figure 9.

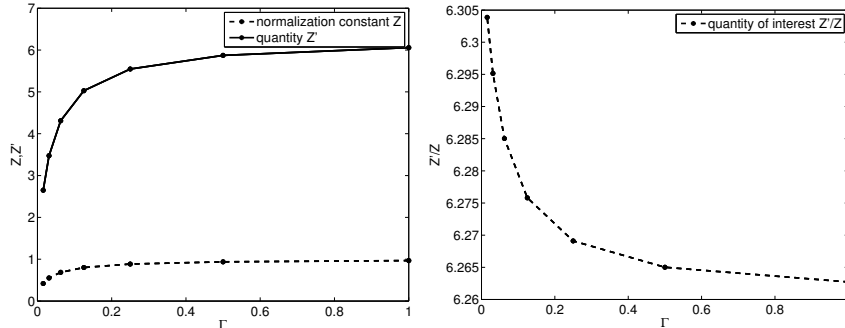


FIGURE 9. Approximated values of the quantities Z_{Γ_k} and Z'_{Γ_k} (left) and the conditional expectation $\mathbb{E}_{\Gamma_k}^{\mu^\delta}[\phi]$ (right) with observational noise Γ_k , $k = 1, \dots, 7$.

Using as reference value the approximation of $\mathbb{E}^{\mu^\delta}[\phi]$ computed by curvature-rescaled, adaptive Smolyak quadrature (with absolute error tolerances $1e - 10$), the results of the extrapolation based on (24) are summarized in Table 2.

#	Synthetic Variances $\Gamma_1, \dots, \Gamma_k$							
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 2. Romberg table of the relative errors for the extrapolation to the zero observation variance Γ_{obs} .

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 10 shows the improvement in convergence gained by the “curvature-rescaling” preconditioning of the adaptive Smolyak quadrature.

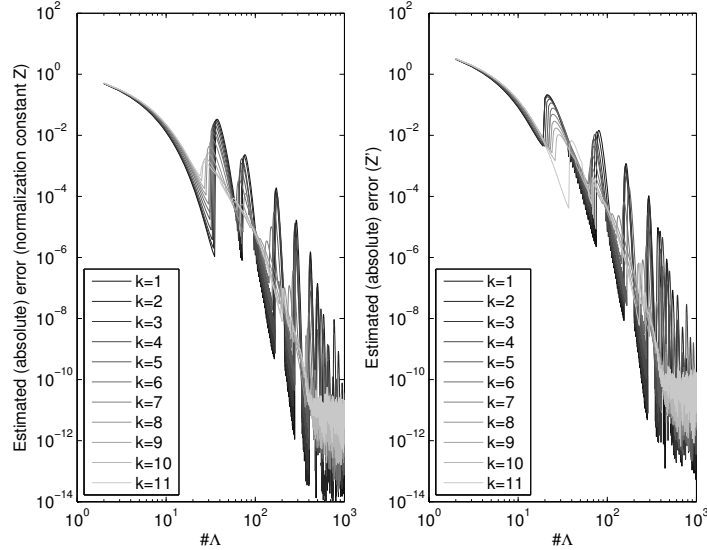


FIGURE 10. Comparison of the estimated (absolute) error using the (curvature-preconditioned) adaptive Smolyak approach (Gauss-Hermite points) for the computation of the quantities Z_{Γ_k} (left) and Z'_{Γ_k} (right) for various synthetic observation noise variances Γ_k .

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 $\mathbb{E}_{\Gamma_k}^{\mu^\delta}[\phi]$. The numerical results are presented in Table 3.

#	Synthetic Variances $\Gamma_1, \dots, \Gamma_k$										
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	0.00005542	0.00007397	0.00008692	0.00009050	0.00009179	0.00009233	0.00009257	0.00009269		
10	0.00025547	0.00006885	0.00009000	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

TABLE 3. Romberg table of the relative errors for the extrapolation to the zero observation variance Γ_{obs} .

In summary, rescaling of the posterior densities (16), (18) in the Bayesian estimate (19) based on second order information at the extremum of these (assumed unimodal) posterior densities to “precondition” the adaptive Smolyak quadrature leads to a significant improvement of the convergence properties. In particular, under a dimension truncation and a nondegeneracy assumption, this “curvature-rescaling” of the posterior renders the adaptive Smolyak quadrature approach 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, cp. 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 8.

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 $S(\mathbf{y}_0)$), we consider Bayesian inversion of the model parametric elliptic boundary value problem (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 $y_j \sim \mathcal{N}(0, 1)$, $j \in \mathbb{J} = \{1, \dots, 5\}$. The eigenvalues of the approximated Hessian of the Bayesian potential by QN with SR1 updates indicate that the posterior concentrates in 2 coordinates. According to the strategy discussed in §4.1, we shift the origin to the minimizer of the posterior density, rotate the coordinate system and rescale the 2 integration coordinates to remove the degeneracy of the integrand. The results of the preconditioned Smolyak method compared to the performance of the Smolyak applied to the original integrand is shown in Fig. 11

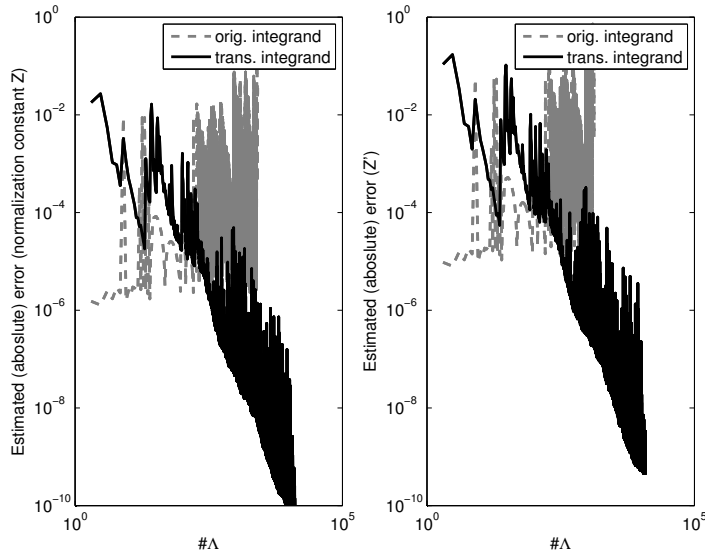


FIGURE 11. Comparison of the estimated (absolute) error using the Smolyak approach (Gauss-Hermite points) for the original integrand (gray) and the transformed integrand (black) for the computation of $Z_{\Gamma_{obs}}$ (left) and $Z'_{\Gamma_{obs}}$ (right) with observational noise $\Gamma_{obs} = 0.01^2 I$, $J = 5$.

6. DISCUSSION AND CONCLUSIONS

We presented an adaptive quadrature approach to high-dimensional, parametric Bayesian inversion. The proposed approach is based on a deterministic representation of the posterior densities, as a holomorphic function of possibly countably many parameters.

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 $\Gamma_{obs} > 0$ 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 . Via posterior consistency, these methods are also relevant in the large observation number limit, an aspect that is not discussed here in detail (cp. [9, Sec.4]). We also remark that the possibility of reparametrization the posterior to alleviate ill-conditioning due to, eg., concentration effects, has been suggested in [24], albeit with a different approach related to optimal transport.

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 (19).

Generalized Richardson extrapolation in the case of *degenerate asymptotics* (26) 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 (40) in case that (53) holds with $N > 0$. For $N = 0$ in (53), however, geometric synthetic observation noise covariance sequences $\{\Gamma_j\}_{j \geq 0}$ will result in a stable extrapolation scheme for the leading term in (26), even without explicit knowledge of the exponents $r_k \in \mathbb{Q}$.

7. ACKNOWLEDGEMENTS

This work is supported by Swiss National Science Foundation (SNF) and by the European Research Council (ERC) under FP7 Grant AdG247277. Work of CIS was mostly performed at ETH funded by the aforementioned ERC grant. The authors thank Andrew M. Stuart for helpful comments on an earlier version of this work.

8. APPENDIX A: LAPLACE'S METHOD

The asymptotic structure of integrals of the form

$$F(\lambda) = \int_U \phi(\mathbf{y}) \exp[\lambda S(\mathbf{y})] d\mathbf{y} \quad (46)$$

as the parameter $\lambda \rightarrow \infty$ is obtained by Laplace's method. In (46), U is a bounded domain in \mathbb{R}^J , with parameter dimension $J < \infty$, and $\mathbf{y} = (y_1, \dots, y_J) \in U$, $\lambda \in \mathbb{R}$ is a parameter and $S(\mathbf{y})$ is a sufficiently smooth, real-valued function.

We are interested in the asymptotic behaviour of $F(\lambda)$ for $\lambda \rightarrow \infty$, having in mind $\lambda = \Gamma^{-1}$. We first consider the nondegenerate case, and address briefly the degenerate case in §8.3. The proof is, in fact, *recursive by dimension* with the recursion based on the one-dimensional result from [12, Chap. II.2.1] for parametric integrand functions; we present this case first.

8.1. Parametric One-dimensional Case

Consider the asymptotic behaviour of integrals of the form

$$F(\lambda, \alpha) = \int_a^b f(x, \alpha) \exp(\lambda S(x, \alpha)) dx, \quad \lambda \rightarrow \infty, \quad (47)$$

where $\alpha = (\alpha_1, \dots, \alpha_k) \in \mathbb{R}^k$ 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 (47), $f(x, \alpha)$ 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 $\alpha \in G$:

$$\forall \alpha \in G: \quad -S_{xx}(x_0(\alpha), \alpha) \geq \delta_0 > 0.$$

For all $\alpha \in G$ holds $x_0(\alpha) \in [a', b'] \subset [a, b]$ for some fixed $a < a' < b' < b$.

Proposition 8.1. ([12, Thm. II.2.1]) *Assume (A1) - (A3). Then the function $F(\lambda, \alpha)$ in (47) admits the asymptotic expansion*

$$F(\lambda, \alpha) \sim \exp(\lambda S(x_0, \alpha)) \sum_{j=0}^{\infty} c_j(\alpha) \lambda^{-j-1/2}, \quad \lambda \rightarrow \infty, \quad (48)$$

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

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

The first term in the asymptotic expansion (48) 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 §2.3.2, there arise improper integrals where the function $S(\mathbf{y})$ in (20) depends generally (analytically) on Γ at $\Gamma = 0$ (cp. (29)). In place of (47), we consider therefore the asymptotics of

$$F(\lambda) = \int_{-a(\lambda)}^{a(\lambda)} f(x, \lambda) \exp[S(x, \lambda)] dx, \quad \lambda \rightarrow \infty \quad (49)$$

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

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

Proposition 8.2. ([12, Thm.II.2.2]) *Assume that there exists a real-valued function $\mu(\lambda) > 0$ with $\mu(\lambda) \rightarrow \infty$ as $\lambda \rightarrow \infty$ such that, as $\lambda \rightarrow \infty$, in (49) holds*

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

$$f(x, \lambda) = f(x_0(\lambda), \lambda)[1 + o(1)]$$

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, as $\lambda \rightarrow \infty$, there holds

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

8.2. Nondegenerate Case in finite Dimension $J > 1$

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

Proposition 8.3. (cp. [12, Thm. II.4.1]) Consider $F(\lambda)$ as in (46) under the assumptions

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

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

$$\exp[-\lambda S(\mathbf{y}_0)] F(\lambda) \sim \lambda^{-J/2} \sum_{k=0}^{\infty} a_k \lambda^{-k}. \quad (50)$$

In particular, as $\lambda \rightarrow \infty$ the leading term a_0 in (50) has the exact representation

$$F(\lambda) = \exp[\lambda S(\mathbf{y}_0)] (2\pi/\lambda)^{J/2} \frac{\phi(\mathbf{y}_0)}{\sqrt{|\det(S''(\mathbf{y}_0))|}} (1 + O(\lambda^{-1})). \quad (51)$$

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

Proof. The proof is recursively by dimension, using the asymptotic expansion in Proposition 8.1 of parametric, univariate integrals. The argument elucidates the curvature rescaling preconditioning in §4.1, so that we indicate here the derivation of (51). Choosing $\varepsilon > 0$ sufficiently small, we partition the domain U of integration in (46) into $U = B_\varepsilon(\mathbf{y}_0) + (U \setminus B_\varepsilon(\mathbf{y}_0)) =: B_\varepsilon(\mathbf{y}_0) + V_\varepsilon$ and we may write the integral (46) accordingly as $F(\lambda) = F_\varepsilon(\lambda) + G_\varepsilon(\lambda)$ with F_ε defined as in (46) but with the domain of integration being $B_\varepsilon(\mathbf{y}_0)$ in place of U . Using that \mathbf{y}_0 is the global maximum of the integrand function in \bar{U} , there exists $\delta' > 0$ (depending on $\varepsilon > 0$ and on S but independent of $\lambda > 0$) such that $|G_\varepsilon(\lambda)| = O(\exp(\lambda(S(\mathbf{y}_0) - \delta')))$ for all $\lambda > 0$. This integral, being of (exponentially) lower asymptotic order as $\lambda \rightarrow \infty$, will not contribute to (51). Thus the asymptotics of $F(\lambda)$ as $\lambda \rightarrow \infty$ in turn is completely determined by $F_\varepsilon(\lambda)$. To prove (51), we perform a C^3 -diffeomorphic change of variables $x = \varphi(\mathbf{y})$ for all $\mathbf{y} \in B_\varepsilon(\mathbf{y}_0)$ such that a) $\varphi(\mathbf{y}_0) = \mathbf{y}_0$, b) $\varphi'(\mathbf{y}_0) = \mathbf{1}$, c) $\varphi(B_\varepsilon(\mathbf{y}_0)) = \widetilde{B_\varepsilon(\mathbf{y}_0)} \subset U$ and such that

$$(S \circ \varphi)(x) = S(\mathbf{y}_0) + \frac{1}{2} \sum_{j=1}^J \mu_j x_j^2. \quad (52)$$

Here, $\mu_j < 0$ are the eigenvalues of the Hessian S''_{yy} at \mathbf{y}_0 (which are negative according to assumptions 2. and 4.). The existence of a C^3 -diffeomorphism φ with (52) (which, like the function $S(\mathbf{y})$, is independent of λ) follows from the Morse Lemma by Assumption 3. Changing variables in the integral $F_\varepsilon(\lambda)$, gives

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

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

$$F_\varepsilon^1(\lambda) = \exp(\lambda S(\mathbf{y}_0)) \int_{|x_1| < \varepsilon'/2} \exp(\lambda \mu_1 \frac{x_1^2}{2}) Y(x_1) dx_1$$

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

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

As $\lambda \rightarrow \infty$, the Laplace asymptotics Proposition 8.1 for the univariate integral apply to $F_\varepsilon^1(\lambda)$. This yields

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

Iterating this reasoning for $Y(y_{0,1})$ which is a $J - 1$ -dimensional integral of the same type $J - 1$ times implies (50). An analysis of the spherical integral in polar-coordinates finally yields (51) (see [12, Prop. 4.1]). \square

8.3. Degenerate Case

The nondegeneracy of $S''(\mathbf{y}_0)$ is essential for the validity of the asymptotic expansion (50) and for obtaining the explicit formula (51). Based on the (generic) compactness of the Hessian in the countably-parametric setting (cp. §9), for large, finite truncation dimension J , the parametric Hessian will generically degenerate. In the case when Assumption 4. in Proposition 8.3 does not hold, (51) becomes invalid, but a generalization of (50) can be established. The result is less explicit than (50).

Proposition 8.4. *Under assumptions 1. - 3. of Proposition 8.3, there exists $N \in \mathbb{N}$ and constants $a_{kl} \in \mathbb{R}$ such that for $\lambda \rightarrow \infty$, for $F(\lambda)$ as in (46) there holds the power-logarithmic asymptotics*

$$\exp(-\lambda S(\mathbf{y}_0)) F(\lambda) \sim \sum_{k=0}^{\infty} \left(\sum_{l=0}^N a_{kl} \lambda^{-r_k} (\log \lambda)^l \right) . \quad (53)$$

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

For the proof, we refer to [12, Section II.3.4]. The derivation and justification of (53) is considerably more involved than the classical asymptotics (50) as its proof is based on “resolution of singularities” [1] (see also [12, Thm. II.4.3]). We also remark that, contrary to (50), to date the leading term a_{00} in (53) is *not* known in closed form, in general, and the exponents $r_k \in \mathbb{Q}$ in (53) depend on J .

9. APPENDIX B: COMPACTNESS OF THE HESSIAN FOR UNIFORM PRIOR π_0

For uniform prior π_0 on the parameter sequences \mathbf{y} with bounded ranges, ie. $U = [-1, 1]^{\mathbb{N}}$, we investigate the Hessian $S''(\mathbf{y})$ in (20), (21). At a critical point $\mathbf{y}_0 \in \text{int}(U)$ we have $r'(\mathbf{y}_0) = 0$ so that we find, in the case (22) of incompatible data, ie. when $\Phi_\Gamma(\mathbf{y}_0, \delta) > 0$, since G depends linearly on f

$$S''(\mathbf{y}_0) = -r(\mathbf{y}_0) r''(\mathbf{y}_0) = -r(\mathbf{y}_0) \mathcal{O}(G''(\mathbf{y}_0)) = -r(\mathbf{y}_0) \mathcal{O}(q''(\mathbf{y}_0)) . \quad (54)$$

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 $(\varphi_j)_{j \in \mathbb{N}}$, 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 $(\partial_{y_i y_j}^2 q(\mathbf{y}))_{i,j \in \mathbb{N}} \in \mathcal{X}^{\mathbb{N} \times \mathbb{N}}$.

Proposition 9.1. *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 (5) with $|y_j| \leq 1$ and $\psi_j = c_j \varphi_j$ where, for some $C, \eta > 0$, there holds the bound*

$$\forall j \in \mathbb{N} : \quad c_j := \|\psi_j\|_X \leq C j^{-1-\eta} . \quad (55)$$

Then, for every $\mathbf{y} \in U$, and for every observation functional $\mathcal{O}(\cdot) \in \mathcal{X}'$, the bi-infinite matrix $(\mathcal{O}((\partial_{y_i y_j}^2 q)(\mathbf{y})))_{i,j \in \mathbb{N}}$ corresponding to the second differential $(D_u^2 \mathcal{G})(u)$ of the uncertainty-to-observation map with u as in (5) induces a compact operator on $\ell^2(\mathbb{N})$.

Proof. Fix $\mathbf{y} \in U$ 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 Hesse matrix $H(\mathbf{y}) := (\mathcal{O}((\partial_{y_i y_j}^2 q)(\mathbf{y})))_{i,j \in \mathbb{N}}$ can be approximated, in norm, by a sequence $(H^J(\mathbf{y}))_{J \in \mathbb{N}}$ 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’’ $\|\cdot\|_F$ given by $\|B\|_F^2 = \sum_{i,j} |B_{ij}|^2$ (which majorizes the induced spectral-norm), it is sufficient to approximate $H(\mathbf{y})$ in $\|\cdot\|_F$ 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 , ie.

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

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

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

Using that there exists $C > 0$ such that for every $\mathbf{y} \in U$

$$\|\partial_{y_i y_j}^2 q(\mathbf{y})\|_{\mathcal{X}} \leq C c_i c_j \quad \forall i, j \in \mathbb{N} \quad (56)$$

we find with the assumption (55) on the size of c_j that

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

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

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

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

$$\sup_{\mathbf{y} \in U} \|H(\mathbf{y}) - H^J(\mathbf{y})\|_F \lesssim \|\mathcal{O}\|_{\mathcal{X}'} J^{-\eta} \quad J \rightarrow \infty . \quad (57)$$

□

Remark 9.2. Condition (55) implies that the sequence $c = (c_j)_{j \geq 1} \in \ell^p(\mathbb{N})$ for $1/(1+\eta) < p < 1$, ie. for $\eta < 1/p - 1$ which, for small $\eta > 0$, is close to the minimal condition $p = 1$ for the parametrization (5) to remain meaningful. Compare to condition (8) in (iii) of the affine perturbation Assumption 2.1. We also note that the truncation $H^J(\mathbf{y})$ of the Hessian corresponds to J -term truncation of the uncertainty parametrization (5). When (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 (57) 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 (57) with (13), (14).

Remark 9.3. 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 [23, Sections 4.3 and 4.4] for numerical evidence and to [2] for an analogous compactness result in inverse shape problems in 2-d, acoustic scattering; both these applications are covered by Proposition 9.1.

10. APPENDIX C: QUASI-NEWTON POSTERIOR CURVATURE ANALYSIS

Based on the Theorems in §3.1, 3.2 and on the compactness result Proposition 9.1, in the regular case discussed in Proposition 8.3, the zero observation noise variance limit of the Bayesian estimate (19) is, upon dimension-truncation, governed by a finite-dimensional, nonlinear least-squares (LSQ) problem

$$S(\mathbf{y}) \stackrel{!}{=} \max \iff f(\mathbf{y}) := \|r(\mathbf{y})\|_2^2 \stackrel{!}{=} \min \quad (58)$$

where $S(\mathbf{y})$ is as in (20), (28), and where the max resp. min is to be taken over $\mathbf{y} \in U$. In the present section, we review Quasi-Newton methods for the numerical determination of a maximizer \mathbf{y}_0 of $S(\mathbf{y})$. In order to use the proposed asymptotics in the case of small observation noise, we are interested in the optimal point \mathbf{y}_0 as well as in the second order derivative information $S''(\mathbf{y}_0)$. Due to these requirements, standard approaches like Gauss-Newton or Levenberg-Marquardt methods are not feasible as they rely on an approximation of the second order derivatives $S''(\mathbf{y}_0)$ given by (21) with $-r'(\mathbf{y})r'(\mathbf{y})^\top$, ie. the second summation term is neglected which leads to a good approximation of $S''(\mathbf{y}_0)$ in the small residual case. Due to possibly large residuals in the incompatible case, this approach is not valid for the application of interest. Here, we will focus on Quasi-Newton methods combined with symmetric rank-1 Hessian updates (“QN-SR1” for short) and review results on superlinear convergence of iterates of both, parameter values and Hessians, to the critical points and the metric tensor (ie., to \mathbf{y}_0 resp. to $S''(\mathbf{y}_0)$) from [7].

Throughout we place ourselves under the assumptions 1. - 4. of Proposition 8.3. We assume, in particular, that the parametric forward problem has been dimensionally truncated to parameter domain of finite dimension J as outlined in §2.4. We also assume that, due to assumption 4. of Proposition 8.3 and due to (54), that for every value J of the truncation dimension, the Hessian $S''_0(\mathbf{y}_0)$ of the Bayesian potential for the dimensionally truncated forward model is nondegenerate with $\mathbf{y}_0 \in \text{int}U$. Under these assumptions, we recall from (25), (20) that for every $J < \infty$, the limit $\Gamma \downarrow 0$ is determined by the nonlinear least-squares problem (58).

The preceding assumptions ensure that the Hessian $S''_0(\mathbf{y}_0)$ for the dimensionally truncated forward model is nondegenerate. As a consequence of the p -sparsity assumption (iii) of Assumption 2.1 which, by Proposition 9.1, implies compactness of the Hessian $\mathcal{O}(D_u^2 q)(\mathbf{y})$, for large, finite truncation dimension J , the corresponding dimensionally truncated Hessians $H^J(\mathbf{y})$ will, in general, be positive definite, but near degenerate. To overcome the ill-posedness of the finite dimensional optimization problems, we propose a preconditioned trust-region strategy which furthermore yields the necessary globalization of the Quasi-Newton method. Based on the results presented in Proposition 9.1, the size of the updates in each component is controlled according to (56), cp. Remark 10.2. Efficient methods for the numerical solution of (58) have been developed over the past decades; we refer to [11, Chap. 10] and the references there. In the context of inverse problems

for operator equations with uncertain, distributed parameters, we refer to [2] and the references there. One key question in (Quasi-) Newton approaches for the solution of the least-squares problem is the acquisition of first and second order derivative information, in particular on the Hessian $S''(\mathbf{y})$. It is also crucial in determining the nature of the asymptotic expansion in Proposition 8.3. An efficient alternative to finite differencing is the adjoint technique, which allows to compute gradients with a computational effort independent of the number of unknown parameters J , cp. eg. [15, 17]. Second order derivatives may be evaluated either directly from the (discretized) variational definition in the PDE (see, eg. [17, 23]), or, as we shall now next, be ‘accumulated’ during a Quasi-Newton iteration in $S(\mathbf{y})$ for the computation of \mathbf{y}_0 . As is well-known, the quadratic and superlinear convergence of Newton- and Quasi-Newton methods for the numerical solution of (58) depend crucially on second derivative information of the function to be minimized. In particular, QN methods do not require explicit information about the objective functions’ Hessians and, for SR1 [7] and PSB [14] rank-1 updates, generate sequences $\{H_k\}_{k \geq 0}$ of approximate Hessians which will converge superlinearly with the iterates to the true Hessian $S''(\mathbf{y}_0)$. Basic results on this are [14] and [7, Thm.2, eqn. (36)]. We show in Theorem 4.1 that the *curvature information produced as a “byproduct” during the QN iteration for the determination of \mathbf{y}_0 can be exploited to achieve Γ -independent performance of adaptive Smolyak (and other) quadrature algorithms* for positive, but small observation noise variance $\Gamma > 0$. This finding is consistent with recent work [16, 23] where QN updates have been found to accelerate burn-in and achieve variance reduction in MCMC methods.

We recapitulate next the main results on the Hessian approximation; to do so, we place ourselves under the assumptions 1. - 4. of Proposition 8.3. Then QN methods construct, based on a starting point $\mathbf{y}^{(0)} \in U$ which is sufficiently close to the global minimum \mathbf{y}_0 of f a sequence $\{\mathbf{y}^{(k)}\}_{k \geq 0}$ which converges to \mathbf{y}_0 at least superlinearly, ie.

$$\mathbf{y}_0 = \lim_{k \rightarrow \infty} \mathbf{y}^{(k)}, \quad \|\mathbf{y}_0 - \mathbf{y}^{(k)}\|_2 \leq C \|\mathbf{y}_0 - \mathbf{y}^{(k-1)}\|_2^s$$

for all $k \geq k_0 \geq 1$ and some $C > 0$ and $s > 1$ independent of k .

Algorithm 10.1. Given $\mathbf{y}^{(0)} \in U$ sufficiently close to \mathbf{y}_0 , repeat until convergence

Step 1: Set $k = 0$, $H_0 := I$ and compute $\mathbf{g}_0 = \nabla f(\mathbf{y}^{(0)})$.

Step 2: Find a direction vector \mathbf{s}_k .

Step 3: Set

$$\mathbf{y}^{(k+1)} := \mathbf{y}^{(k)} + \mathbf{s}_k,$$

and compute

$$\mathbf{g}_{k+1} := \nabla f(\mathbf{y}^{(k+1)}), \quad \mathbf{x}^{(k)} := \mathbf{g}_{k+1} - \mathbf{g}_k.$$

Step 4: update H_k to a new matrix H_{k+1} , such that $\text{rank}(H_{k+1} - H_k) = 1$.

Step 5: set $k := k + 1$ and go to Step 2.

In Step 4, we consider the symmetric rank-1 (SR1) update

$$H_{k+1} := H_k + \frac{r_k \mathbf{s}_k^\top}{r_k^\top \mathbf{s}_k} \quad (59)$$

with $r_k := \mathbf{x}^{(k)} - H_k \mathbf{s}_k$.

QN methods with SR1 updates are known to produce function values and approximate Hessians which converge superlinearly locally, ie., near nondegenerate extrema of the objective function.

Proposition 10.1. ([7]) *Assume that*

(AS.1) $f(\mathbf{y}) \in C^2$,

(AS.2) $H(\mathbf{y}) = f''(\mathbf{y})$ is Lipschitz,

(AS.3) the sequence $\{\mathbf{y}^{(k)}\}_{k \geq 0}$ converges to a limit point \mathbf{y}_* ,

(AS.4) the sequence $\{s_k\}_{k \geq 0}$ is uniformly linearly independent, ie. there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ the matrices

$$S_k^* := \left(\frac{s_k}{\|s_k\|}, \dots, \frac{s_{k+J-1}}{\|s_{k+J-1}\|} \right)$$

are nonsingular

where $\{\mathbf{y}^{(k)}\}_{k \geq 0}$ is a sequence of iterates generated by Algorithm 10.1 with SR1 update (59). Assume also that for all k there holds

$$|r_k^\top s_k| \geq c_1 \|r_k\| \|s_k\| ,$$

with $c_1 \in (0, 1)$. Then, there exists a constant $c > 0$ such that, for $k \geq k_0$,

$$\|H_{k+J+1} - H(\mathbf{y}_0)\| \leq c\epsilon_k,$$

where, for $m \geq 1$ we defined $\epsilon_k := \max\{\|\mathbf{y}^{(l)} - \mathbf{y}_0\| : k \leq l \leq k + J + 1\}$, and there holds $\lim_{k \rightarrow \infty} \|H_k - H(\mathbf{y}_0)\| = 0$. If, moreover, for sufficiently large k ,

$$\|\mathbf{y}^{(k+1)} - \mathbf{y}_0\| \leq c\|\mathbf{y}^{(k)} - \mathbf{y}_0\|, \quad \text{then} \quad \|H_{k+m+1} - H(\mathbf{y}_0)\| \leq c\|\mathbf{y}^{(k)} - \mathbf{y}_0\| .$$

ie., for these values of k holds superlinear convergence of the accumulated SR1 updated Hessians.

The stated results follow from [7, Thm. 2, Cor. 3, Cor.4]. We mention that assumption (AS.4) above corresponds to the (stronger) assumption (AS.4b) in [7]. In what follows, we apply the QN Algorithm 10.1 with SR1 update (59) to

$$\min_{\mathbf{y} \in U} \Phi_\Gamma(\mathbf{y}; \delta) .$$

Remark 10.2. The updates in Step 3 in Algorithm 10.1 are computed by a trust region strategy leading to a globalization of the Quasi-Newton method, ie. to obtain the new iterate $\mathbf{y}^{(k+1)}$, we seek a solution of the following subproblem

$$\min f(\mathbf{y}_k) + g_k^\top s + \frac{1}{2} s^\top H_k s \quad \text{s.t.} \quad \|s\| \leq \Delta_k , \quad (60)$$

where $\Delta_k > 0$ is the trust-region radius. We refer to [8] for details on trust-region methods. In particular, we consider a weighted norm in (60) of the form $\|Ds\| \leq \Delta_k$ with a nonsingular, diagonal and positive scaling matrix D in order to overcome the inherently present ill-posedness of the underlying least-squares problem due to the compactness of Hessians shown in Proposition 9.1. Based on the bounds on the second order derivatives given in (56), we propose $D := \text{diag}(c_1^2, \dots, c_J^2)$ reducing the effects of poor scaling, cp. [26, Chapter 4,10].

We note that the smoothness assumption (AS.2) is satisfied due to the Bayesian potential being a quadratic functional of the uncertainty-to-observation map, and due to the analytic dependence of $q(\mathbf{y})$ on the coordinates of \mathbf{y} .

Due to the local nature of the convergence result Proposition 10.1, we always assume that the starting point of our iteration is feasible and sufficiently close to \mathbf{y}_0 .

Remark 10.3. Under the assumption of the existence of a unique nondegenerate minimizer $\mathbf{y}_0 \in \text{int}U$ and of a starting point sufficiently close to $\mathbf{y}_0 \in \text{int}U$, we may omit the box constraints on the parameter \mathbf{y} and consider the unconstrained optimization problem $\min f(\mathbf{y})$. In the case $\mathbf{y} \in \partial U$ where the box constraints are active at \mathbf{y}_0 a nonlinear programming problem

$$\min f(\mathbf{y}) \quad \text{s.t.} \quad -1 \leq y_j \leq 1, \quad j = 1, \dots, J$$

needs to be solved. In [6], a SR1 QN method with projection to the feasible set in each iteration is proven to converge superlinearly.

Remark 10.4. Under the nondegeneracy assumption, Proposition 10.1 implies superlinear local convergence of iterates and updated Hessians. For large dimension J , termination of the QN iteration could well be reached in less than J iterations. In this case, the rank of the converged QN Hessian could be used as an estimate of the effective concentration dimension J_* of the posterior at the MAP point.

11. APPENDIX D: ADAPTIVE SMOLYAK QUADRATURE IN SPHERICAL POLAR COORDINATES

The asymptotic analysis presented in §3 ensures that, in the limit $\Gamma \downarrow 0$, only terms up to order 2 of the parametric posterior density contribute to the Bayesian estimate $\mathbb{E}^{\mu^\delta}[\phi]$. This asymptotically spherical behavior of the posterior density upon rescaling could be clearly observed in the numerical results, cp. Fig. 2 and Fig. 5. Theorem 4.1 also indicates that the curvature rescaled posterior density is an analytic function of the parameters, with uniform (w.r. to $\Gamma_{obs} > 0$) domain of analyticity.

For Gaussian prior π_0 , however, we found that the Smolyak algorithm based on Gauss-Hermite quadrature could not exploit this behavior due to an additional transformation factor arising from the weight function $\exp(-0.5\|z\|^2)$, ie. to compute the Bayesian estimate Z'/Z with

$$Z = C \int_{\mathbb{R}^J} \exp\left(-\frac{1}{2}\|\delta - \mathcal{G}(QM^{-\frac{1}{2}}z + \mathbf{y}_0)\|_\Gamma^2\right) \exp\left(-\frac{1}{2}\|QM^{-\frac{1}{2}}z + \mathbf{y}_0\|^2\right) dz,$$

and

$$Z' = C \int_{\mathbb{R}^J} \phi(QM^{-\frac{1}{2}}z + \mathbf{y}_0) \exp\left(-\frac{1}{2}\|\delta - \mathcal{G}(QM^{-\frac{1}{2}}z + \mathbf{y}_0)\|_\Gamma^2\right) \exp\left(-\frac{1}{2}\|QM^{-\frac{1}{2}}z + \mathbf{y}_0\|^2\right) dz,$$

a factor $\exp(0.5\|z\|^2) \exp(-0.5\|z\|^2)$ needs to be introduced

$$\begin{aligned} Z &= C \int_{\mathbb{R}^J} \exp\left(-\frac{1}{2}\|\delta - \mathcal{G}(QM^{-\frac{1}{2}}z + \mathbf{y}_0)\|_\Gamma^2\right) \\ &\quad \times \exp\left(-\frac{1}{2}\|QM^{-\frac{1}{2}}z + \mathbf{y}_0\|^2\right) \exp(0.5\|z\|^2) \exp(-0.5\|z\|^2) dz \\ &= C \int_{\mathbb{R}^J} \exp\left(-\frac{1}{2}\|\delta - \mathcal{G}(QM^{-\frac{1}{2}}z + \mathbf{y}_0)\|_\Gamma^2\right) \exp\left(-\frac{1}{2}\|QM^{-\frac{1}{2}}z + \mathbf{y}_0\|^2\right) \\ &\quad \times \exp(0.5\|z\|^2) \mu_0(dz), \end{aligned}$$

which will not preserve the asymptotically quadratic w.r. to z posterior density. This effect is illustrated in Fig. 12 showing the transformed posterior of the diffusion problem with lognormal coefficients ($J = 2$) discussed in §5, cp. 5.

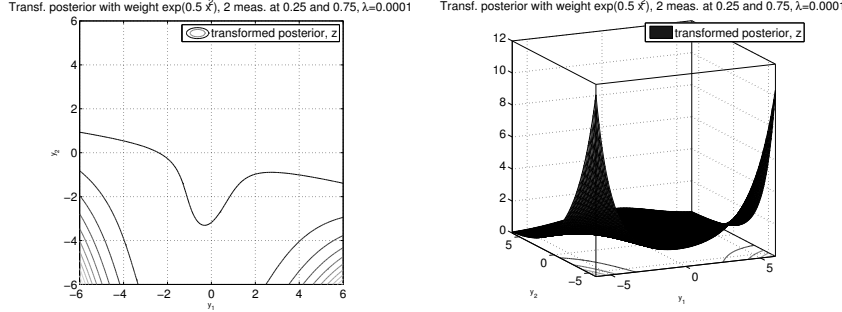


FIGURE 12. Plots of the transformed posterior density with additional weight function $\exp(0.5\|z\|^2)$, observational noise $\Gamma_{obs} = 0.01^2 I$. The asymptotically spherical level-lines of the rescaled posterior density, cp. 5, are not preserved due to an additional transformation factor arising from the use of Gauss-Hermite quadrature formulas.

In concentration dimension $J_* = 2$, we found numerical integration in polar coordinates able to exploit the asymptotic behavior of the posterior density, as illustrated in Fig. 7. We will generalize this idea in the following to the case of posterior concentration in $J_* > 2$ concentration variables. Specifically, we propose an adaptive Smolyak algorithm on the sphere S^{J_*} tailored to exploit the analytic structure of the curvature-rescaled posterior density. We are interested in the computation of improper integrals of the form

$$\int_{\mathbb{R}^J} f(z) \exp\left(-\frac{1}{2}\|\varphi(z)\|^2\right) dz, \quad (61)$$

where φ denotes the curvature-rescaling transformation (cp. (34)) $\varphi(z) = \mathbf{y}_0 + \Gamma^{\frac{1}{2}} Q M^{-\frac{1}{2}} z$. Introducing spherical polar coordinates $\theta_1, \theta_2, \dots, \theta_{J-1}, r$ defined by

$$\begin{aligned} z_1 &= r \prod_{i=1}^{J-1} \sin \theta_i \\ z_j &= r \cos \theta_{j-1} \prod_{i=j}^{J-1} \sin \theta_i, \quad j = 2, \dots, J \end{aligned}$$

with $0 \leq \theta_1 \leq 2\pi, 0 \leq \theta_j \leq \pi, j = 2, \dots, J-1$ and $0 \leq r < \infty$, (61) can be transformed to

$$\int_0^\infty \int_0^\pi \dots \int_0^\pi \int_0^{2\pi} J(\theta, r) f(\theta, r) \exp\left(-\frac{1}{2}\|\varphi(\theta, r)\|^2\right) d\theta_1 d\theta_2 \dots d\theta_{J-1} dr,$$

where $\theta = (\theta_1, \dots, \theta_{J-1})$ and the Jacobian $J(\theta, r)$ of the transformation is given by

$$J(\theta_1, \dots, \theta_{J-1}, r) = r^{J-1} \sin \theta_2 \sin^2 \theta_3 \dots \sin^{J-2} \theta_{J-1}, \quad (64)$$

see eg. [25]. Note that the transformation to spherical polar coordinates leads to a tensor-product structure of the Jacobian and the hierarchical structure with respect to the angles $\theta_1, \dots, \theta_{J-1}$ is amenable to dimension-adaptive Smolyak quadrature as described in [27, 28] in the ‘‘concentrating variable’’ $y_j, j = 1, \dots, J_*$ with $1 \leq J_* \leq J$ (cp. (37) and Remark 4.2). By analogy to the two-dimensional example discussed in §5, we suggest the trapezoidal rule to integrate with respect to the azimuth angle θ_1 and Gauss-Laguerre quadrature to integrate with respect to the radius r . Specifically, Gauss-Laguerre formulas with weight function $r \exp(-r)$, leading to an additional factor $\exp(r)$, which preserves the spherical behavior of the posterior density. We

remark that the integrand function explicitly depends on the active dimension J_* via the Jacobian of the transformation (64). This means that, using an adaptive control of the active dimension in the Smolyak algorithm, the previous estimate of the Smolyak algorithm has to be corrected by the factor r in the integrand. This can be done efficiently by storing the forward solutions for each quadrature point, so that there is no need of recomputing the quadrature points when enlarging the active dimension. In the remaining dimensions $\theta_2, \dots, \theta_{J-1}$, we propose to use Gauss-Jacobi formulas based on weights $\sin^{j-1} \theta_j = \sqrt{1 - \cos^2 \theta_j}^{j-1}$ for $j = 2, \dots, J - 1$.

Using the Gauss-Laguerre formulas with weight functions $r \exp(-r)$ and the Gauss-Jacobi formulas as described above, we construct an adaptive Smolyak algorithm with dimension dependent quadrature formulas, optimized to exploit the structure of the integrand based on the asymptotic analysis presented. To determine the dimension J of the truncated parameter space, ie. the most sensitive parameters with respect to the given observational data δ , the additional information provided by the Quasi-Newton method with SR1 updates based on the strategy discussed in §4.1 can be used. The hyperspherical adaptive Smolyak to compute the Bayesian quantities Z' and Z can be summarized as follows:

- Use a QN-Algorithm 10.1 with SR1 updates to determine the MAP \mathbf{y}_0 point and the effective dimension J_* based on the eigenvalues of the QN-approximated Hessian.
- Apply the adaptive Smolyak algorithm in hyperspherical coordinates with dimension dependent quadrature formulas to the integrals Z' and Z with curvature-rescaled densities to compute the Bayesian estimate $\mathbb{E}^{\mu^\delta}[\phi]$.

REFERENCES

- [1] M. Atyiah, Resolution of Singularities and division of distributions, *Comm. Pure Appl. Math.* **23** 145-150, 1970.
- [2] 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** 055002, 2012.
- [3] J. Charrier, Strong and weak error estimates for the solutions of elliptic PDEs with random coefficients *SIAM J. Numer. Analysis* **50**(1), 216D246, 2013.
- [4] A. Chkifa, A. Cohen and C. Schwab, Breaking the curse of dimensionality in sparse polynomial approximation of parametric PDEs *Journ. Math. Pures Appliquees*, 2014.
- [5] A. Cohen, R. DeVore and C. Schwab, Analytic regularity and polynomial approximation of parametric and stochastic elliptic PDEs *Analysis and Applications* **9** 1-37, 2011.
- [6] A. Conn, N. Gould and P. Toint, Global Convergence of a Class of Trust Region Algorithms for Optimization with Simple Bounds *SIAM Journal on Numerical Analysis* **25** 433-460, 1988.
- [7] A. Conn, N. Gould and P. Toint, Convergence of Quasi-Newton matrices generated by the symmetric rank-one update *Mathematical Programming* **50**(1991) 177-195, 1991.
- [8] A. Conn, N. Gould and P. Toint, Trust Region Methods *MPS-SIAM Series on Optimization*, SIAM, 2000.
- [9] M. Dashti, K. Law, A. M. Stuart and J. Voss, MAP Estimators and their consistency in Bayesian nonparametric inverse problems *Inverse Problems* **29** 095017, 2013.
- [10] M. Dashti and A. M. Stuart, The Bayesian Approach to Inverse Problems ArXiv arXiv:1302.6989v2 Version 7 Aug 2014.
- [11] J. E. Dennis, Jr., and R. B. Schnabel, Numerical methods for unconstrained optimization and nonlinear equations *Classics in Applied Mathematics* **16** SIAM, 1996.
- [12] M. V. Fedoryuk, Asimptotika: integraly i ryady Spravochnaya Matematicheskaya Biblioteka. [Mathematical Reference Library] (in russian). Nauka Publ., Moscow, Russia, 1987.
- [13] R. Fletcher, An Overview of Unconstrained Optimization in Algorithms for Continuous Optimization, ed. Spedicato, E, Springer Netherlands, 109-14, 1994.
- [14] R. P. Ge and M. J. D. Powell, The convergence of variable metric matrices in unconstrained optimization, *Mathematical Programming* **27** 123-143, 1983.
- [15] M. B. Giles and N. A. Pierce, An Introduction to the Adjoint Approach to Design *Flow, Turbulence and Combustion* **65** 393-415, 2000.
- [16] M. Girolami and B. Calderhead, Riemann manifold Langevin and Hamiltonian Monte Carlo methods *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**(2) 123- 214, 2011.
- [17] A. Griewank and A. Walther, Evaluating Derivatives: Principles and Techniques of Algorithmic Differentiation *SIAM 2nd edn.*, 2008.
- [18] M. Hansen and C. Schwab, Analytic regularity and best N -term approximation of high dimensional parametric initial value problems *Vietnam Journ. Mathematics*, 2013.

- [19] V. H. Hoang and C. Schwab, Regularity and Generalized Polynomial Chaos Approximation of Parametric and Random Second-Order Hyperbolic Partial Differential Equations *Analysis and Applications (Singapore)* **10**(3), 2012.
- [20] V. H. Hoang and C. Schwab, Analytic regularity and polynomial approximation of stochastic, parametric elliptic multiscale PDEs, *Analysis and Applications (Singapore)* **11**(1) 1350001, 2013.
- [21] J. Kaipio and E. Somersalo, *Statistical and computational inverse problems* Applied Mathematical Sciences **160** Springer, 2005.
- [22] B. T. Knapik, A. W. van der Vaart and J. H. van Zanten, Bayesian inverse problems with Gaussian priors *The Annals of Statistics* **39** 2626-2657, 2011.
- [23] J. Martin, L. C. Wilcox, C. Burstedde and O. Ghattas, A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion *SIAM Journal on Scientific Computing* **34** A1460-A1487, 2012.
- [24] T. A. Moselhy and Y. M. Marzouk, Bayesian inference with optimal maps *Journal of Computational Physics* **231** 7815-7850, 2012.
- [25] D. Mustard, Numerical integration over the n-dimensional spherical shell *Math. Comp.* **18** 578-589, 1964.
- [26] J. Nocedal and S. J. Wright Numerical Optimization *Springer*, 2nd Edn., 1999.
- [27] C. Schillings and C. Schwab, Sparse, Adaptive Smolyak Algorithms for Bayesian Inverse Problems, *Inverse Problems* **29**(6), 2013.
- [28] C. Schillings and C. Schwab, Sparse Quadrature Approach to Bayesian Inverse Problems *Inverse Problems* **30**(6), 2014.
- [29] C. Schwab and A. M. Stuart, Sparse Deterministic Approximation of Bayesian Inverse Problems *Inverse Problems* **28** 045003, 2012.
- [30] A. Sidi, On a generalization of the Richardson extrapolation process *Numer. Math.* **57** 365-377, 1990.
- [31] A. M. Stuart, Inverse problems: a Bayesian approach *Acta Numerica* **19** 451-559, 2010.

Recent Research Reports

Nr.	Authors/Title
2014-16	P. Grohs and S. Keiper and G. Kutyniok and M. Schaefer α -Molecules
2014-17	A. Hildebrand and S. Mishra Efficient computation of all speed flows using an entropy stable shock-capturing space-time discontinuous Galerkin method
2014-18	D. Conus and A. Jentzen and R. Kurniawan Weak convergence rates of spectral Galerkin approximations for SPDEs with nonlinear diffusion coefficients
2014-19	J. Doelz and H. Harbrecht and Ch. Schwab Covariance regularity and H-matrix approximation for rough random fields
2014-20	P. Grohs and S. Hosseini Nonsmooth Trust Region Algorithms for Locally Lipschitz Functions on Riemannian Manifolds
2014-21	P. Grohs and A. Obermeier Optimal Adaptive Ridgelet Schemes for Linear Transport Equations
2014-22	S. Mishra and Ch. Schwab and J. Sukys Multi-Level Monte Carlo Finite Volume methods for uncertainty quantification of acoustic wave propagation in random heterogeneous layered medium
2014-23	J. Dick and Q. T. Le Gia and Ch. Schwab Higher order Quasi Monte Carlo integration for holomorphic, parametric operator equations
2014-24	C. Sanchez-Linares and M. de la Asuncion and M. Castro and S. Mishra and J. Šukys Multi-level Monte Carlo finite volume method for shallow water equations with uncertain parameters applied to landslides-generated tsunamis
2014-25	R.N. Gantner and Ch. Schwab Computational Higher Order Quasi-Monte Carlo Integration