Underdetermined Bayesian experimental designs and the Laplace approximation

Quan Long[*]; Marco Scavino[†] [*]; Raúl Tempone[*]; Suojin Wang[††]

[*] SRI Center for Uncertainty Quantification
Computer, Electrical and Mathematical Sciences & Engineering Division,
King Abdullah University of Science and Technology, KSA
[†]Unidad de Estadística – IESTA, Facultad de Ciencias Económicas y de Administración,
UdelaR, Montevideo, Uruguay
[††]Department of Statistics, Texas A&M University, USA

Abstract

In this work we address the problem to determine the information degree of a statistical experiment, planned in a Bayesian design framework, to calibrate the parameters of a chosen statistical model or to predict some quantity of interest.

Along the lines of our previous work (Long, 2013) we develop a fast approach based on the Laplace approximation to estimate the expected information gain (expected Kullback–Leibler divergence) for model parameter inferences or the prediction of quantities of interest, in the case of underdetermined statistical models.

We demonstrate the accuracy, efficiency and robustness of the proposed method via some numerical examples, such as the design of a scalar parameter in a model expressed through a one dimensional cubic polynomial function with two indistinguishable parameters. Our method clearly outperforms the estimation techniques that rely solely on Monte Carlo methods for multidimensional integration.

Keywords: Bayesian experimental design, information gain, Laplace approximation, Monte Carlo sampling, uncertainty quantification.

1 Introduction

In Bayesian experimental design the information of a proposed experiment is usually measured by the expected information gain, i.e., the expected log ratio between the posterior and prior distribution for the parameters in the statistical model (see, for instance, Lindley, 1956, Chaloner, 1995, Ginebra, 2007). The computation of the expected information gain is commonly analytically intractable and computationally very expensive, particularly when the outcomes of the experiment are modeled as functions of the solution of Partial Differential Equations (PDEs). Using the Laplace approximation we proposed in (Long, 2013) a fast approach for the estimation of the expected information gain and we analyzed the rates of different dominant error terms with respect to the amount of data in each experiment scenario, provided that the parameters can be determined completely through the experiments. When both the determinant of the posterior covariance matrix and the prior probability density functions (pdf) satisfy certain regularity conditions with respect to the random parameters, we demonstrated, by means of several nonlinear examples, also involving the solution of PDEs, that sparse quadrature techniques can be employed to carry out the integration steps with high efficiency.

In this work, we extend the methodology developed in (Long, 2013) to the cases where the random parameters are not determined completely through the experiments, i.e., models with underdetermined parameters. We assume the existence of an embedded manifold on which the parameters are not distinguishable by the data. In this context, the posterior pdf will start to concentrate around this manifold as the amount of data increases. The key innovation of our novel extension consists in performing the normality approximation for the conditional posterior pdf given a fixed point on the manifold, and in the use of Laplace approximation for the evaluation of the conditional expected information gain. Both approximations are carried out along the directions which are orthogonal to the indistinguishable manifold. Asymptotic expansions of the expected Kullback–Leibler divergence between the posterior and prior pdfs for determined models have been derived by several authors using the likelihood ratio process. See, for instance, (Clarke, 1991) for an interesting connection with an information–theoretic framework of the Bayesian Central Limit Theorem, (Polson, 1992) for an extension to non i.i.d. regular models for experimental designs, and (Ghosal, 1997) for the analysis of non–regular models when the posterior distribution is consistent. Other works were inspired by (Bernardo, 1979), whose motivation was to justify the use of certain prior distributions in Bayesian statistical analysis by maximizing the Shannon mutual information between the parameter vector and the data, also in the presence of nuisance parameters in the model. These works, see for example (Polson, 1988, Clarke, 1993 y 2004), making an explicit distinction between parameters of interest and nuisance parameters in the model, can be used, in principle, to obtain asymptotic expressions of the expected information gain for underdetermined models. However, their applicability is confined to the case where the indistinguishable manifold can be explicitly parametrized in terms of the nuisance parameters. Instead, our approach does not require such an explicit representation of the underlying manifold where the posterior distribution concentrates. On the other hand, an explicit representation of the manifold is practically not possible. In Section 2, we formulate our new methodology for parameter inference: we first introduce the information gain and the expected information gain. We then reparametrize the prior and posterior pdfs, using two set of local parameters, $\boldsymbol{t}$ and $\boldsymbol{s}$, separately. The $\boldsymbol{t}$ direction parametrizes the indistinguishable manifold, while $\boldsymbol{s}$ parametrizes the directions orthogonal to the manifold. Next, the Laplace approximation is carried out along the $\boldsymbol{s}$ direction, conditioned on a fixed $\boldsymbol{t}$ value. Finally, the information gain is expressed as an integral along the $\boldsymbol{t}$ direction. By an extra integral over the data, we obtain the asymptotic formulation of the expected information gain. Section 3 applies similar ideas to the prediction of quantities of interest. In the Section 4 we describe briefly the numerical methods used to approximate the integrals that appear in the asymptotic expressions of the expected information gain and the expected conditional entropy. Some numerical examples are presented in Section 5, including the designs of a scalar parameter in a one dimensional function with two indistinguishable parameters.

2 Estimation of the expected information gain for an underdetermined model

In this section we describe our new methodology for parameter inference.

2.1 The general model, information gain and expected information gain

To introduce the main ideas, we consider the following model of both multidimensional parameters and multidimensional outputs. For simplicity of exposition only, let us assume additive Gaussian measurement noise,

$$\boldsymbol{y}_i = \boldsymbol{g}(\boldsymbol{\theta}_0, \boldsymbol{\xi}) + \boldsymbol{\epsilon}_i \,,$$

where $\boldsymbol{g}(\boldsymbol{\theta}_0, \boldsymbol{\xi})$ is the mean vector of $\boldsymbol{y}_i$, which is a known deterministic model in our case, $\boldsymbol{\theta}_0$ is the d dimensional vector of "true" parameters used to generate the synthetic data, $\boldsymbol{\xi}$ is the vector of design parameters, also known as the experimental set–up, $\boldsymbol{y}_i$ is the i^{th} observation vector, and $\boldsymbol{\epsilon}_i$ is assumed to be additive independent and identically distributed (i.i.d.) Gaussian noise, $\boldsymbol{\epsilon}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\epsilon)$, corresponding to the i^{th} observation. Assume now that we are able to collect M observations using the same experimental set–up and that $\bar{\boldsymbol{y}} = \{\boldsymbol{y}\}_{i=1}^{M}$ is a set of observed data points. Essentially, the observations $\bar{\boldsymbol{y}}$ are i.i.d., given specific values of $\boldsymbol{\theta}_0$ and $\boldsymbol{\xi}$.

The Kullback–Leibler (K–L) divergence (information gain) and expected K–L divergence (expected information gain) for a given experiment $\boldsymbol{\xi}$ are defined as follows:

$$D_{KL}(\bar{\boldsymbol{y}}) = \int_{\boldsymbol{\Theta}} \log\left(\frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}})}{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}\right) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}}) d\boldsymbol{\theta}\,,$$

$$I = \int_{\mathcal{Y}} \int_{\boldsymbol{\Theta}} \log\left(\frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}})}{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}\right) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}) p(\bar{\boldsymbol{y}}) d\boldsymbol{\theta} d\bar{\boldsymbol{y}}\,,$$

where $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})$ is the prior pdf of the unknown random parameter $\boldsymbol{\theta}$, $p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}})$ is the posterior pdf of the unknown random parameter $\boldsymbol{\theta}$, and $p(\bar{\boldsymbol{y}})$ is the so–called Bayesian evidence, defined as the marginal likelihood over the parameter, $p(\bar{\boldsymbol{y}}) = \int_{\boldsymbol{\Theta}} p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}\,.$

2.2 The manifold of indistinguishable parameters

For an underdetermined model, given $\boldsymbol{\theta}_0$, we define the manifold

Definition 1.

$$T(\boldsymbol{\theta}_0) := \{\boldsymbol{\theta} \in \boldsymbol{R}^d \,:\, g(\boldsymbol{\theta}) - g(\boldsymbol{\theta}_0) = \boldsymbol{0}\}\,, \tag{1}$$

where d is the dimension of the parameter vector $\boldsymbol{\theta}_0$, and a subdomain that contains $T(\boldsymbol{\theta}_0)$:

Definition 2.

$$\Omega_M(\boldsymbol{\theta}_0) := \{\boldsymbol{x} \in \mathbb{R}^d : dist(\boldsymbol{x}, T(\boldsymbol{\theta}_0)) \leq O(M^{-\alpha})\}\,.$$

$O(\cdot)$ denotes the usual big O notation. Intuitively, as we obtain more data, the posterior pdf of $\boldsymbol{\theta}$ concentrates at the vicinity of $T(\boldsymbol{\theta}_0)$, in such a way that the integration of the information gain can be approximated using firstly the Laplace approximations along the directions orthogonal to $T(\boldsymbol{\theta}_0)$ and secondly an integration over the manifold $T(\boldsymbol{\theta}_0)$. Thus, we define a subdomain $\Omega_M(\boldsymbol{\theta}_0) \subset \boldsymbol{R}^d$ by extending $T(\boldsymbol{\theta}_0)$ along its normal directions, and assume that the length of the extension shrinks to zero at a slower rate than the concentration of the posterior pdf $p(\boldsymbol{\theta}|\bar{\boldsymbol{y}})$. Therefore, the volume of $\Omega_M(\boldsymbol{\theta}_0)$ shrinks to zero at a slower rate than the square root of the number of replicate experiments M, i.e., $\alpha \in (0,\, 0.5)$.

Lemma 1. $\mathbb{P}_r(\boldsymbol{\theta} \in \mathbb{R}^d/\Omega_M(\boldsymbol{\theta}_0)|\bar{\boldsymbol{y}}) = O_P\left(M^{(\alpha-\frac{1}{2})p}\right)$ *when* $M \to \infty$

We leave the proof of this lemma to Section 2.5 after introducing the new local parameters. p denotes the highest order of the available statistical moment of $p(\boldsymbol{\theta}|\bar{\boldsymbol{y}})$. Note that $\alpha < 0.5$ is necessary to ensure a decreasing probability mass outside $\Omega_M(\boldsymbol{\theta}_0)$. For example, if $p = 2$ and $\alpha = 0$, we have $O_P\left(M^{(\alpha-\frac{1}{2})p}\right) = O_P\left(M^{-1}\right)$; if $p = 4$ and $\alpha = 0$, we have $O_P\left(M^{(\alpha-\frac{1}{2})p}\right) = O_P\left(M^{-2}\right)$. From now on, we assume that $O_P\left(M^{(\alpha-\frac{1}{2})p}\right)$ is smaller than $O_P\left(M^{-1}\right)$.

The K–L divergence can then be written as the sum of two terms, namely

$$D_{KL}(\bar{\boldsymbol{y}}) = \int_{\Omega_M(\boldsymbol{\theta}_0)} \log\left(\frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}})}{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}\right) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}}) d\boldsymbol{\theta} + \epsilon_{\Omega_M} \tag{2}$$

with

$$\epsilon_{\Omega_M} = \int_{\boldsymbol{\Theta}-\Omega_M(\boldsymbol{\theta}_0)} \log\left(\frac{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}})}{p_{\boldsymbol{\Theta}}(\boldsymbol{\theta})}\right) p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}}) d\boldsymbol{\theta} = O_P\left(M^{(\alpha-\frac{1}{2})p}\right), \quad \text{when} \quad M \to \infty\,.$$

The rate of ϵ_{Ω_M} is a direct consequence of Lemma 1.

2.3 Local coordinates and weight functions

For the purpose of conciseness, we use T and Ω_M instead of $T(\boldsymbol{\theta}_0)$ and $\Omega_M(\boldsymbol{\theta}_0)$, respectively, in the remainder of this work. We define a new set of parameters $\boldsymbol{t}$ and $\boldsymbol{s}$ for our estimation of D_{KL}: $\boldsymbol{t}$ parameterizes the manifold T and $\boldsymbol{s}$ parameterizes the direction orthogonal to T. Specifically, the direction orthogonal to T is defined as the orthonormal complement of the kernel of the Jacobian of our model $\boldsymbol{g}$, i.e., $\mathbf{Ker}(\boldsymbol{J}_g)^{\perp}$. Observe that the $\mathbf{Ker}(\boldsymbol{J}_g)$ contains the directions tangent to the manifold at $\boldsymbol{t}$ given $\boldsymbol{\theta} \in T$. We define the following diffeomorphism mapping:

Definition 3.

$$\boldsymbol{f} : \Omega_{Ms,t} \to \Omega_M\,, \tag{3}$$

where $\Omega_{M\boldsymbol{s},\boldsymbol{t}}$ is the $(\boldsymbol{s},\boldsymbol{t})$ space, which is asymptotically a rectangular, ignoring possible boundary effects, i.e., $\Omega_{M\boldsymbol{s},\boldsymbol{t}} = [-O(M^{-\alpha}), O(M^{-\alpha})] \times T_{\boldsymbol{t}}$.

Here, $[-O(M^{-\alpha}), O(M^{-\alpha})]$ is the range of parameter $\boldsymbol{s}$, and $T_{\boldsymbol{t}}$ is the set containing all the values of the parameter $\boldsymbol{t}$. Observe that all these objects depend on $\boldsymbol{\theta}_0$. For the purpose of conciseness, we do not write this dependence explicitly, so instead of $T_{\boldsymbol{t}}(\boldsymbol{\theta}_0)$ we write $T_{\boldsymbol{t}}$ in the remainder of this work. Here, we let $\boldsymbol{J}$ be the Jacobian of this mapping with respect to $(\boldsymbol{s},\boldsymbol{t})$. Figure 1 illustrates such an indistinguishable manifold T, the orthogonal direction S, and the subdomain Ω_M.

Generally we are not able to give an explicit parameterization of the manifold T. Nevertheless, we can give the explicit form of the local coordinate $\boldsymbol{s}$ as follows. We first define a log likelihood cost function F given by

$$F(\boldsymbol{\theta}) := \frac{1}{2}(\boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{g}(\boldsymbol{\theta}_0))^T \boldsymbol{\Sigma}_{\epsilon}^{-1}(\boldsymbol{g}(\boldsymbol{\theta}) - \boldsymbol{g}(\boldsymbol{\theta}_0))\,.$$

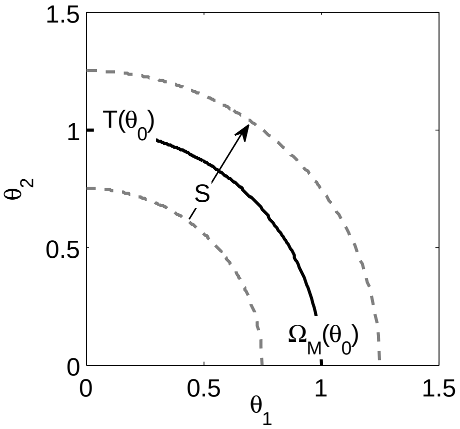


Figure 1: An illustrative indistinguishable manifold in two dimensional parameter space.

We subsequently perform the eigenvalue decomposition of the Hessian of $F(\boldsymbol{\theta})$ at $\boldsymbol{f}(\boldsymbol{0},\boldsymbol{t})$ on the manifold (by the construction we have that $\boldsymbol{s}=\boldsymbol{0}$ on the manifold T) as follows:

$$\boldsymbol{H}(\boldsymbol{f}(\boldsymbol{0},\boldsymbol{t})) = [\boldsymbol{U}\,\boldsymbol{V}]\,\boldsymbol{\Lambda}\,[\boldsymbol{U}\,\boldsymbol{V}]^T \ , \tag{4}$$

where $\boldsymbol{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\boldsymbol{H}(\boldsymbol{f}(\boldsymbol{0},\boldsymbol{t}))$, $\boldsymbol{U}$ is a matrix whose columns are the basis corresponding to the positive eigenvalues of $\boldsymbol{H}(\boldsymbol{f}(\boldsymbol{0},\boldsymbol{t}))$, and $\boldsymbol{V}$ is a matrix whose columns are the basis corresponding to the zero eigenvalues of $\boldsymbol{H}(\boldsymbol{f}(\boldsymbol{0},\boldsymbol{t}))$. Then, we can locally define $\boldsymbol{s}$ at the vicinity of the point $\boldsymbol{f}(\boldsymbol{0},\boldsymbol{t})$ as follows

Definition 4.

$$\boldsymbol{s} = \boldsymbol{U}^T(\boldsymbol{\theta} - \boldsymbol{f}(\boldsymbol{0},\boldsymbol{t}))\,,$$

which is a vector of length $r = rank(\boldsymbol{H}(\boldsymbol{f}(\boldsymbol{0},\boldsymbol{t})))$.
Meanwhile, $\boldsymbol{t}$ is a vector of length $d-r$. In this work we assume that r does not change value w.r.t. $\boldsymbol{\theta}$. Observe now that with this decomposition, we can express Ω_M equivalently as follows:

$$\Omega_M = \left\{\boldsymbol{\theta} : \boldsymbol{\theta} \in \boldsymbol{R}^d;\, \boldsymbol{\theta} = \boldsymbol{\theta}_t + \boldsymbol{\theta}_s;\, \boldsymbol{\theta}_t \in T;\, \boldsymbol{\theta}_s = \boldsymbol{s}\boldsymbol{U}(\boldsymbol{\theta}_t);\, ||\boldsymbol{\theta}_s|| < O(M^{-\alpha})\right\}$$

for some $\alpha \in (0, 0.5)$.
Keeping in mind that we intend to carry out the Laplace approximation in $\boldsymbol{s}$ direction, expressing the related pdfs, e.g., $p(\boldsymbol{\theta})\,, p(\boldsymbol{\theta}|\bar{\boldsymbol{y}})$, in terms of local coordinates $\boldsymbol{s}$ and $\boldsymbol{t}$ is necessary. We consequently define two weight functions of $(\boldsymbol{s},\boldsymbol{t})$ through a change of variables from the pdfs of $\boldsymbol{\theta}$:

Definition 5.

$$p(\boldsymbol{s},\boldsymbol{t}) := p_{\boldsymbol{\Theta}}(\boldsymbol{f}(\boldsymbol{s},\boldsymbol{t}))|\boldsymbol{J}|\,, \tag{5}$$

and

Definition 6.

$$p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}}) := p_{\boldsymbol{\Theta}}(\boldsymbol{f}(\boldsymbol{s},\boldsymbol{t})|\bar{\boldsymbol{y}})|\boldsymbol{J}|\,. \tag{6}$$

Here, $p(\boldsymbol{s},\boldsymbol{t})$ and $p(\boldsymbol{s},\boldsymbol{t}|\{\boldsymbol{y}_i\})$ are the prior and posterior weight functions of $(\boldsymbol{s},\boldsymbol{t})$, respectively. Observe that (5) and (6) are like a standard change of variables. However, the integrations of both weight functions over $\Omega_{M\boldsymbol{s},\boldsymbol{t}}$ do not equal to 1. i.e., $\mathbb{P}(\boldsymbol{\theta}\in\Omega_m) < 1$ and $\mathbb{P}(\boldsymbol{\theta}\in\Omega_m|\bar{\boldsymbol{y}}) < 1$ for an M of finite size. Also note that both $p(\boldsymbol{s},\boldsymbol{t})$ and $p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})$ depend on $\boldsymbol{\theta}_0$. Nevertheless, we note that $p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})$ is asymptotically a pdf, since the posterior pdf $p(\boldsymbol{\theta}|\bar{\boldsymbol{y}})$ concentrates in Ω_M. In addition, since

$$p_{\boldsymbol{\Theta}}(\boldsymbol{\theta}|\bar{\boldsymbol{y}}) = \frac{p(\bar{\boldsymbol{y}}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{p(\bar{\boldsymbol{y}})} \quad \text{for} \quad \boldsymbol{\theta}\in\Omega_{\mathrm{M}}\,,$$

we have

$$p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}}) = \frac{p(\bar{\boldsymbol{y}}|\boldsymbol{s},\boldsymbol{t})p(\boldsymbol{s},\boldsymbol{t})}{p(\bar{\boldsymbol{y}})} \quad \text{for} \quad (\boldsymbol{s},\boldsymbol{t})\in\Omega_{\mathrm{M}\boldsymbol{s},\boldsymbol{t}}\,. \tag{7}$$

We use these asymptotic relations in the following derivations.
Substituting (5) and (6) into (2), we obtain

$$\begin{aligned} D_{KL}(\bar{\boldsymbol{y}}) &= \int_{T_{\boldsymbol{t}}}\int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} \log\left(\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right) p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{s}d\boldsymbol{t} + \epsilon_{\Omega_M} \\ &= \int_{T_{\boldsymbol{t}}}\int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} \log\left(\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right) p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})p(\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{s}d\boldsymbol{t} + \epsilon_{\Omega_M}\,. \end{aligned} \tag{8}$$

2.4 Laplace approximation for the conditional information gain

For a given $\boldsymbol{t}$, the posterior pdfs are expected to concentrate at $\hat{\boldsymbol{s}}$ (note that $\hat{\boldsymbol{s}}$ is the maximum likelihood estimator of the "true" parameter) as the number of observations M increases. We approximate $p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})$, $p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})$ and $p(\boldsymbol{s},\boldsymbol{t})$ by taking the exponential of the second order Taylor series expansion of their corresponding logarithms at $\hat{\boldsymbol{s}}$, for a given value of $\boldsymbol{t}$, namely by the following Gaussian posterior pdfs and the local expansion of $p(\boldsymbol{s},\boldsymbol{t})$ at $\hat{\boldsymbol{s}}$

(9)

$$\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) = \frac{1}{(\sqrt{2\pi})^r|\boldsymbol{\Sigma}_{s|t}|^{1/2}} \exp\left[-\frac{(\boldsymbol{s}-\hat{\boldsymbol{s}})^T\boldsymbol{\Sigma}_{s|t}^{-1}(\boldsymbol{s}-\hat{\boldsymbol{s}})}{2}\right],$$

(10)

$$\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}}) = p(\hat{\boldsymbol{s}},\boldsymbol{t}|\bar{\boldsymbol{y}}) \exp\left[-\frac{(\boldsymbol{s}-\hat{\boldsymbol{s}})^T\boldsymbol{\Sigma}_{s|t}^{-1}(\boldsymbol{s}-\hat{\boldsymbol{s}})}{2}\right],$$

$$(11)\quad \tilde{p}(\boldsymbol{s},\boldsymbol{t}) = p(\hat{\boldsymbol{s}},\boldsymbol{t}) \exp\left[\nabla \log p(\hat{\boldsymbol{s}},\boldsymbol{t})(\boldsymbol{s}-\hat{\boldsymbol{s}}) + \frac{(\boldsymbol{s}-\hat{\boldsymbol{s}})^T H_p(\hat{\boldsymbol{s}},\boldsymbol{t})(\boldsymbol{s}-\hat{\boldsymbol{s}})}{2}\right].$$

where $H_p(\hat{\boldsymbol{s}},\boldsymbol{t})$ is the Hessian of $h_p(\hat{\boldsymbol{s}},\boldsymbol{t})$, $h_p(\boldsymbol{s}\,,\boldsymbol{t})$ is the logarithm of prior weight function. i.e., $\log\left[p(\boldsymbol{s},\boldsymbol{t})\right]$. In order to compute $\boldsymbol{\Sigma}_{s|t}$, we carry out the second order Taylor expansion of $\tilde{F} = -\log(p(\boldsymbol{\theta}|\bar{\boldsymbol{y}}))$ at $\boldsymbol{f}(\hat{\boldsymbol{s}},\boldsymbol{t})$ as follows:

$$\tilde{F}(\boldsymbol{f}(\hat{\boldsymbol{s}},\boldsymbol{t})+\boldsymbol{U}\boldsymbol{s}) = \tilde{F}(\boldsymbol{f}(\hat{\boldsymbol{s}},\boldsymbol{t})) + \frac{(\boldsymbol{s}-\hat{\boldsymbol{s}})^T\boldsymbol{U}^T\tilde{\boldsymbol{H}}(\boldsymbol{f}(\hat{\boldsymbol{s}},\boldsymbol{t}))\boldsymbol{U}(\boldsymbol{s}-\hat{\boldsymbol{s}})}{2} + O(||\boldsymbol{s}-\hat{\boldsymbol{s}}||^3).$$

Therefore, we obtain the conditional covariance matrix after the change of variables

$$(12)\qquad \boldsymbol{\Sigma}_{s|t} = (\boldsymbol{U}^T(\tilde{\boldsymbol{H}}(\boldsymbol{f}(\hat{\boldsymbol{s}},\boldsymbol{t}))\boldsymbol{U})^{-1}.$$

Now we have the following lemma regarding to the approximation of K-L divergence:

Lemma 2. *The information gain D_{KL} can be approximated by*

(13)

$$D_{KL} = \int_{T_{\boldsymbol{t}}}\int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} \log\left(\frac{\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{\tilde{p}(\boldsymbol{s},\boldsymbol{t})}\right)\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}p(\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{t} + \epsilon_{laplace} + \epsilon_{\Omega_M},$$

where $\epsilon_{laplace} = O_P(\frac{1}{M^2})$ (its proof is given in Appendix A), $\epsilon_{\Omega_M} = O_P\left(M^{(\alpha-\frac{1}{2})p}\right)$, we assume it is smaller than $O_P(\frac{1}{M^2})$.

2.5 Laplace approximation for the expected information gain

We first introduce the following definition

Definition 7.

$$D_{\boldsymbol{s}|\boldsymbol{t}} = h(\hat{\boldsymbol{s}},\boldsymbol{t}) - h_p(\hat{\boldsymbol{s}},\boldsymbol{t}) - \frac{r}{2}\,. \tag{14}$$

where $h(\boldsymbol{s}\,,\boldsymbol{t})$ denotes the logarithm of the posterior weight function. i.e., $\log\left[p(\boldsymbol{s},\boldsymbol{t}|\{\boldsymbol{y}_i\})\right]$. Note that

$$\int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} \log\left(\frac{\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{\tilde{p}(\boldsymbol{s},\boldsymbol{t})}\right)\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s} = D_{\boldsymbol{s}|\boldsymbol{t}} + O_P\left(\frac{1}{M}\right)\,, \tag{15}$$

where the error term $O_P(\frac{1}{M})$ is dominated by $\frac{\mathbf{tr}(\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}H_p(\hat{\boldsymbol{s}},\boldsymbol{t}))}{2}$.
$\mathbf{tr}(A)$ denotes the trace of the matrix A.
The asymptotic form of the expected information gain is given below:

$$\begin{aligned} I &= \int_{\mathcal{Y}} D_{KL}\, p(\bar{\boldsymbol{y}})d\bar{\boldsymbol{y}} \\ &= \int_{\mathcal{Y}}\int_{\Omega_{M\boldsymbol{s},\boldsymbol{t}}} D_{\boldsymbol{s}|\boldsymbol{t}}p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{s}d\boldsymbol{t}p(\bar{\boldsymbol{y}})d\bar{\boldsymbol{y}} + O\left(\frac{1}{M}\right)\,, \end{aligned}$$

where we have assumed that the error term in (15) is integrable w.r.t. the data $\bar{\boldsymbol{y}}$. Furthermore, we carry out a change of parameters such that

$$\begin{aligned} I &= \int_{\mathcal{Y}}\int_{\Omega_M} D_{\boldsymbol{s}|\boldsymbol{t}}p(\boldsymbol{\theta}_0|\bar{\boldsymbol{y}})d\boldsymbol{\theta}_0p(\bar{\boldsymbol{y}})d\bar{\boldsymbol{y}} + O\left(\frac{1}{M}\right) \\ &= \int_{\mathcal{Y}}\int_{\boldsymbol{\Theta}} \mathbf{1}_{\Omega_M}D_{\boldsymbol{s}|\boldsymbol{t}}p(\boldsymbol{\theta}_0|\bar{\boldsymbol{y}})d\boldsymbol{\theta}_0p(\bar{\boldsymbol{y}})d\bar{\boldsymbol{y}} + O\left(\frac{1}{M}\right) \\ &= \int_{\boldsymbol{\Theta}}\int_{\mathcal{Y}} \mathbf{1}_{\Omega_M}D_{\boldsymbol{s}|\boldsymbol{t}}p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}_0)d\bar{\boldsymbol{y}}p(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0 + O\left(\frac{1}{M}\right)\,, \qquad (16) \end{aligned}$$

where the $\boldsymbol{t}$ in $D_{\boldsymbol{s}|\boldsymbol{t}}$ is implicitly given by $\boldsymbol{\theta}_0$.

Next, we rewrite the first two terms in $D_{\boldsymbol{s}|\boldsymbol{t}}$ using (7):

$$\begin{aligned} &\log\left[p(\hat{\boldsymbol{s}},\boldsymbol{t}|\bar{\boldsymbol{y}})\right] - \log\left[p(\hat{\boldsymbol{s}},\boldsymbol{t})\right] = \log\left[p(\bar{\boldsymbol{y}}|\hat{\boldsymbol{s}},\boldsymbol{t})\right] - \log\left[p(\bar{\boldsymbol{y}})\right] \\ &\qquad = \log\left[p(\bar{\boldsymbol{y}}|\hat{\boldsymbol{s}},\boldsymbol{t})\right] - \log\left[\int_{T_{\boldsymbol{t}}}\int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} p(\bar{\boldsymbol{y}}|\boldsymbol{s},\boldsymbol{t})p(\boldsymbol{s},\boldsymbol{t})d\boldsymbol{s}d\boldsymbol{t}\right] + O_P\left(M^{(\alpha-\frac{1}{2})p}\right)\,. \end{aligned} \tag{17}$$

Furthermore, the Laplace approximation for the above inner integration of $\boldsymbol{s}$ and the independence between $\boldsymbol{t}$ and $\bar{\boldsymbol{y}}$ given $\boldsymbol{s}$ (note that the tangent hyperplane to the manifold T is the kernel of the Jacobian of the model, i.e., $\mathbf{Ker}(\boldsymbol{J}_g)$) lead to

$$
\begin{aligned}
&-\log\left[\int_{T_{\boldsymbol{t}}}\int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} p(\bar{\boldsymbol{y}}|\boldsymbol{s},\boldsymbol{t})p(\boldsymbol{s},\boldsymbol{t})d\boldsymbol{s}d\boldsymbol{t}\right]\\
&\quad= -\log\left[\int_{T_{\boldsymbol{t}}} p(\bar{\boldsymbol{y}}|\hat{\boldsymbol{s}},\boldsymbol{t})p(\hat{\boldsymbol{s}},\boldsymbol{t})(\sqrt{2\pi})^r|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t}\right] + O_P\left(\frac{1}{M}\right)\\
&\quad= -\log\left[p(\bar{\boldsymbol{y}}|\hat{\boldsymbol{s}})\right] - \log\left[\int_{T_{\boldsymbol{t}}} p(\hat{\boldsymbol{s}},\boldsymbol{t})(\sqrt{2\pi})^r|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t}\right] + O_P\left(\frac{1}{M}\right).
\end{aligned}
$$

Substituting this expression back into (17), we can write

$$
\log\left[p(\hat{\boldsymbol{s}},\boldsymbol{t}|\bar{\boldsymbol{y}})\right] - \log\left[p(\hat{\boldsymbol{s}},\boldsymbol{t})\right] = -\log\left[\int_{T_{\boldsymbol{t}}} p(\hat{\boldsymbol{s}},\boldsymbol{t})(\sqrt{2\pi})^r|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t}\right] + O_p\left(\frac{1}{M}\right)
$$

where $p_{\boldsymbol{s}}(\hat{\boldsymbol{s}}) = \int_{\boldsymbol{T_t}} p(\hat{\boldsymbol{s}},\boldsymbol{t})d\boldsymbol{t}$, which depends on $\boldsymbol{\theta}_0$. Replacing the new expression of $D_{\boldsymbol{s}|\boldsymbol{t}}$ back into (16), we have the following theorem regarding the approximation of the expected information gain.

Theorem 1. *The expected information gain can be expressed as*

$$
I = \int_{\boldsymbol{\Theta}}\int_{\mathcal{Y}} \mathbf{1}_{\Omega_M}\left[-\log\left(\int_{T_{\boldsymbol{t}}} p(\hat{\boldsymbol{s}},\boldsymbol{t})|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t}\right) - \frac{r}{2}\log(2\pi) - \frac{r}{2}\right] p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}_0)p(\boldsymbol{\theta}_0)d\bar{\boldsymbol{y}}d\boldsymbol{\theta}_0
$$

(18)

$$
+ O\left(\frac{1}{M}\right).
$$

We can furthermore approximate the maximum posterior solution of $\boldsymbol{s}$ for a given value of $\boldsymbol{t}$, i.e., $\hat{\boldsymbol{s}}$, by $\mathbf{0}$. Theorem 1 can be simplified to the following Theorem 2.

Theorem 2. *The expected information gain can be approximated by*

(19)

$$
I = \int_{\boldsymbol{\Theta}} -\log\left(\int_{T_{\boldsymbol{t}}} p(\mathbf{0},\boldsymbol{t})|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t}\right)p(\boldsymbol{\theta}_0)d\boldsymbol{\theta}_0 - \frac{r}{2}\log(2\pi) - \frac{r}{2} + O\left(\frac{1}{M}\right).
$$

2.6 Simplification of the integration over the manifold $T_{\boldsymbol{t}}$

In (19), there still exists a double integral, due to the manifold integral in the logarithmic integrand. Specifically, the outer one is over the space of

$\boldsymbol{\theta}_0$, while the inner one is on the manifold $T_{\boldsymbol{t}}$: $\int_{T_{\boldsymbol{t}}} p(\mathbf{0},\boldsymbol{t})|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t}$. The integrand of the outer integral is a non-trivial function of the inner integral. Therefore, as a whole, this double–integral can not be viewed as a single loop with higher dimension. We now make some further simplifications of the manifold integral in (19). We first state the following lemma:

Lemma 3.

$$\int_{T_{\boldsymbol{t}}} p(\mathbf{0},\boldsymbol{t})|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}d\boldsymbol{t} = |\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2}\int_{T_{\boldsymbol{t}}} p(\mathbf{0},\boldsymbol{t})d\boldsymbol{t} + O\left(\frac{1}{M^{\frac{3}{2}}}\right),$$

where $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}$ is an approximation of $\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}$ which does not depend on $\boldsymbol{t}$.

The proof of Lemma 3 is as follows. We know (Long, 2013) that the Hessian of the negative log posterior can be expressed as

(20)

$$\tilde{\boldsymbol{H}} = \boldsymbol{H}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s + M\boldsymbol{J}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{J}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t})) - \boldsymbol{H}_p(\boldsymbol{f}(\mathbf{0},\boldsymbol{t})),$$

where $\boldsymbol{J}_g$ is the Jacobian of model $\boldsymbol{g}$ w.r.t. the parameter $\boldsymbol{\theta}$, $\boldsymbol{H}_g$ is the Hessian of model $\boldsymbol{g}$ w.r.t. the parameter $\boldsymbol{\theta}$, and $\boldsymbol{E}_s$ denotes the sum of the data residuals, i.e., $\boldsymbol{E}_s = \sum_{i=1}^{M}\boldsymbol{r}_i$.
Substituting (20) into (12) leads to

$$\begin{aligned}\boldsymbol{\Sigma}_{s|t} =& \left\{\boldsymbol{U}^T\left[M\boldsymbol{J}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{J}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))\right]\boldsymbol{U}\right\}^{-1} + O\left(\frac{1}{M^{\frac{3}{2}}}\right)\\ =&\tilde{\boldsymbol{\Sigma}}_{s|t} + O\left(\frac{1}{M^{\frac{3}{2}}}\right). \qquad (21)\end{aligned}$$

The order of the error term can be derived using the Woodbury matrix identity (Hager,1989). Let us consider the following eigendecomposition of

$$\boldsymbol{U}^T\left[\boldsymbol{H}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s - \nabla\nabla h_p(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))\right]\boldsymbol{U} = \boldsymbol{RCL},$$

where $\boldsymbol{R}$ is the column eigenvector matrix, $\boldsymbol{L}$ is the row eigenvector matrix, and $\boldsymbol{C}$ is the diagonal matrix containing eigenvalues. Let

$$\boldsymbol{A} = \boldsymbol{U}^T\left[M\boldsymbol{J}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{J}_g(\boldsymbol{f}(\mathbf{0},\boldsymbol{t}))\right]\boldsymbol{U}.$$

Thus, the conditional covariance matrix $\boldsymbol{\Sigma}_{s|t}$ can be rewritten using Woodbury matrix identity as follows

$$\boldsymbol{\Sigma}_{s|t} = (\boldsymbol{A}+\boldsymbol{RCL})^{-1} = \boldsymbol{A}^{-1} - \boldsymbol{A}^{-1}\boldsymbol{L}(\boldsymbol{C}^{-1}+\boldsymbol{RA}^{-1}\boldsymbol{L})^{-1}\boldsymbol{RA}^{-1}. \qquad (22)$$

Since the order of $\boldsymbol{C}$ is $O(\sqrt{M})$ and the order of $\boldsymbol{A}$ is $O(M)$, the error term $\boldsymbol{A}^{-1}\boldsymbol{L}(\boldsymbol{C}^{-1}+\boldsymbol{RA}^{-1}\boldsymbol{L})^{-1}\boldsymbol{RA}^{-1}$ is $O\left(\frac{1}{M^{\frac{3}{2}}}\right)$.

By construction, the model $\boldsymbol{g}$ is only a function of $\boldsymbol{s}$, therefore, the simplification $\tilde{\boldsymbol{\Sigma}}_{s|t}$ does not depend on $\boldsymbol{t}$. Now we complete the proof of Lemma 3.

Let $p_{\boldsymbol{s}}(\mathbf{0}) = \int_{T_{\boldsymbol{t}}} p(\mathbf{0}, \boldsymbol{t}) d\boldsymbol{t}$ be the marginal of prior pdf of parameter $\boldsymbol{s}$.

In addition, it is in general difficult to compute the marginal pdf $p_{\boldsymbol{s}}(\mathbf{0})$. Trying to introduce an error as small as possible, we first linearize the manifold locally at $\boldsymbol{\theta}^*(\boldsymbol{\theta}_0)$ in the case of a single modal prior or at the modes of the set $\{\boldsymbol{\theta}^*(\boldsymbol{\theta}_0)\}$, which contains all the local optima in the case of a multimodal prior:

$$\boldsymbol{\theta}^*(\boldsymbol{\theta}_0) := \arg\max_{\boldsymbol{\theta}\in T} \{p(\boldsymbol{\theta})\} \ .$$

A linear transformation of variables of the prior leads to the approximated marginal $\tilde{p}_{\boldsymbol{s}}(\mathbf{0})$. Note that by linearizing the manifold at the maximum point/points, we minimizes the $O(1)$ error here. For instance, we can obtain $\tilde{p}_{\boldsymbol{s}}(\mathbf{0})$ easily for Gaussian or Gaussian mixture prior.

By carrying out both simplifications, we introduce an error of order $O(1)$. However, our numerical example shows that it is still a negligible error. Thus, we have a further simplified estimation of the expected information gain:

Theorem 3.

(23)
$$I = \int_{\boldsymbol{\Theta}} \left[-\log\left[\tilde{p}_{\boldsymbol{s}}(\mathbf{0})\right] - \frac{1}{2}\log|\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}| - \frac{r}{2}\log(2\pi) - \frac{r}{2} \right] p(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 + O\left(1\right) \ .$$

with $\tilde{\boldsymbol{\Sigma}}_{\boldsymbol{s}|\boldsymbol{t}}$ *in* (21) *and* $\tilde{p}_{\boldsymbol{s}}(\mathbf{0})$ *computed by linearizing the manifold.*

3 Estimation of the expected information gain for a quantity of interest

In (Long, 2013) we considered the approximation of the posterior pdf of a quantity of interest through a Gaussian distribution, in the case of a model whose parameters can be completely determined by the data. We now focus on the prediction of a physical quantity of interest for an underdetermined model. The quantity of interest is commonly defined as a function of $\boldsymbol{\theta}$ plus some independent error, i.e.,

$$Q = \tau(\boldsymbol{\theta}) + \epsilon_Q \, ,$$

The prediction error ϵ_Q is assumed to be independent of $\boldsymbol{\theta}$. We can reparameterize τ using $(\boldsymbol{s},\, \boldsymbol{t})$ defined in Section 2.3. i.e.,

$$Q = \tau(\ f(\boldsymbol{s},\, \boldsymbol{t})) + \epsilon_Q = \hat{\tau}(\boldsymbol{s},\, \boldsymbol{t}) + \epsilon_Q \quad \text{for}\, (\boldsymbol{s}, \boldsymbol{t}) \in \Omega_{M\boldsymbol{s},\boldsymbol{t}} \, .$$

Given $\boldsymbol{t}$, a Taylor expansion of τ at $(\hat{\boldsymbol{s}},\, \boldsymbol{t})$ along $\boldsymbol{s}$ leads to

$$\hat{\tau}(\boldsymbol{s},\, \boldsymbol{t}) = \hat{\tau}(\hat{\boldsymbol{s}},\, \boldsymbol{t}) + (\nabla_{\boldsymbol{s}}\hat{\tau}(\hat{\boldsymbol{s}},\, \boldsymbol{t}))(\boldsymbol{s} - \hat{\boldsymbol{s}}) + O_P(||\boldsymbol{s} - \hat{\boldsymbol{s}}||^2)\,.$$

Since the conditional posterior pdf $p(\boldsymbol{s}|\boldsymbol{t}, \bar{\boldsymbol{y}})$ can be approximated by a Gaussian distribution concentrated around $(\hat{\boldsymbol{s}},\, \boldsymbol{t})$ as discussed in the previous section, we can apply a small noise approximation to propagate randomness from $\boldsymbol{s}$ to the quantity of interest Q.
The resulting approximated distribution of Q, given $\boldsymbol{t}$ and $\bar{\boldsymbol{y}}$, is also Gaussian:

(24)
$$p(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}}) = \frac{1}{\sqrt{2\pi}\sigma_{Q|\boldsymbol{t},\,\bar{\boldsymbol{y}}}} \exp\left[-\frac{(Q - \hat{\tau}(\hat{\boldsymbol{s}},\, \boldsymbol{t}))^2}{2\sigma^2_{Q|\boldsymbol{t},\,\bar{\boldsymbol{y}}}}\right] + O_P\left(\frac{1}{M^2}\right) = \hat{p}(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}}) + O_P\left(\frac{1}{M^2}\right)\,,$$

where

$$\sigma^2_{Q|\boldsymbol{t},\,\bar{\boldsymbol{y}}} = (\nabla_{\boldsymbol{s}}\hat{\tau})^T \boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}} \nabla_{\boldsymbol{s}}\hat{\tau} + \sigma^2_{\epsilon_Q}\,.$$

Here, $\sigma^2_{\epsilon_Q}$ denotes the variance of ϵ_Q, which is assumed to be a known constant.
The approximated posterior pdf of Q is then

(25)
$$p(Q|\{\boldsymbol{y}_i\}) = \int_{T_{\boldsymbol{t}}} \hat{p}(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}}) p(\boldsymbol{t}|\bar{\boldsymbol{y}}) d\boldsymbol{t} + O_P\left(\frac{1}{M^2}\right) = \hat{p}(Q|\{\boldsymbol{y}_i\}) + O_P\left(\frac{1}{M^2}\right)\,.$$

We can furthermore write $p(\boldsymbol{t}|\bar{\boldsymbol{y}})$ as a marginal over $\boldsymbol{s}$:

$$p(\boldsymbol{t}|\bar{\boldsymbol{y}}) = \int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}}) d\boldsymbol{s} + O_P\left(M^{(\alpha-\frac{1}{2})p}\right)\,.$$

Note that $p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})$ should concentrate at $\boldsymbol{s} = \hat{\boldsymbol{s}}$ in Ω_M as the amount of data increases, which leads to the following Laplace approximation of $p(\boldsymbol{t}|\bar{\boldsymbol{y}})$:

$$p(\boldsymbol{t}|\bar{\boldsymbol{y}}) = p(\hat{\boldsymbol{s}},\boldsymbol{t}|\bar{\boldsymbol{y}})(\sqrt{2\pi})^r |\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2} + O_P\left(\frac{1}{M}\right) = \hat{p}(\boldsymbol{t}|\bar{\boldsymbol{y}}) + O_P\left(\frac{1}{M}\right)\,. \tag{26}$$

Substituting (26) back into (25), we obtain

$$\begin{aligned} p(Q|\{\boldsymbol{y}_i\}) =& \int_{T_{\boldsymbol{t}}} \hat{p}(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}}) p(\hat{\boldsymbol{s}},\boldsymbol{t}|\bar{\boldsymbol{y}})(\sqrt{2\pi})^r |\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2} d\boldsymbol{t} + O_P\left(\frac{1}{M}\right) \\ =& \hat{p}(Q|\{\boldsymbol{y}_i\}) + O_P\left(\frac{1}{M}\right)\,. \end{aligned} \tag{27}$$

Using the Bayes theorem we derive $\hat{p}(Q|\{\boldsymbol{y}_i\})$ up to a scaling factor:

$$\begin{aligned}\hat{p}(Q|\{\boldsymbol{y}_i\}) &\propto \int_{T_{\boldsymbol{t}}} \hat{p}(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}})(\sqrt{2\pi})^r |\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2} p(\bar{\boldsymbol{y}}|\hat{\boldsymbol{s}},\boldsymbol{t}) p(\hat{\boldsymbol{s}},\boldsymbol{t}) d\boldsymbol{t} \\ &\propto \int_{T_{\boldsymbol{t}}} \hat{p}(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}}) |\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2} p(\hat{\boldsymbol{s}},\boldsymbol{t}) d\boldsymbol{t} \\ &= \int_{T_{\boldsymbol{t}}} \int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} \mathbf{1}_{\hat{\boldsymbol{s}}} \hat{p}(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}}) |\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2} p(\boldsymbol{s},\boldsymbol{t}) d\boldsymbol{s} d\boldsymbol{t}\,, \qquad (28)\end{aligned}$$

where $\mathbf{1}_{\hat{\boldsymbol{s}}}$ denotes an indicator function, which takes value 1 when $\boldsymbol{s} = \hat{\boldsymbol{s}}$, otherwise, it takes value 0.
Since Q is a scalar, we can obtain the scaling factor for $\hat{p}(Q|\{\boldsymbol{y}_i\})$ once (28) is computed using a one dimensional grid of Q whose range is defined in Section 4. We can next compute the expected conditional entropy by

(29)

$$H(Q|\bar{\boldsymbol{y}}) = \int_{\boldsymbol{\Theta}} \int_{\mathcal{Y}} \int_{Q} \log\left[\hat{p}(Q|\bar{\boldsymbol{y}})\right] \hat{p}(Q|\bar{\boldsymbol{y}}) dQ p(\bar{\boldsymbol{y}}|\boldsymbol{\theta}_0) d\bar{\boldsymbol{y}} p(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 + O\left(\frac{1}{M}\right)$$

If we approximate $\hat{\boldsymbol{s}}$ using $\mathbf{0}$, we have the following approximated pdfs:

$$\hat{p}(Q|\boldsymbol{t},\boldsymbol{\theta}_0) = \frac{1}{\sqrt{2\pi}\sigma_{Q|\boldsymbol{t},\boldsymbol{\theta}_0}} \exp\left[-\frac{(Q-\hat{\tau}(\mathbf{0},\boldsymbol{t}))^2}{2\sigma^2_{Q|\boldsymbol{t},\boldsymbol{\theta}_0}}\right] = \hat{p}(Q|\boldsymbol{t},\, \bar{\boldsymbol{y}}) + O_P\left(\frac{1}{\sqrt{M}}\right)\,,$$

with

$$\sigma_{Q|\boldsymbol{t},\boldsymbol{\theta}_0} = (\nabla_{\boldsymbol{s}}\hat{\tau})^T \boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}(\mathbf{0}) \nabla_{\boldsymbol{s}}\hat{\tau} + \sigma^2_{\epsilon_Q}\,.$$

and

$$\hat{p}(Q|\;\boldsymbol{\theta}_0) \propto \int_{T_{\boldsymbol{t}}} \int_{[-O(M^{-\alpha}),O(M^{-\alpha})]} \mathbf{1}_{\mathbf{0}} \hat{p}(Q|\boldsymbol{t},\, \boldsymbol{\theta}_0) |\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|^{1/2} p(\boldsymbol{s},\boldsymbol{t}) d\boldsymbol{s} d\boldsymbol{t}\,. \qquad (30)$$

Replacing $\hat{p}(Q|\bar{\boldsymbol{y}})$ in (29) by $\hat{p}(Q|\boldsymbol{\theta}_0)$ and integrating out the $O_P\left(\frac{1}{\sqrt{M}}\right)$ term using $\mathcal{Y}$ (this term has mean zero; see Appendix B for details), we obtain the following asymptotic results for the expected conditional entropy:

$$H(Q|\bar{\boldsymbol{y}}) = \int_{\boldsymbol{\Theta}} \int_{Q} \log\left[\hat{p}(Q|\boldsymbol{\theta}_0)\right] \hat{p}(Q|\boldsymbol{\theta}_0) dQ p(\boldsymbol{\theta}_0) d\boldsymbol{\theta}_0 + O\left(\frac{1}{M}\right)\,, \qquad (31)$$

The rates of the dominant errors are derived in Appendix C. It is observed that we have a double loop integral for the computation of $H(Q|\bar{\boldsymbol{y}})$. We show the details of the numerical computation in the next section. Eventually the expected information gain for the quantity of interest Q can be approximated by

$$I = H(Q) - H(Q|\bar{\boldsymbol{y}})\,,$$

where $H(Q) = -\int_Q \log\left[p(Q)\right] p(Q)dQ$ is the prior entropy of Q. Observe that we do not need to compute $H(Q)$ in order to select the best experimental set–up, since $H(Q)$ does not depend on the set–up parameter $\boldsymbol{\xi}$. A more detailed description about $H(Q)$ and its computation can be found in (Long, 2013). Thus, we will focus on the numerical computation of (31) in Section 4.

4 Numerical integration of the asymptotic forms

In most practical scenarios, we need to compute numerically the expected information gain (18) for parameter inferences and the expected conditional entropy (29) for predictions of quantities of interest.

We can approximate the integral in (18) using numerical quadratures or Monte Carlo sampling depending on the regularity of the integrand. In the case of using quadratures, the numerical integration adopts the following form:

$$(32) \qquad I_Q = \sum_{i=1}^{NQ} \left[-\log\left[p_{\boldsymbol{s}}(\boldsymbol{0})\right] - \frac{1}{2}\log\left(|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|\right) \right] w_i - \frac{r}{2} - \frac{r}{2}\log\left(2\pi\right) \,,$$

where NQ is the number of quadrature points, w_i is the weight for the i^{th} quadrature points. $p_{\boldsymbol{s}}(\boldsymbol{0})$ and $\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}$ are computed using the i^{th} quadrature of the "true" parameter $\boldsymbol{\theta}_{0i}$. We can adopt sparse quadrature abscissas and weights for high dimensional parameter $\boldsymbol{\theta}_0$ and a corresponding integrand with certain regularity. A review of sparse grids can be found, for instance, in (Barthelmann, 2000, Nobile, 2008, Long, 2013, Garcke, 2013). On the other hand, if there is a lack of regularity in the marginal prior $p_{\boldsymbol{s}}(\boldsymbol{0})$ or the conditional covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}$, we can use Monte Carlo sampling:

$$(33) \qquad I_{MC} = \frac{1}{NS}\sum_{j=1}^{NS} \left[-\log\left[p_{\boldsymbol{s}}(\boldsymbol{0})\right] - \frac{1}{2}\log\left(|\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}|\right) \right] - \frac{r}{2} - \frac{r}{2}\log(2\pi) \,,$$

where NS denotes the number of samples. Both $p_{\boldsymbol{s}}(\boldsymbol{0})$ and $\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}}$ are computed using the j^{th} random sample of the "true" parameter $\boldsymbol{\theta}_{0j}$.

Due to the discontinuous indicator function in the integral of $\hat{p}(Q|\boldsymbol{\theta}_0)$ (30), we use Monte Carlo sampling for its numerical estimation. For the sake of convenience, we also adopt Monte Carlo sampling for the numerical estimation of the expected conditional entropy of the quantity of interest (31), such that we can reuse the same set of samples in both integrals. Specifically, we use the sample average w.r.t. the $\boldsymbol{\theta}$ prior and one dimensional binning for

Q as follows

$$H(Q|\bar{\boldsymbol{y}})_{MC} = \frac{1}{NS}\sum_{j=1}^{NS}\sum_{k=1}^{NS1} \log\left[\hat{p}(Q_k|\boldsymbol{\theta}_{0j})\right]\hat{p}(Q_k|\boldsymbol{\theta}_{0j})\Delta_k \tag{34}$$

with

$$\hat{p}(Q_k|\boldsymbol{\theta}_{0j}) \propto \frac{1}{NS}\sum_{l=1}^{NS} \mathbf{1}_\Omega \hat{p}(Q_k|\boldsymbol{t}_l,\, \boldsymbol{\theta}_{0j})\,, \tag{35}$$

where $NS1$ is the number of points in a one dimensional mesh partitioning the domain of scalar Q, and Δ_k is the length of the k^{th} segment.
We define the domain of Q as $[\min\tau(\boldsymbol{\theta}), \max\tau(\boldsymbol{\theta})]$, where $\boldsymbol{\theta}$ takes a value from the NS samples. We use the same collection of samples for $\boldsymbol{\theta}$ in both (34) and (35). $\boldsymbol{t}_l$ is the local coordinate corresponding to the l^{th} sample of $\boldsymbol{\theta}$. Ideally, this sample is supposed to be on the manifold T. In practice, we approximately consider all the samples in Ω as ones on the manifold T. Ω, which depends on $\boldsymbol{\theta}_0$, is defined by

$$\Omega(\boldsymbol{\theta}_0) = \{\boldsymbol{\theta} \in \boldsymbol{R}^d \,:\, ||g(\boldsymbol{\theta}) - g(\boldsymbol{\theta}_0)|| \leq C\}\,. \tag{36}$$

where C is a small constant. We used 10^{-3} in our computations.

5 Numerical Examples

5.1 Model with two indistinguishable parameters

We apply our projective Laplace method to the second example in (Long, 2013). The measurement y reads

$$y = (\alpha\theta_1 + \beta\theta_2)^3\xi^2 + (\alpha\theta_1 + \beta\theta_2)\exp[-|0.2 - \xi|] + \epsilon\,.$$

It is a single output experiment with a model of two parameters. The model is not sensitive to a change in the parameters along the direction of $(\beta,\, -\alpha)$, where α and β are two given constants. The measurement noise is assumed to be Gaussian, i.e., $\epsilon \sim \mathcal{N}(0, \sigma_m^2)$. We firstly assume a uniform prior for the parameters $\boldsymbol{\theta} = [\theta_1,\, \theta_2]^T$, i.e.,

$$\boldsymbol{\theta} \sim \mathcal{U}(\boldsymbol{\theta}_l,\, \boldsymbol{\theta}_u)\,, \quad \text{with} \quad \boldsymbol{\theta}_l = [0,0]^T \quad \text{and} \quad \boldsymbol{\theta}_u = [1,1]^T\,.$$

Note that the method in (Long, 2013) is not applicable here since there is no single mode in the posterior due to the non–informative prior. The Jacobian of this model with respect to $\boldsymbol{\theta}$ is

$$\boldsymbol{J} = [3\alpha(\alpha\theta_1{+}\beta\theta_2)^2\xi^2{+}\alpha\exp(-|0.2{-}\xi|), \quad 3\beta(\alpha\theta_1{+}\beta\theta_2)^2\xi^2{+}\beta\exp(-|0.2{-}\xi|)]\,.$$

Consider the particular case of $\alpha = 1$ and $\beta = 1$. We know that the linear manifold is defined by $\boldsymbol{V} = \left[\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right]^T$, and its orthogonal direction is $\boldsymbol{U} = \left[\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right]^T$. According to our theory, the following transformation is carried out:

$$\begin{aligned} s &= \frac{\sqrt{2}}{2}(\theta_1 - \theta_{10}) + \frac{\sqrt{2}}{2}(\theta_2 - \theta_{20})\,, \\ t &= \frac{\sqrt{2}}{2}(\theta_1 - \theta_{10}) - \frac{\sqrt{2}}{2}(\theta_2 - \theta_{20})\,. \end{aligned}$$

We can easily obtain $p_s(0)$ as

$$p_s(0) = \begin{cases} \sqrt{2}(\theta_{10} + \theta_{20})\,, & 0 < \theta_{10} + \theta_{20} \leq 1\,, \\ \sqrt{2}(2 - \theta_{10} - \theta_{20})\,, & 1 < \theta_{10} + \theta_{20} \leq 2\,. \end{cases}$$

In Figure 2, we compare the information gains computed using our projective Laplace approximation with sparse grid (LA + SG) numerical integration, Monte Carlo (LA + MC) sampling, or double–loop Monte Carlo (DLMC) in the scenario when $M = 10$ and $\xi = 0.3$. The projective Laplace approximations have no bias and exhibit faster convergence compared to the DLMC. The DLMC requires at least 10^2 times number of samples to reach the same precision as the projective Laplace approximation. Note that the likelihood evaluation, in this particular case, dominates the CPU time of the DLMC. Therefore, the CPU time spent on the estimation is actually proportional to the square of number of samples, when the number of samples in the outer loop equals to the number of samples in the inner loop, in the DLMC of our case. In this sense, our method is at least 10^4 times faster than the DLMC. In Figure 3, we compare the absolute relative error of the expected information gain for the sparse quadrature and Monte Carlo sampling when $\xi = 0.3$. The value of the expected information gain, 3.87, computed using 10^7 number of samples in the Monte Carlo sampling is taken as the reference, in order to compute the absolute relative error in Figure 3. Due to the lack of smoothness of the integrand function, the convergence rate of LA + SG is similar to LA + MC. A snapshot of the surface of marginal p_s is visualized in Figure 4. Note that there is a “kink” along the diagonal connecting $(1, 0)$ and $(1, 0)$. This makes numerical integration difficult in principle.

5.1.1 Mixture Gaussian prior

To further evaluate the robustness of (18), we set the prior as a mixture Gaussian which adopts the following form:

$$p(\boldsymbol{\theta}) = 0.5 \times p_1(\boldsymbol{\theta}) + 0.5 \times p_2(\boldsymbol{\theta}) \tag{37}$$

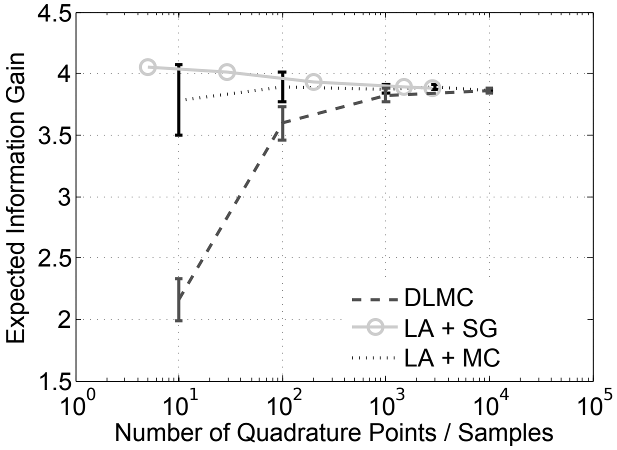


Figure 2: The convergence of the expected information gain computed using the LA + MC or LA + SG, or using a DLMC in Example 5.1. A uniform prior was used. The same set of samples was used for both the inner and outer loops in the DLMC. The number of samples of DLMC is associated to this set of samples.

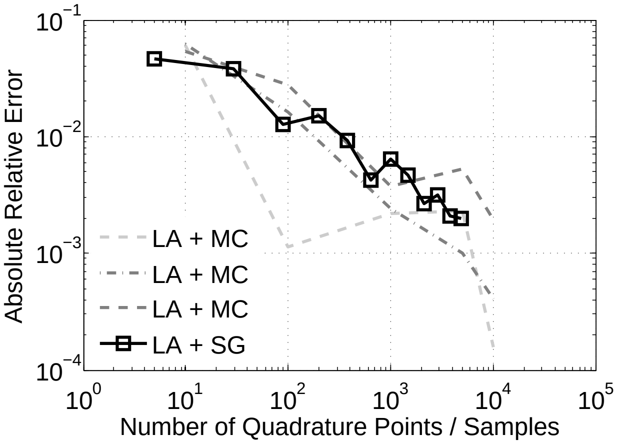


Figure 3: The absolute relative error of the LA + MC or LA + SG in Example 5.1. The three curves of LA + MC represent three independent runs of this method.

where $p_1(\boldsymbol{\theta})$ and $p_2(\boldsymbol{\theta})$ are the pdfs of two multivariate Gaussian with mean vectors $[2, 0]^T$ and $[0, 2]^T$, respectively, and covariance matrix

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

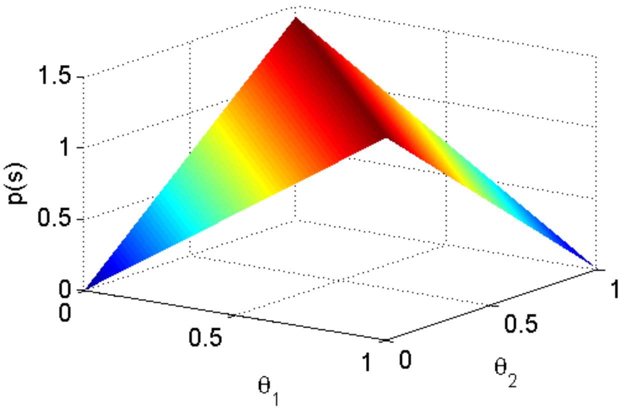


Figure 4: The tent function $p(\boldsymbol{s})$ in Example 5.1.

The salient features of this prior are the two separated modes; see Figures 5(a) and 5(b) for the visualization. We compare the results obtained using LA + SG and LA + MC, and DLMC. Figure 6(a) shows the performances of the three methods in terms of the number of quadrature points (LA + SG) or sample points (LA + MC and DLMC), when $\xi = 0.3$. Our approximation integrated by either sparse grid or Monte Carlo is significantly faster than the DLMC. Similarly, Figure 6(b) shows the convergeces of these methods when $\xi = 1$. We used an auxiliary Gaussian pdf, with mean vector $[2, 0]^T$ and covariance matrix

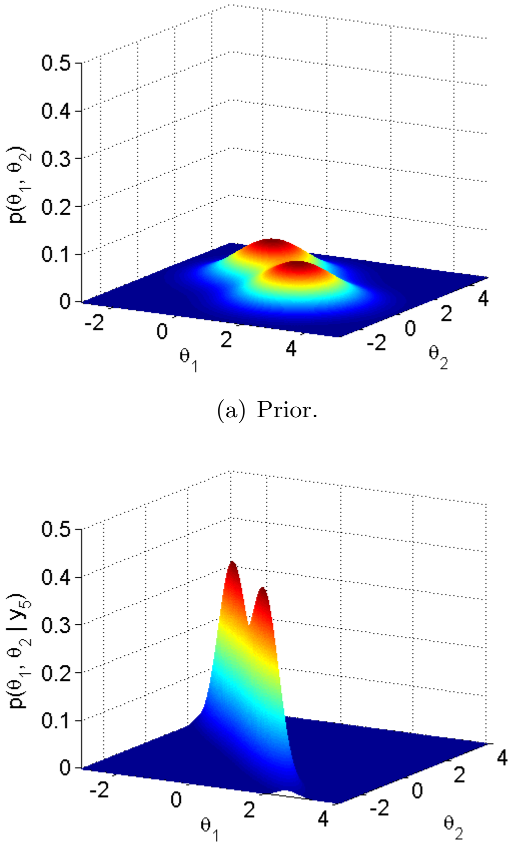
$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

in the sparse grid numerical integration.

5.1.2 Mixture Log Gaussian prior

We define a new set of parameters $\boldsymbol{\gamma} = \log \boldsymbol{\theta}$. We assume $\boldsymbol{\gamma}$ is the same mixture of two Gaussian pdfs as in 5.1.1. The indistinguishable manifold of $\boldsymbol{\gamma}$ is not a straight line anymore (see Figure 7 for the visualization of two posterior pdfs.), and needs to be computed implicitly by the eigenvalue decomposition of the Hessian matrix (4).

As seen in Figure 8, the convergence of DLMC is substantially slower than our approaches in this case (at least 10^5 times slower in terms of number of samples) due to the change of variable. Additionally, we split the integral (18) against the mixture Gaussian into two integrals with two Gaussian pdfs separately, so that an auxiliary measure is not needed. Same numbers of quadrature points are used in both integrations. We note that the splitted integration reaches a high precision when the total number of quadratures



(a) Prior.

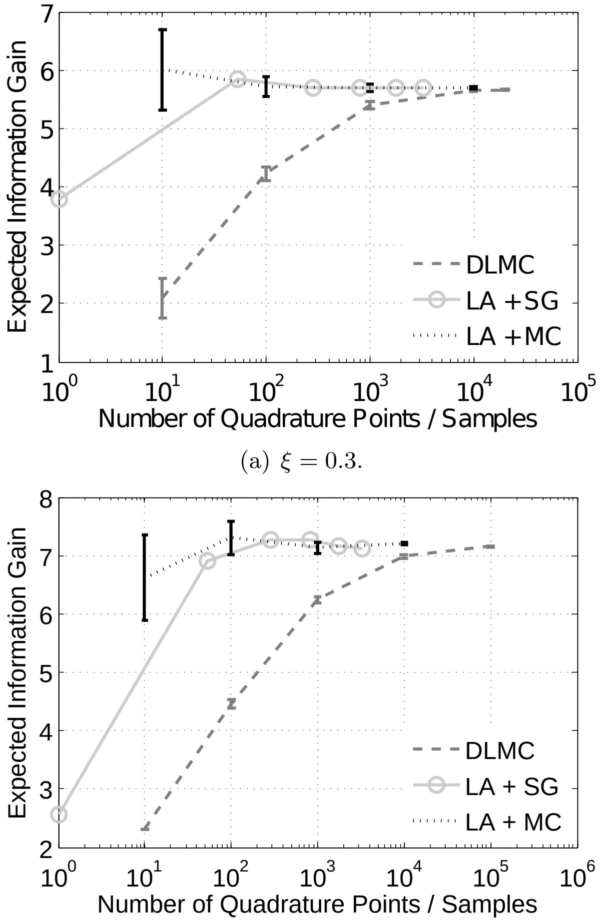
(b) Posterior. M=5.

Figure 5: The prior and posterior pdfs of mixture Gaussian with two separated modes in Example 5.1.

is smaller than 10 (see Figure 9). The convergence of both approaches in terms of the absolute consecutive difference are shown in Figure 10.

6 Conclusion

In this work, we have extended the Bayesian experimental design methodology based on the Laplace approximation from determined cases to underdetermined cases. Instead of carrying out the Laplace approximation at the single well–defined posterior mode, we conduct the Laplace approximation in the orthogonal directions of the unidentifiable manifold for the conditional posterior weight functions, under a local transformation of parameters. Eventually, the expected information gain can be approximated

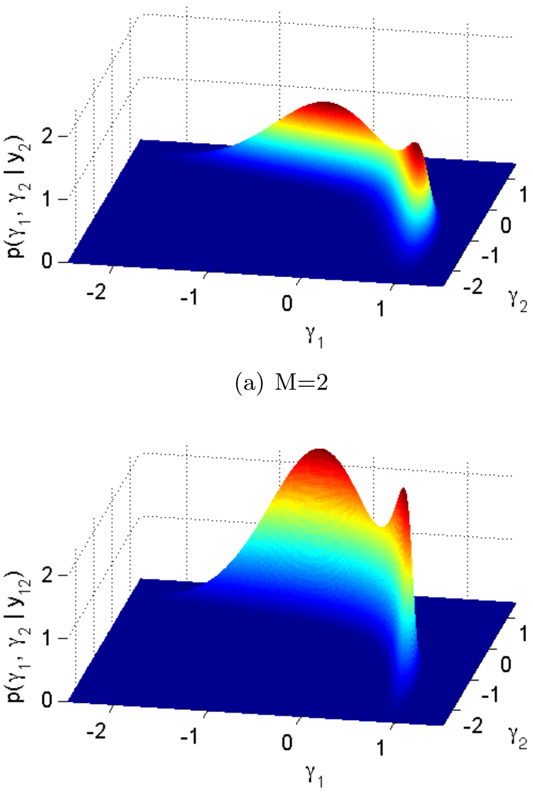


(a) $\xi = 0.3$.

(b) $\xi = 1$.

Figure 6: The convergence of the expected information gain computed using the LA + MC or LA + SG, or using DLMC in Example 5.1. A mixture Gaussian prior was used.

asymptotically as an integration over the prior parameter domain similar to the cases where the model parameters are determined completely by the experiment. One extra step is to project the Hessian onto the orthonormal complement space of the kernel of the cost function Jacobian. Furthermore, we approximate the marginal pdf of the parameter orthogonal to the indistinguishable manifold, using a linearized manifold at the modes found by a constrained optimization. By doing this, we have a dominant error $O(1)$. We also developed the techniques for the prediction of quantities of interest based on the same strategy. The proposed formula is able to deal with a model of unidentifiable manifold of parameters and multimodal or noninfor-



(a) M=2

(b) M=12

Figure 7: The posterior pdfs of $\boldsymbol{\gamma}$ with two separated modes in Example 5.1.

mative (uniform) priors. In order to carry out the numerical integration, we can use sparse quadrature techniques or Monte Carlo sampling depending on the regularity of the integrand function. We have demonstrated the efficiency and accuracy of our method using numerical examples that include the design of the scalar experimental set–up in an one dimensional cubic polynomial function with two indistinguishable parameters.

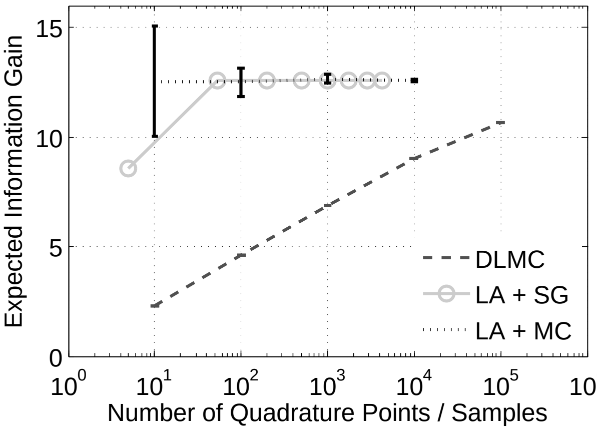


Figure 8: The convergence of the expected information gain computed using the LA + MC or LA + SG, or using DLMC. $\xi = 1$ in Example 5.1. A mixture log Gaussian prior was used.

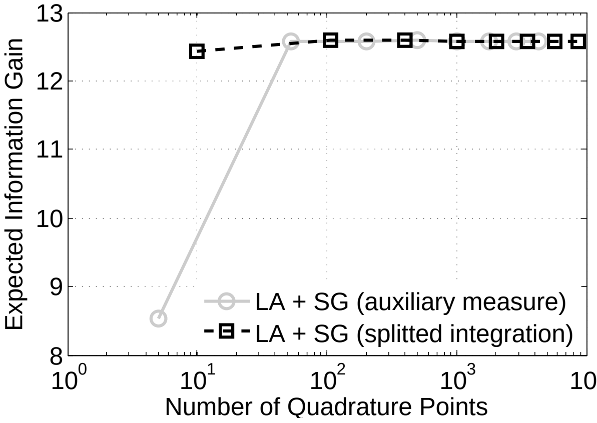


Figure 9: The convergence of expected information gain computed by LA + SG in Example 5.1. $\xi = 1$. A mixture log Gaussian prior was used.

A Proof of the error estimate in Equation (13).

The Laplace approximation error $\epsilon_{laplace}$ in Equation (13) can be expressed as follows:

$$\int_{\boldsymbol{T}}\int_{\boldsymbol{S}} \log\left[\frac{\tilde{p}(\boldsymbol{s},\boldsymbol{t})}{p(\boldsymbol{s},\boldsymbol{t})}\right]\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}d\boldsymbol{t} + \int_{\boldsymbol{T}}\int_{\boldsymbol{S}} \log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}\right]\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}d\boldsymbol{t}$$
$$+\int_{\boldsymbol{T}}\int_{\boldsymbol{S}} \log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right]\left[p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) - \tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})\right]d\boldsymbol{s}d\boldsymbol{t} = E_1 + E_2 + E_3\,.$$

To prove the error order for E_1, we consider the inner integration of $\boldsymbol{s}$ for a fixed value of $\boldsymbol{t}$. We write the Taylor series of $h_p(\boldsymbol{s},\boldsymbol{t})$, defined in Section

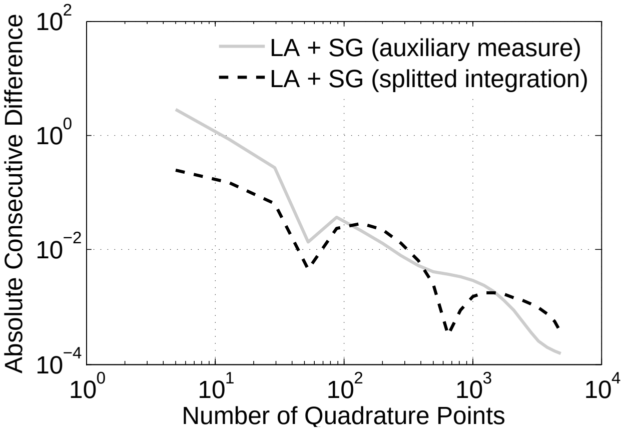


Figure 10: The absolute consecutive difference of expected information gain computed by LA + SG in Example 5.1. $\xi = 1$. A mixture log Gaussian prior was used.

2.4, at the vicinity of $\hat{\boldsymbol{s}}$ as follows

$$h_p(\boldsymbol{s},\boldsymbol{t}) = \sum_{|\boldsymbol{\alpha}|\leq 4} \frac{D^{\boldsymbol{\alpha}} h_p(\hat{\boldsymbol{s}},\boldsymbol{t})}{\boldsymbol{\alpha}!}(\boldsymbol{s}-\hat{\boldsymbol{s}})^{\boldsymbol{\alpha}} + O_P\left(\|\boldsymbol{s}-\hat{\boldsymbol{s}}\|^5\right) ,$$

where we use the multi–index notation $\boldsymbol{\alpha}$ with the following properties:

$$|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d \, , \, \alpha! = \alpha_1! \cdots \alpha_d! \, , \, (\boldsymbol{s})^{\boldsymbol{\alpha}} = s_1^{\alpha_1} \cdots s_d^{\alpha_d} \, .$$

The odd central moments of the multivariate Gaussian are zero and the parameter posterior covariance $\boldsymbol{\Sigma}$ is of $O_P\left(\frac{1}{M}\right)$. It is straightforward to see that the fourth and sixth moments of this multivariate Gaussian are $O_P\left(\frac{1}{M^2}\right)$ and $O_P\left(\frac{1}{M^3}\right)$, respectively. Consequently the conditional expectation of $h_p(\boldsymbol{s},\boldsymbol{t})$ is

$$\begin{aligned}\int_{\boldsymbol{S}} h_p(\boldsymbol{s},\boldsymbol{t})\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s} =& h_p(\hat{\boldsymbol{s}},\boldsymbol{t}) + \frac{\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}} : \nabla\nabla h_p(\hat{\boldsymbol{s}},\boldsymbol{t})}{2} \\ &+ \frac{1}{4!}\sum_{i,j,k,l}(\partial_{ijkl}h_p)(\Sigma_{\boldsymbol{s}|\boldsymbol{t},ij}\Sigma_{\boldsymbol{s}|\boldsymbol{t},kl} + \Sigma_{\boldsymbol{s}|\boldsymbol{t},ik}\Sigma_{\boldsymbol{s}|\boldsymbol{t},jl} + \Sigma_{\boldsymbol{s}|\boldsymbol{t},il}\Sigma_{\boldsymbol{s}|\boldsymbol{t},jk}) + O_P\left(\frac{1}{M^3}\right) \\ =& h_p(\hat{\boldsymbol{s}},\boldsymbol{t}) + \frac{\boldsymbol{\Sigma}_{\boldsymbol{s}|\boldsymbol{t}} : \nabla\nabla h_p(\hat{\boldsymbol{s}},\boldsymbol{t})}{2} + O_P\left(\frac{1}{M^2}\right) \\ =& \int_{\boldsymbol{S}} \log(\tilde{p}(\boldsymbol{s},\boldsymbol{t}))\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s} + O_P\left(\frac{1}{M^2}\right) ,\end{aligned}$$

with i, j, k, l = 1,..., dim($\boldsymbol{s}$) and $\partial_{ijkl}h = \frac{\partial^4 h(\hat{\boldsymbol{s}},\boldsymbol{t})}{\partial s_i \partial s_j \partial s_k \partial s_l}$. Therefore, $E_1 = O_p(\frac{1}{M^2})$.

Next, regarding E_2, we observe that

$$\log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}\right] = \sum_{|\boldsymbol{\alpha}|=3,4} \frac{D^{\boldsymbol{\alpha}} h(\hat{\boldsymbol{s}},\boldsymbol{t})}{\boldsymbol{\alpha}!}(\boldsymbol{s}-\hat{\boldsymbol{s}})^{\boldsymbol{\alpha}} + O_P\left(\|\boldsymbol{s}-\hat{\boldsymbol{s}}\|^5\right) .$$

Similar to the analysis of the expectation of $h_p(\boldsymbol{s},\boldsymbol{t})$ above, the expectation of this log ratio is

$$\int_{\boldsymbol{S}} \log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{\tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}\right] \tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) d\boldsymbol{s} = \frac{1}{4!}\sum_{i,j,k,l} (\partial_{ijkl} h_p)\left(\Sigma_{\boldsymbol{s}|\boldsymbol{t},ij}\Sigma_{\boldsymbol{s}|\boldsymbol{t},kl} + \Sigma_{\boldsymbol{s}|\boldsymbol{t},ik}\Sigma_{\boldsymbol{s}|\boldsymbol{t},jl} + \Sigma_{\boldsymbol{s}|\boldsymbol{t},il}\Sigma_{\boldsymbol{s}|\boldsymbol{t},jk}\right) + O_P\left(\frac{1}{M^3}\right) = O_P\left(\frac{1}{M^2}\right) .$$

Finally, regarding to the third term E_3, we have

$$\begin{aligned}&\int_{\boldsymbol{T}}\int_{\boldsymbol{S}} \log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right] (p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) - \tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})) d\boldsymbol{s} d\boldsymbol{t} \\ &= \int_{\boldsymbol{T}}\int_{\boldsymbol{S}} \log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right] \left\{\exp\left[\sum_{|\boldsymbol{\alpha}|=3}\frac{D^{\boldsymbol{\alpha}} h_p(\hat{\boldsymbol{s}})}{\boldsymbol{\alpha}!}(\boldsymbol{s}-\hat{\boldsymbol{s}})^{\boldsymbol{\alpha}} + O_P(||\boldsymbol{s}-\hat{\mathbf{s}}||^4)\right] - 1\right\} \tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) d\boldsymbol{s} d\boldsymbol{t} .\end{aligned}$$

After the first order expansion of the exponential term, we obtain

$$\int_{\boldsymbol{T}}\int_{\boldsymbol{S}} \log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right] \left\{\sum_{|\boldsymbol{\alpha}|=3}\frac{D^{\boldsymbol{\alpha}} h_p(\hat{\boldsymbol{s}})}{\boldsymbol{\alpha}!}(\boldsymbol{s}-\hat{\boldsymbol{s}})^{\boldsymbol{\alpha}} + O_P(||\boldsymbol{s}-\hat{\mathbf{s}}||^4)\right\} \tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}}) d\boldsymbol{s} d\boldsymbol{t} .$$

Since $\log\left[\frac{p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})}{p(\boldsymbol{s},\boldsymbol{t})}\right]$ is $O_P(1)$ in $\boldsymbol{s}$ and the third moment of a multivariate Gaussian is zero, the rate of this error is dominant by

$$\int_{\boldsymbol{T}}\int_{\boldsymbol{S}} O_P(||\boldsymbol{s}-\hat{\mathbf{s}}||^4)\tilde{p}(\boldsymbol{s}|\boldsymbol{t}\bar{\boldsymbol{y}}) d\boldsymbol{s} d\boldsymbol{t} ,$$

which has already been shown to be inversely proportional to M^2. Now, it is straightforward to observe that the dominating term is $O_P\left(\frac{1}{M^2}\right)$. These three error terms are similar to the error terms in Appendices A, B and C in (Long, 2013), respectively.

B Proof of the error estimate of the conditional maximum posterior estimator $\hat{\boldsymbol{s}}$.

Let

$$R(\boldsymbol{s}) = \frac{1}{2}M(\boldsymbol{g}(\boldsymbol{\theta}_0) - \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{s},\boldsymbol{t}))^T \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{g}(\boldsymbol{\theta}_0) - \boldsymbol{g}(\boldsymbol{f}(\boldsymbol{s},\boldsymbol{t}))) + \boldsymbol{E}_s{}^T \boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{s},\boldsymbol{t})) - \boldsymbol{g}(\boldsymbol{\theta}_0)) - h_p(\boldsymbol{s},\boldsymbol{t}) .$$

We then have that

$$\nabla R(\boldsymbol{s}) = M\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}(\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{s},\boldsymbol{t})) - \boldsymbol{g}(\boldsymbol{\theta}_0)) + \boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s - \nabla h_p(\boldsymbol{s},\boldsymbol{t})\,,$$

where $\boldsymbol{J}_s$ is the Jacobian of $\boldsymbol{g}$ w.r.t. the parameters $\boldsymbol{s}$, and $\boldsymbol{E}_s$ is the summation of residual vectors defined in (Long, 2013). If we ignore the higher order terms, the first order expansion of $\nabla R(\boldsymbol{s})$ at $(\boldsymbol{0},\boldsymbol{t})$ reads

$$\nabla R(\boldsymbol{s}) = \nabla R(\boldsymbol{0}) + \nabla\nabla R(\boldsymbol{0})\boldsymbol{s}\,.$$

Therefore, using Newton's method $\nabla R(\hat{\boldsymbol{s}}) = \boldsymbol{0}$ implies that $\nabla R(\boldsymbol{0})+\nabla\nabla R(\boldsymbol{0})\hat{\boldsymbol{s}} = \boldsymbol{0}$. Thus,

$$\begin{aligned}\hat{\boldsymbol{s}} =& \boldsymbol{0} - (M\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^T\boldsymbol{J}_s + \boldsymbol{H}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s - \nabla\nabla h_p(\boldsymbol{0},\boldsymbol{t}))^{-1}(\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s - \nabla h_p(\boldsymbol{0},\boldsymbol{t}))\\ =& \boldsymbol{0} - (M\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^T\boldsymbol{J}_s + \boldsymbol{H}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s - \nabla\nabla h_p(\boldsymbol{0},\boldsymbol{t}))^{-1}(\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s)\\ &+ (M\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^T\boldsymbol{J}_s + \boldsymbol{H}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s - \nabla\nabla h_p(\boldsymbol{0},\boldsymbol{t}))^{-1}\nabla h_p(\boldsymbol{0},\boldsymbol{t})\,,\end{aligned}$$

where $\boldsymbol{H}_s$ is the Hessian of the model $\boldsymbol{g}$ w.r.t. the parameter $\boldsymbol{s}$.
As $M \to \infty$, $\hat{\boldsymbol{s}} = \boldsymbol{0}-(M\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^T\boldsymbol{J}_s+\boldsymbol{H}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s-\nabla\nabla h_p(\boldsymbol{0},\boldsymbol{t}))^{-1}(\boldsymbol{J}_s^T\boldsymbol{\Sigma}_\epsilon^{-1}\boldsymbol{E}_s)+ O_P\left(\frac{1}{M}\right) = O_P\left(\frac{1}{\sqrt{M}}\right)$.

C Proof of the error estimate in Equation (29)

We express the posterior distribution of the quantity of interest, Q, as

$$\begin{aligned}p(Q|\bar{\boldsymbol{y}}) =& \int_{\boldsymbol{T}} p(Q|\boldsymbol{t},\bar{\boldsymbol{y}})p(\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{t}\\ =& \int_{\boldsymbol{T}} \tilde{p}(Q|\boldsymbol{t},\bar{\boldsymbol{y}})p(\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{t} + \int_{\boldsymbol{T}} \left[p(Q|\boldsymbol{t},\bar{\boldsymbol{y}}) - \tilde{p}(Q|\boldsymbol{t},\bar{\boldsymbol{y}})\right] p(\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{t}\\ =& \int_{\boldsymbol{T}} \tilde{p}(Q|\boldsymbol{t},\bar{\boldsymbol{y}})(\int_{\boldsymbol{S}} \tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{s})d\boldsymbol{t} + \int_{\boldsymbol{T}} \tilde{p}(Q|\boldsymbol{t},\bar{\boldsymbol{y}})(\int_{\boldsymbol{S}} p(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}}) - \tilde{p}(\boldsymbol{s},\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{s})d\boldsymbol{t}\\ &+ \int_{\boldsymbol{T}} \left[p(Q|\boldsymbol{t},\bar{\boldsymbol{y}}) - \tilde{p}(Q|\boldsymbol{t},\bar{\boldsymbol{y}})\right] p(\boldsymbol{t}|\bar{\boldsymbol{y}})d\boldsymbol{t}\end{aligned}$$

(38)

$$= \tilde{p}(Q|\bar{\boldsymbol{y}}) + \int_{\boldsymbol{T}} \left[p(Q|\boldsymbol{t},\bar{\boldsymbol{y}}) - \tilde{p}(Q|\boldsymbol{t},\bar{\boldsymbol{y}})\right] p(\boldsymbol{t}|\bar{\boldsymbol{y}})dt + O_P\left(\frac{1}{M^2}\right)\,,$$

where the error rate is obtained in a way similar to the derivation of E_3. Furthermore, we have

$$\begin{aligned}
&p(Q|\boldsymbol{t},\bar{\boldsymbol{y}})\\
&\quad=\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t},\bar{\boldsymbol{y}})p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}\\
&\quad=\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t},\bar{\boldsymbol{y}})\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}+\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t},\bar{\boldsymbol{y}})\left[p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})\right]d\boldsymbol{s}\\
&\quad=\int_{\boldsymbol{S}} \exp\left\{\frac{[Q-\tau(\hat{\boldsymbol{s}},\boldsymbol{t})-\nabla\tau(\hat{\boldsymbol{s}},\boldsymbol{t})(\boldsymbol{s}-\hat{\boldsymbol{s}})]^2-[Q-\tau(\boldsymbol{s},\boldsymbol{t})]^2}{2\sigma_Q^2}\right\}\tilde{p}(Q|\boldsymbol{s},\boldsymbol{t},\bar{\boldsymbol{y}})\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}\\
&\qquad+\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t},\bar{\boldsymbol{y}})\left[p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})\right]d\boldsymbol{s}\\
&\quad=\int_{\boldsymbol{S}} \exp\left\{[(\boldsymbol{s}-\hat{\boldsymbol{s}})^T\nabla\nabla\tau(\hat{\boldsymbol{s}},\boldsymbol{t})(\boldsymbol{s}-\hat{\boldsymbol{s}})+O_P(||\boldsymbol{s}-\hat{\boldsymbol{s}}||^3)]O_P(||\boldsymbol{s}-\hat{\boldsymbol{s}}||)\right\}\tilde{p}(Q|\boldsymbol{s},\boldsymbol{t},\bar{\boldsymbol{y}})\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}\\
&\qquad+\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t})\left[p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})\right]d\boldsymbol{s}\\
&\quad=\int_{\boldsymbol{S}} \left\{[(\boldsymbol{s}-\hat{\boldsymbol{s}})^T\nabla\nabla\tau(\hat{\boldsymbol{s}},\boldsymbol{t})(\boldsymbol{s}-\hat{\boldsymbol{s}})+O_P(||\boldsymbol{s}-\hat{\boldsymbol{s}}||^3)]O_P(||\boldsymbol{s}-\hat{\boldsymbol{s}}||)+1\right\}\tilde{p}(Q|\boldsymbol{s},\boldsymbol{t})\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}\\
&\qquad+\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t})\left[p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})\right]d\boldsymbol{s}\\
&\quad=\int_{\boldsymbol{S}} \tilde{p}(Q|\boldsymbol{s},\boldsymbol{t})\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})d\boldsymbol{s}+\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t})\left[p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})\right]d\boldsymbol{s}+O_P\left(\frac{1}{M^2}\right).
\end{aligned}$$

By reusing the expansion of $p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})$ derived in E_3, we obtain

$$p(Q|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(Q|\boldsymbol{t},\bar{\boldsymbol{y}})=\int_{\boldsymbol{S}} p(Q|\boldsymbol{s},\boldsymbol{t})\left[p(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})-\tilde{p}(\boldsymbol{s}|\boldsymbol{t},\bar{\boldsymbol{y}})\right]d\boldsymbol{s}+O_P\left(\frac{1}{M^2}\right)=O_P\left(\frac{1}{M^2}\right).$$

Therefore,

$$p(Q|\bar{\boldsymbol{y}})=\tilde{p}(Q|\bar{\boldsymbol{y}})+O_P\left(\frac{1}{M^2}\right). \tag{39}$$

In this connection, the first error term in (29), i.e., the error term before marginalizing over data, can be obtained as

$$\begin{aligned}
&\int_{\mathcal{Q}} \log\left[\frac{p(Q|\bar{\boldsymbol{y}})}{\tilde{p}(Q|\bar{\boldsymbol{y}})}\right]\tilde{p}(Q|\bar{\boldsymbol{y}})dQ+\int_{\mathcal{Q}} p(Q|\bar{\boldsymbol{y}})\left(p(Q|\bar{\boldsymbol{y}})-\hat{p}(Q|\bar{\boldsymbol{y}})\right)dQ\\
&\quad=\int_{\mathcal{Q}} \log\left[\frac{p(Q|\bar{\boldsymbol{y}})-\tilde{p}(Q|\bar{\boldsymbol{y}})}{\tilde{p}(Q|\bar{\boldsymbol{y}})}+1\right]\tilde{p}(Q|\bar{\boldsymbol{y}})dQ+O_P\left(\frac{1}{M^2}\right)\\
&\quad=\int_{\mathcal{Q}} \left[\frac{p(Q|\bar{\boldsymbol{y}})-\tilde{p}(Q|\bar{\boldsymbol{y}})}{\tilde{p}(Q|\bar{\boldsymbol{y}})}+O_P\left(\frac{p(Q|\bar{\boldsymbol{y}})-\tilde{p}(Q|\bar{\boldsymbol{y}})}{\tilde{p}(Q|\bar{\boldsymbol{y}})}\right)^2\right]\tilde{p}(Q|\bar{\boldsymbol{y}})dQ+O_P\left(\frac{1}{M^2}\right)\\
&\quad=O_P\left(\frac{1}{M^2}\right).
\end{aligned}$$

References

1. Long Q., Scavino M., Tempone R., Wang S. (2013). “Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations”, *Computer Methods in Applied Mechanics and Engineering*, 259, 24–39.

2. Lindley D. (1956). “On a measure of information provided by an experiment”, *The Annals of Mathematical Statistics*, 27, 986–1005.

3. Chaloner K., Verdinelli I. (1995). “Bayesian experimental design: a review”, *Statistical Science*, 10, 273–304.

4. Ginebra J. (2007). “On the measure of the information in a statistical experiment”, *Bayesian Analysis*, 2, 167–212.

5. Clarke B.S., Barron A.R. (1991). “Entropy risk and the Bayesian central limit theorem”, Tech. Report 91-56, Department of Statistics, Purdue University, September.

6. Polson N.G. (1992). “On the expected amount of information from a non-linear model”, *Journal of the Royal Statistical Society*, Series B, 54, 889–895.

7. Ghosal S., Samanta T. (1997). “Expansion of Bayes risk for entropy loss and reference prior in nonregular cases”, *Statistics & Decisions*, 15, 129–140.

8. Bernardo J.M. (1979). “Reference posterior distributions for Bayesian inference”, *Journal of the Royal Statistical Society*, Series B, 41, 113–147.

9. Polson N.G. (1988). Bayesian Perspectives on Statistical Modelling, PhD thesis, Department of Mathematics, University of Nottingham, October.

10. Clarke B.S., Wasserman L. (1993). “Noninformative priors and nuisance parameters”, *Journal of the American Statistical Association*, 88, 1427–1432.

11. Clarke B.S., Yuan A. (2004). “Partial information reference priors: derivation and interpretations”, *Journal of Statistical Planning and Inference*, 123, 313–345.

12. Hager W.W. (1989). “Updating the inverse of a matrix”, *SIAM Review*, 31, 221–239.

13. Barthelmann V., Novak E., Ritter K. (2000). “High dimensional polynomial interpolation on sparse grids”, *Advances in Computational Mathematics*, 12, 273–288.

14. Garcke J. (2013). “Sparse Grids in a Nutshell” en Garcke J., Griebel M. (eds.), *Sparse Grids and Applications*, Lecture Notes in Computational Science and Engineering 88, Springer.

15. Nobile F., Tempone R., Webster C.G. (2008). “A sparse grid stochastic collocation method for partial differential equations with random input data”, *SIAM Journal on Numerical Analysis*, 46, 2309–2345.