



# Non-asymptotic error estimates for the Laplace approximation in Bayesian inverse problems

Tapio Helin<sup>1</sup> · Remo Kretschmann<sup>1,2</sup>

Received: 22 December 2020 / Revised: 26 October 2021 / Accepted: 10 December 2021 /  
Published online: 15 January 2022  
© The Author(s) 2022

## Abstract

In this paper we study properties of the Laplace approximation of the posterior distribution arising in nonlinear Bayesian inverse problems. Our work is motivated by Schillings et al. (Numer Math 145:915–971, 2020. <https://doi.org/10.1007/s00211-020-01131-1>), where it is shown that in such a setting the Laplace approximation error in Hellinger distance converges to zero in the order of the noise level. Here, we prove novel error estimates for a given noise level that also quantify the effect due to the nonlinearity of the forward mapping and the dimension of the problem. In particular, we are interested in settings in which a linear forward mapping is perturbed by a small nonlinear mapping. Our results indicate that in this case, the Laplace approximation error is of the size of the perturbation. The paper provides insight into Bayesian inference in nonlinear inverse problems, where linearization of the forward mapping has suitable approximation properties.

**Mathematics Subject Classification** 47J06 · 62E17 · 62E20 · 62F15 · 65D30 · 65D32 · 65J20 · 65J22

## 1 Introduction

The study of Bayesian inverse problems [8,26] has gained wide attention during the last decade as the increase in computational resources and algorithmic development have enabled uncertainty quantification in numerous new applications in science and

---

✉ Remo Kretschmann  
remo.kretschmann@mathematik.uni-wuerzburg.de  
Tapio Helin  
tapio.helin@lut.fi

<sup>1</sup> School of Engineering Science, LUT University, P.O. Box 20, 53851 Lappeenranta, Finland

<sup>2</sup> Institute of Mathematics, University of Würzburg, Emil-Fischer-Str. 30, 97074 Würzburg, Germany

engineering. Large-scale problems, where the computational burden of the likelihood is prohibitive, are, however, still a subject of ongoing research.

In this paper we study the Laplace approximation of the posterior distribution arising in nonlinear Bayesian inverse problems. The Laplace approximation is obtained by replacing the log-posterior density with its second order Taylor approximation around the maximum a posteriori (MAP) estimate and renormalizing the density. This produces a Gaussian measure centered at the maximum a posteriori (MAP) estimate with a covariance corresponding to the Hessian of the negative log-posterior density (see, e.g., [3, Section 4.4]).

The asymptotic behavior of the parametric Laplace approximation in the small noise or large data limit has been studied extensively in the past (see, e.g., [30]). We note that in terms of approximation properties with respect to taking a posterior expectation over a given function, there is a long line of research which we discuss below. Our work is parallel to this effort in that we aim to estimate the total variation (TV) distance between the two probability measures. On the one hand, the error in TV distance bounds the error of the expectation of any function with respect to the Laplace approximation. On the other hand, it is a measure of the non-Gaussianity of the posterior distribution. Thus, our results describe and quantify how the nonlinearity of the forward mapping translates into non-Gaussianity of the posterior distribution.

Our work is motivated by a recent result by Schillings, Sprungk, and Wacker in [25], where the authors show that in the context of Bayesian inverse problems, the Laplace approximation error in Hellinger distance converges to zero in the order of the noise level. In practice, one is, however, often interested in estimating the error for a given, *fixed* noise level. It can, e.g., be unclear if the noise level is small enough in order to dominate the error estimate. Indeed, the nonlinearity of the forward mapping (more generally, the non-Gaussianity of the likelihood) or a large problem dimension can have a significant contribution to the constant appearing in the asymptotic estimates. Therefore, it is of interest to quantify such effects in non-asymptotic error estimates for the Laplace approximation. This is the main goal of our work.

## 1.1 Our contributions

The main contribution of this work is threefold:

1. In Theorem 3.4, we derive our central error estimate for the total variation distance of the Laplace posterior approximation in nonlinear Bayesian inverse problems. The error bound consists of two error terms for which we derive an implicit optimal balancing rule in Proposition 3.13. We assume uniform bounds on the third differentials of log-likelihood and log-prior density as well as a quadratic lower bound on the log-posterior density to control the error. Given such bounds the error estimate can be numerically evaluated.
2. In Theorem 4.1, we derive a further estimate for the Laplace approximation error that makes the effect of noise level, the bounds specified above and the dimension of the problem explicit. This error estimate readily implies linear rates of convergence for fixed problem dimension both in the small noise limit and when the third differential of the log-likelihood goes to zero, see Corollary 4.4. It fur-

thermore leads to a convergence rate for increasing problem dimension in terms of noise level, problem dimension, and aforementioned bounds aligned with [15], see Corollary 4.6.

3. In Theorem 5.3, we quantify the error of the Laplace approximation in terms of the nonlinearity of the forward mapping for linear inverse problems with small nonlinear perturbation and Gaussian prior distribution. We assume uniform bounds on the differentials of the nonlinear perturbation of up to third order to control the error. This error estimate immediately implies linear convergence in terms of the size of the perturbation. Moreover, such a result provides insight into Bayesian inference in nonlinear inverse problems, where linearization of the forward mapping has suitable approximation properties.

## 1.2 Relevant literature

The asymptotic approximation of general integrals of the form  $\int e^{\lambda f(x)} g(x) dx$  by Laplace's method is presented in [21,30]. Non-asymptotic error bounds for the Laplace approximation of such integrals have been stated in the univariate [20] and multivariate case [7,16]. The Laplace approximation error and its convergence in the limit  $\lambda \rightarrow \infty$  have been estimated in the multivariate case when the function  $f$  depends on  $\lambda$  or the maximizer of  $f$  is on the boundary of the integration domain [10]. A representation of the coefficients appearing in the asymptotic expansion of the approximated integral utilizing ordinary potential polynomials is given in [18].

The error estimates on the Laplace approximation in TV distance are closely connected to the so-called Bernstein–von Mises (BvM) phenomenon that quantifies the convergence of the scaled posterior distribution toward a Gaussian distribution in the large data or small noise limit. Parametric BvM theory is well-understood [12,29]. Our work is inspired by a BvM result by Lu [15], where a parametric BvM theorem for nonlinear Bayesian inverse problems with an increasing number of parameters is proved. Similar to our objectives, he quantifies the asymptotic convergence rate in terms of noise level, nonlinearity of the forward mapping and dimension of the problem. However, our emphasis differs from [15] (and other BvM results) in that we are not restricted to considering the vanishing noise limit, but are more interested in quantifying the effect of small nonlinearity or dimension at a fixed noise level. We also point out that BvM theory has been developed for non-parametric Bayesian inverse problems (see, e.g., [6,17,19]), where the convergence is quantified in a distance that metrizes the weak convergence.

Let us conclude by briefly emphasizing that the Laplace approximation is widely utilized for different purposes in computational Bayesian statistics including, i.a., the celebrated INLA algorithm [22]. It has also recently gained popularity in optimal Bayesian experimental design (see, e.g., [1,14,23]). Moreover, it provides a convenient reference measure for numerical quadrature [4,24] or importance sampling [2].

### 1.3 Organization of the paper

Before we present the aforementioned three main results in Sects. 3 to 4 to 5, we introduce our set-up and notation, Laplace's method, and the total variation metric in Sect. 2. In Sect. 3, we introduce our central error bound for the Laplace approximation and explain the idea behind its proof. In Sect. 4, we derive an explicit error estimate for the Laplace approximation and describe its asymptotic behavior. In Sect. 5, we prove the error estimate for inverse problems with small nonlinearity in the forward mapping and Gaussian prior distribution.

## 2 Preliminaries and set-up

We consider for  $\varepsilon > 0$  the inverse problem of recovering  $x \in \mathbb{R}^d$  from a noisy measurement  $y \in \mathbb{R}^d$ , where

$$y = G(x) + \sqrt{\varepsilon}\eta,$$

$\eta \in \mathbb{R}^d$  is random noise with standard normal distribution  $\mathcal{N}(0, I_d)$ , and  $G: \mathbb{R}^d \rightarrow \mathbb{R}^d$  is a possibly nonlinear mapping. In the following,  $|\cdot|$  denotes the Euclidean norm on  $\mathbb{R}^d$ . If we assume a prior distribution  $\mu$  on  $\mathbb{R}^d$  with Lebesgue density  $\exp(-R(x))$ , then Bayes' formula yields a posterior distribution  $\mu^y$  with density

$$\mu^y(dx) \propto \exp\left(-\frac{1}{2\varepsilon}|y - G(x)|^2\right) \mu(dx) = \exp\left(-\frac{1}{2\varepsilon}|y - G(x)|^2 - R(x)\right) dx. \quad (2.1)$$

For all  $x, y \in \mathbb{R}^d$ , we denote the *scaled* negative log-likelihood by

$$\Phi(x) = \frac{1}{2}|y - G(x)|^2.$$

If

$$x \mapsto \Phi(x) + \varepsilon R(x)$$

has a unique minimizer in  $\mathbb{R}^d$ , we call this minimizer the *maximum a posteriori (MAP) estimate* and denote it by  $\hat{x} = \hat{x}(y)$ . Furthermore, we set

$$I(x) := \Phi(x) + \varepsilon R(x) - \Phi(\hat{x}) - \varepsilon R(\hat{x})$$

for all  $x \in \mathbb{R}^d$ . This way,  $I$  is nonnegative, the MAP estimate  $\hat{x}$  minimizes  $I$  and satisfies  $I(\hat{x}) = 0$ . Moreover, we can express the posterior density as

$$\mu^y(dx) = \frac{1}{Z} \exp\left(-\frac{1}{\varepsilon}I(x)\right) dx \quad (2.2)$$

with a normalization constant  $Z$ .

Laplace’s method approximates the posterior distribution by a Gaussian distribution  $\mathcal{L}_{\mu^y}$  whose mean and covariance are chosen in such a way that its log-density agrees, up to a constant, with the second order Taylor polynomial around  $\hat{x}$  of the log-posterior density. If  $I \in C^2(\mathbb{R}^d, \mathbb{R})$ , the Laplace approximation of  $\mu^y$  is defined as

$$\mathcal{L}_{\mu^y} := \mathcal{N}(\hat{x}, \varepsilon \Sigma),$$

where  $\Sigma := (D^2I(\hat{x}))^{-1}$ . Here,  $DI$  denotes the differential of  $I$ , and we identify  $D^2I(\hat{x})$  with the Hessian matrix  $\{D^2I(\hat{x})(e_j, e_k)\}_{j,k=1}^d$ . The Lebesgue density of  $\mathcal{L}_{\mu^y}$  is given by

$$\begin{aligned} \mathcal{L}_{\mu^y}(dx) &= \frac{1}{Z} \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2\right) dx \\ &= \frac{1}{Z} \exp\left(-\frac{1}{2\varepsilon} D^2I(\hat{x})(x - \hat{x}, x - \hat{x})\right) dx, \end{aligned}$$

where

$$\tilde{Z} = \int_{\mathbb{R}^d} \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2\right) dx = \varepsilon^{\frac{d}{2}} (2\pi)^{\frac{d}{2}} \sqrt{\det \Sigma}. \tag{2.3}$$

Since  $I(\hat{x}) = 0$  and  $DI(\hat{x}) = 0$ ,  $\frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2$  is precisely the truncated Taylor series of  $I/\varepsilon$  around  $\hat{x}$ .

The *total variation (TV) distance* between two probability measures  $\nu$  and  $\mu$  on  $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$  is defined as

$$d_{TV}(\nu, \mu) = \sup_{A \in \mathcal{B}(\mathbb{R}^d)} |\nu(A) - \mu(A)|,$$

see Section 2.4 in [27]. It has the alternative representation

$$d_{TV}(\nu, \mu) = \frac{1}{2} \sup_{\|f\|_{\infty} \leq 1} \left\{ \int_{\mathbb{R}^d} f d\nu - \int_{\mathbb{R}^d} f d\mu \right\} = \frac{1}{2} \int_{\mathbb{R}^d} \left| \frac{d\nu}{d\rho} - \frac{d\mu}{d\rho} \right| d\rho$$

where  $\|f\|_{\infty} := \sup_{x \in \mathbb{R}^d} |f(x)|$  and  $\rho$  can be any probability measure dominating both  $\mu$  and  $\nu$ , see Remark 5.9 in [27] and equation (1.12) in [11]. The total variation distance is valuable for the purpose of uncertainty quantification because it bounds the error of any credible region when using a measure  $\nu$  instead of another measure  $\mu$ . It can, moreover, be used to bound the difference in expectation of any bounded function  $f$  on  $\mathbb{R}^d$  with respect to  $\mu$  and  $\nu$ , respectively, by

$$|\mathbb{E}^{\nu}[f] - \mathbb{E}^{\mu}[f]| \leq 2\|f\|_{\infty} d_{TV}(\nu, \mu),$$

see Lemma 1.32 in [11]. By Kraft’s inequality

$$d_H(\mu, \nu)^2 \leq d_{TV}(\mu, \nu) \leq \sqrt{2}d_H(\mu, \nu),$$

the total variation distance bounds the square of the *Hellinger distance*

$$d_H(\mu, \nu) = \left( \frac{1}{2} \int_{\mathbb{R}^d} \left| \sqrt{\frac{d\nu}{d\rho}} - \sqrt{\frac{d\mu}{d\rho}} \right|^2 d\rho \right)^{\frac{1}{2}},$$

see Definition 1.28 and Lemma 1.29 in [11] or [9], while both metrics induce the same topology. The *bounded Lipschitz metric*

$$d_{BL}(\mu, \nu) = \frac{1}{2} \sup_{\|f\|_\infty + \|f\|_{Lip} \leq 1} \left\{ \int_{\mathbb{R}^d} f d\mu - \int_{\mathbb{R}^d} f d\nu \right\},$$

which induces the topology of weak convergence of probability measures, is trivially bounded by the total variation distance. Here, we denote

$$\|f\|_{Lip} := \sup_{x, y \in \mathbb{R}^d, x \neq y} \frac{|f(x) - f(y)|}{|x - y|}.$$

For further information on the relation between the total variation distance and other probability metrics we refer the survey paper [5].

### 3 Central error estimate

We will use the following ideas to bound the error of the Laplace approximation  $\mathcal{L}_{\mu^y}$  for a given realization of the data  $y \in \mathbb{R}^d$ . First, we will prove the fundamental estimate

$$d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) \leq \frac{1}{Z} \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} I(x)\right) - \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_\Sigma^2\right) \right| dx. \tag{3.1}$$

If we have a radial upper bound  $f(\|x - \hat{x}\|_\Sigma)$  for the integrand on the right hand side of (3.1), we can estimate

$$d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) \leq \int_{\mathbb{R}^d} f(\|x - \hat{x}\|_\Sigma) dx = \sqrt{\det \Sigma} \int_{\mathbb{R}^d} f(|u|) du,$$

where we applied a change of variable to a local parameter  $u := \Sigma^{-\frac{1}{2}}(x - \hat{x})$ . This integral, we can now express as a 1-dimensional integral using polar coordinates.

The integrand on the right hand side of (3.1) is very small and flat around  $\hat{x}$ , since  $\frac{1}{2}\|x - \hat{x}\|_\Sigma^2$  is the second order Taylor expansion of  $I(x)$  around  $\hat{x}$ , and it falls off as  $|x| \rightarrow \infty$  because it is integrable. Its mass is thus concentrated in an intermediate distance from  $\hat{x}$ . This can be seen, e.g., in Fig. 1. We exploit this structure by splitting up the integral in (3.1) and bounding the integrand on a  $\Sigma$ -norm ball

$$U(r_0) := \{x \in \mathbb{R}^d : \|x - \hat{x}\|_\Sigma \leq r_0\}$$

around the MAP estimate  $\hat{x}$  and on the remaining space  $\mathbb{R}^d \setminus U(r_0)$  separately. On  $U(r_0)$ , we then control the integrand by imposing uniform bounds on the third order differentials of the log-likelihood and the log-prior density. Outside of  $U(r_0)$ , we control it by imposing a quadratic lower bound on  $I$ .

We make the following assumptions on  $\Phi$ ,  $R$ ,  $I$ ,  $\hat{x}$ , and  $\Sigma$ , which will be further discussed in Remark 3.9.

**Assumption 3.1** We have  $\Phi, R \in C^3(\mathbb{R}^d, \mathbb{R})$ ,  $I$  has a unique global minimizer  $\hat{x} = \hat{x}(y) \in \mathbb{R}^d$  and  $D^2I(\hat{x})$  is positive definite.

**Assumption 3.2** There exists a constant  $K > 0$  such that

$$\max \left\{ \|D^3\Phi(x)\|_{\Sigma}, \|D^3R(x)\|_{\Sigma} \right\} \leq K$$

for all  $x \in \mathbb{R}^d$ , where

$$\|D^3\Phi(x)\|_{\Sigma} := \sup \left\{ \left| D^3\Phi(x)(h_1, h_2, h_3) \right| : \|h_1\|_{\Sigma}, \|h_2\|_{\Sigma}, \|h_3\|_{\Sigma} \leq 1 \right\}.$$

**Assumption 3.3** There exists  $0 < \delta \leq 1$  such that

$$I(x) \geq \frac{\delta}{2} \|x - \hat{x}\|_{\Sigma}^2 \quad \text{for all } x \in \mathbb{R}^d.$$

Let  $\Gamma(z)$  denote the classical gamma function and  $\gamma(a, z)$  the lower incomplete gamma function. Then,

$$\mathcal{E}_d(t) := \frac{\gamma\left(\frac{d}{2}, \frac{t}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \quad \text{for all } t \geq 0, d > 0,$$

describes the probability of a Euclidean ball in  $\mathbb{R}^d$  with radius  $\sqrt{t}$  around 0 under a standard Gaussian measure (see Lemma 3.12).

The main result of this section is the following error estimate.

**Theorem 3.4** *Suppose that Assumptions 3.1–3.3 hold. Then we have*

$$d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) \leq E_1(r_0) + E_2(r_0) \tag{3.2}$$

for all  $r_0 \geq 0$ , where

$$E_1(r_0) := c_d \varepsilon^{-\frac{d}{2}} \int_0^{r_0} f(r) r^{d-1} dr, \tag{3.3}$$

$$E_2(r_0) := \delta^{-\frac{d}{2}} \left( 1 - \mathcal{E}_d\left(\frac{\delta r_0^2}{\varepsilon}\right) \right) \tag{3.4}$$

for all  $r_0 \geq 0$ ,

$$f(r) := \left[ \exp\left(\frac{(1 + \varepsilon)K}{6\varepsilon} r^3\right) - 1 \right] \exp\left(-\frac{1}{2\varepsilon} r^2\right) \tag{3.5}$$

for all  $r \geq 0$ , and

$$c_d := \frac{2^{1-\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}.$$

**Remark 3.5** The two functions  $E_1$  and  $E_2$  are continuous and monotonic with the following asymptotic behavior. The first error term  $E_1(r_0)$  obeys

$$E_1(0) = 0, \quad \lim_{r_0 \rightarrow \infty} E_1(r_0) = \infty,$$

whereas the second error term  $E_2(r_0)$  satisfies

$$E_2(0) = 2\delta^{-\frac{d}{2}}, \quad \lim_{r_0 \rightarrow \infty} E_2(r_0) = 0.$$

This can be seen as follows.

The function  $f(r)r^{d-1}$  is bounded on the interval  $[0, 1]$ , so that the integral  $\int_0^{r_0} f(r)r^{d-1}dr$  converges to 0 as  $r_0 \rightarrow 0$ , and hence also  $E_1(r_0)$ . On the other hand,  $f(r)r^{d-1}$  converges to  $\infty$  as  $r \rightarrow \infty$ , so that the integral  $\int_0^{r_0} f(r)r^{d-1}dr$  and  $E_1(r_0)$  converge to  $\infty$  as  $r \rightarrow \infty$ . Since  $f(r)r^{d-1}$  is positive for all  $r \geq 0$ ,  $E_1$  moreover increases monotonically. By definition of the lower incomplete gamma function,  $\mathcal{E}_d$  increases monotonically and  $\mathcal{E}_d(t) \in [0, 1]$  for all  $t \geq 0$  and  $d > 0$ . Moreover,  $\mathcal{E}_d(t) \rightarrow 0$  as  $t \rightarrow 0$  and  $\mathcal{E}_d(t) \rightarrow 1$  as  $t \rightarrow \infty$ . Consequently,  $E_2(r_0)$  converges toward  $2\delta^{-d/2}$  as  $r_0 \rightarrow 0$ , and toward 0 as  $r_0 \rightarrow \infty$ . The asymptotic behavior of  $E_2$  is described more precisely in Lemma 4.3.

The following three propositions formalize the ideas described in the beginning of this section and constitute the prove of Theorem 3.4.

**Proposition 3.6** (Fundamental estimate) *The Laplace approximation  $\mathcal{L}_{\mu^y}$  of  $\mu^y$  satisfies*

$$d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) \leq \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx),$$

where  $R_2(x) := I(x) - \frac{1}{2}\|x - \hat{x}\|_{\Sigma}^2$  for all  $x \in \mathbb{R}^d$ .

**Proof** For a fixed  $\varepsilon > 0$  we can estimate

$$2d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) = \int_{\mathbb{R}^d} \left| \frac{1}{Z} \exp\left(-\frac{1}{\varepsilon} I(x)\right) - \frac{1}{Z} \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2\right) \right| dx$$



$$\begin{aligned}
 &= \frac{1}{\tilde{Z}} \int_{\mathbb{R}^d} \left| \frac{\tilde{Z}}{Z} \exp\left(-\frac{1}{\varepsilon} I(x)\right) - \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2\right) \right| dx \\
 &\leq \frac{1}{\tilde{Z}} \left| \frac{\tilde{Z}}{Z} - 1 \right| \int_{\mathbb{R}^d} \exp\left(-\frac{1}{\varepsilon} I(x)\right) dx \\
 &\quad + \frac{1}{\tilde{Z}} \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} I(x)\right) - \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2\right) \right| dx \\
 &= \frac{1}{\tilde{Z}} |\tilde{Z} - Z| + \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} I(x) + \frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2\right) - 1 \right| \mathcal{L}_{\mu^y}(dx). \\
 &= \frac{1}{\tilde{Z}} |\tilde{Z} - Z| + \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx).
 \end{aligned}$$

Now, the estimate

$$\begin{aligned}
 |Z - \tilde{Z}| &\leq \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} I(x)\right) - \exp\left(-\frac{1}{2\varepsilon} \|x - \hat{x}\|_{\Sigma}^2\right) \right| dx \\
 &= \tilde{Z} \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx)
 \end{aligned}$$

yields the proposition. □

**Proposition 3.7** (Close range estimate) *Suppose that Assumption 3.2 holds. Then it follows that*

$$\int_{U(r_0)} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx) \leq c_d \varepsilon^{-\frac{d}{2}} \int_0^{r_0} f(r) r^{d-1} dr$$

for all  $r_0 \geq 0$ , where  $f$  and  $c_d$  are defined as in Theorem 3.4.

**Proposition 3.8** (Far range estimate) *Suppose that Assumption 3.3 holds. Then we have*

$$\int_{\mathbb{R}^d \setminus U(r_0)} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx) \leq \delta^{-\frac{d}{2}} \left( 1 - \Xi_d\left(\frac{\delta r_0^2}{\varepsilon}\right) \right) \tag{3.6}$$

for all  $r_0 \geq 0$ .

The proof of Theorem 3.4 is now very short.

**Proof of Theorem 3.4** By Proposition 3.6 we have

$$d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) \leq \int_{\mathbb{R}^d} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx).$$

Now, splitting up this integral into integrals over  $U(r_0)$  and its complement and applying Propositions 3.7 and 3.8 proves the statement. □

- Remark 3.9** 1. Because of  $I(\hat{x}) = 0$  and the necessary optimality condition  $DI(\hat{x}) = 0$ , the function  $R_2(x) = I(x) - \frac{1}{2}\|x - \hat{x}\|_{\Sigma}^2$  defined in Proposition 3.6 is precisely the remainder of the second order Taylor polynomial of  $I$  around  $\hat{x}$ . In Proposition 3.7, Assumption 3.2 is used to control  $R_2$  near the MAP estimate by bounding the third order differential of  $I$ . In Proposition 3.8, in turn, Assumption 3.3 is used to control  $R_2$  at a distance from  $\hat{x}$  by bounding it from below by  $-\frac{1-\delta}{2}\|x - \hat{x}\|_{\Sigma}^2$ .
2. The constant  $K \geq 0$  in Assumption 3.2 quantifies the non-Gaussianity of the likelihood and the prior distribution and can be arbitrarily large. Assumption 3.3 bounds the unnormalized log-posterior density from above by a multiple of the unnormalized log-density of the Laplace distribution, where the constant  $\delta > 0$  represents the scaling factor and can be arbitrarily small. This restricts our results to posterior distributions whose tail does not decay slower than that of a Gaussian distribution. Assumption 3.3 can for example be violated if a prior distribution with heavier than Gaussian tail is chosen such as a Cauchy distribution and if the forward mapping is linear but singular. Our main interest lies on inverse problems with a posterior distribution that is not too different from a Gaussian distribution, since this is a setting in which the Laplace approximation can be expected to yield reasonable results.
3. In case of a linear inverse problem and a Gaussian prior distribution, the Laplace approximation is exact, so that Assumptions 3.2 and 3.3 are trivially satisfied with  $K = 0$  and  $\delta = 1$ . We will see in Sect. 5 that Assumptions 3.2 and 3.3 are satisfied for nonlinear inverse problems with  $\delta$  and  $K$  as given in Propositions 5.6 and 5.7 if the prior distribution is Gaussian and the nonlinearity of the forward mapping is small enough. In this case, the quadratic lower bound on  $I$  in Assumption 3.3 restricts the nonlinearity of the forward mapping to be small enough such that the tail of the posterior distribution does not decay slower than that of a Gaussian distribution.
4. Note that neither in Sect. 3 nor in Sect. 4 we make use of the Gaussianity of the noise. Therefore, the results of these sections remain valid for non-Gaussian noise as long as the log-likelihood satisfies Assumptions 3.2 and 3.3. In case of noise with a log-density  $-\nu \in C^3(\mathbb{R}^d)$ , the negative log-likelihood takes the form  $\Phi(x) = \nu(y - G(x))$  and we have  $I(x) = \nu(y - G(x)) + \varepsilon R(x) + c$ . Consider for example standard multivariate Cauchy noise, where

$$\nu(\eta) = -\ln \left[ \frac{C}{(1 + |\eta|^2)^{\frac{d+1}{2}}} \right] = \frac{d+1}{2} \ln(1 + |\eta|^2) - \ln C$$

for all  $\eta \in \mathbb{R}^d$ . The derivatives of up to third order of  $s \mapsto \ln(1 + s)$ ,  $s \geq 0$ , are bounded since they are continuous and converge to 0 as  $s$  tends to infinity. By the smoothness of  $x \mapsto |x|^2$ ,  $\nu$  is therefore in  $C^3(\mathbb{R}^d)$  and

$$\|D^3 \nu(x)\| := \sup_{|h_1|, |h_2|, |h_3| \leq 1} \left| D^3 \nu(x)(h_1, h_2, h_3) \right|$$

is uniformly bounded. In case of a linear forward mapping, the uniform boundedness transfers to  $\|D^3\Phi(x)\|$  and we can estimate

$$\|D^3\Phi(x)\|_{\Sigma} \leq \left\| \Sigma^{-\frac{1}{2}} \right\|^{-3} \|D^3\Phi(x)\|$$

for any symmetric positive definite matrix  $\Sigma$ .

5. We make Assumptions 3.2 and 3.3 globally, i.e., for all  $x \in \mathbb{R}^d$ , for the sake of simplicity. For a given  $r_0 \geq 0$ , Theorem 3.4 remains valid if Assumption 3.2 only holds for  $\|x - \hat{x}\|_{\Sigma} \leq r_0$  and if Assumption 3.3 only holds for  $\|x - \hat{x}\|_{\Sigma} \geq r_0$ . This allows for prior distributions which are not supported on the whole space  $\mathbb{R}^d$ , as long as the support of the prior contains the set  $U(r_0)$  and  $R \in C^3(U(r_0))$ . In this case,  $R$  and  $I$  are allowed to take values in  $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$  and we follow the convention  $\exp(-\infty) = 0$ .
6. The constant  $K$  in Assumption 3.2 can be replaced by a radial bound  $\rho(\|x - \hat{x}\|_{\Sigma})$  with a monotonically increasing function  $\rho$ . This way, an estimate of the form (3.2) can be obtained with  $f$  replaced by

$$\tilde{f}(r) = \left( \exp\left(\frac{1 + \varepsilon}{6\varepsilon} \rho(r)r^3\right) - 1 \right) \exp\left(-\frac{1}{2\varepsilon}r^2\right).$$

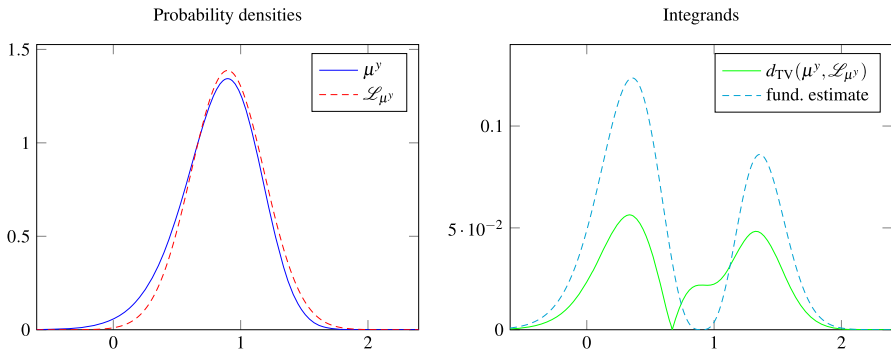
**Remark 3.10** Both the unnormalized posterior density  $\exp(-\frac{1}{\varepsilon}I(x))$  and the unnormalized Gaussian density  $\exp(-\frac{1}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2)$  attain their maximum 1 in  $\hat{x}$ . The densities of  $\mu^y$  and  $\mathcal{L}_{\mu^y}$  themselves, however, take the values  $1/Z$  and  $1/\tilde{Z}$  in  $\hat{x}$  due to the different normalization, see Figure 1. For this reason, the probability of small balls around  $\hat{x}$  under  $\mu^y$  and  $\mathcal{L}_{\mu^y}$  differs asymptotically by a factor of  $\tilde{Z}/Z$ . This has several consequences in case that the normalization constants  $Z$  and  $\tilde{Z}$  differ considerably.

One the one hand, credible regions around  $\hat{x}$  may have considerably different size under the posterior distribution and its Laplace approximation. On the other hand, the integrand

$$\frac{1}{2} \left| \frac{\exp(-\frac{1}{\varepsilon}I(x))}{Z} - \frac{\exp(-\frac{1}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2)}{\tilde{Z}} \right|$$

of the total variation distance  $d_{TV}(\mu^y, \mathcal{L}_{\mu^y})$  may, unlike the integrand of the fundamental estimate (3.1), have a significant amount of mass around  $\hat{x}$ , see Figure 1. This means that a significant portion of the error when approximating the probability of an event under  $\mu^y$  by that under  $\mathcal{L}_{\mu^y}$  may be due to the difference in their densities near the MAP estimate  $\hat{x}$ . So although the Laplace approximation is defined by the local properties of the posterior distribution in the MAP estimate, it is not necessarily a good local approximation around it.

A large difference in the normalization constants  $Z$  and  $\tilde{Z}$  as mentioned above reflects that the log-posterior density cannot be approximated well globally by its second order Taylor polynomial around  $\hat{x}$ . In the proof of Proposition 3.6, we saw that the difference in normalization is in fact bounded by the total variation of the



**Fig. 1** The probability densities of a posterior distribution  $\mu^y$  and its Laplace approximation  $\mathcal{L}_{\mu^y}$  (left), as well as the integrands of the total variation distance between  $\mu^y$  and  $\mathcal{L}_{\mu^y}$  and of the fundamental estimate (3.1) (right)

unnormalized densities. The value of Proposition 3.6 lies in providing an estimate for the total variation error that only involves unnormalized densities.

In the following sections, we present the proofs of our close and far range estimate, and characterize the optimal choice of  $r_0$ .

### 3.1 Proof of Proposition 3.7

We consider the close range integral

$$\int_{U(r_0)} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx)$$

over the  $\Sigma$ -norm ball with radius  $r_0 \geq 0$ . The proof of our close range estimate is based upon the following estimate for the remainder term  $R_2(x)$ .

**Lemma 3.11** *If Assumption 3.2 holds, then we have*

$$|R_2(x)| \leq \frac{1 + \varepsilon}{6} K \|x - \hat{x}\|_{\Sigma}^3 \text{ for all } x \in \mathbb{R}^d. \tag{3.7}$$

**Proof** We set  $h := x - \hat{x}$  and write the remainder of the second order Taylor polynomial of  $\Phi$  in mean-value form as

$$R_{2,\Phi}(x) := \Phi(x) - \Phi(\hat{x}) - D\Phi(\hat{x})(h) - \frac{1}{2} D^2\Phi(\hat{x})(h, h) = \frac{1}{6} D^3\Phi(z)(h, h, h)$$

for some  $z \in \hat{x} + [0, 1]h$ . Since  $\hat{x} + [0, 1]h \subset U(\|x - \hat{x}\|_{\Sigma})$ , we can now use the multilinearity of  $D^3\Phi(z)$  to express  $R_{2,\Phi}$  as

$$R_{2,\Phi}(x) = \frac{1}{6} D^3\Phi(z) \left( \frac{h}{\|h\|_{\Sigma}}, \frac{h}{\|h\|_{\Sigma}}, \frac{h}{\|h\|_{\Sigma}} \right) \|h\|_{\Sigma}^3$$

for some  $z \in U(\|x - \hat{x}\|_\Sigma)$ , and estimate

$$|R_{2,\phi}(x)| \leq \frac{1}{6} \|D^3\Phi(z)\|_\Sigma \|h\|_\Sigma^3 \leq \frac{1}{6} K \|h\|_\Sigma^3$$

using Assumption 3.2. By proceeding similarly for  $R$ , we now obtain

$$|R_2(x)| \leq |R_{2,\phi}(x)| + \varepsilon |R_{2,R}(x)| \leq \frac{1}{6} K \|h\|_\Sigma^3 + \frac{\varepsilon}{6} K \|h\|_\Sigma^3.$$

□

Now, we can prove our close range estimate.

**Proof of Proposition 3.7** By Lemma 3.11 and (2.3), we have

$$\begin{aligned} \int_{U(r_0)} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx) &\leq \frac{1}{Z} \int_{U(r_0)} f(\|x - \hat{x}\|_\Sigma) dx \\ &= \frac{1}{Z} \sqrt{\det \Sigma} \int_{\{u \in \mathbb{R}^d; |u| \leq r_0\}} f(|u|) du \\ &= \varepsilon^{-\frac{d}{2}} (2\pi)^{-\frac{d}{2}} \cdot d\kappa_d \int_0^{r_0} f(r) r^{d-1} dr, \end{aligned}$$

since  $|u| = \|x - \hat{x}\|_\Sigma$ . Here

$$\kappa_d = \frac{\pi^{d/2}}{\Gamma(d/2 + 1)}$$

denotes the *volume* of the  $d$ -dimensional Euclidean unit ball ( $d\kappa_d$  is its surface area). Using the fundamental recurrence  $\Gamma(z + 1) = z\Gamma(z)$ , we can write

$$d\kappa_d = \frac{2\pi^{d/2}}{\Gamma(d/2)}.$$

□

### 3.2 Proof of Proposition 3.8

Now, we consider the integral

$$\int_{\mathbb{R}^d \setminus U(r_0)} \left| \exp\left(-\frac{1}{\varepsilon} R_2(x)\right) - 1 \right| \mathcal{L}_{\mu^y}(dx)$$

over the space outside of a  $\Sigma$ -norm ball with radius  $r_0 \geq 0$ . In the proof of our far range estimate, the following expression is used to describe the probability of  $\mathbb{R}^d \setminus U(r_0)$  under  $\mathcal{L}_{\mu^y}$ . Let  $\Gamma(a, z)$  denote the upper incomplete gamma function.

**Lemma 3.12** *Let  $v = \mathcal{N}(\hat{x}, \delta^{-1}\varepsilon\Sigma)$  with  $\delta > 0$ . Then,*

$$v(\mathbb{R}^d \setminus U(r_0)) = \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{d}{2}, \frac{\delta r_0^2}{2\varepsilon}\right) \text{ for all } r_0 \geq 0.$$

**Proof** We compute the tail integral explicitly using a local parameter and polar coordinates. This yields

$$\begin{aligned} v(\{\|x - \hat{x}\|_{\Sigma} \geq r_0\}) &= \frac{\delta^{d/2}}{\varepsilon^{d/2}(2\pi)^{d/2}\sqrt{\det \Sigma}} \int_{\{\|x - \hat{x}\|_{\Sigma} \geq r_0\}} \exp\left(-\frac{\delta}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2\right) dx \\ &= \frac{\delta^{d/2}}{\varepsilon^{d/2}(2\pi)^{d/2}} \int_{\{|u| \geq r_0\}} \exp\left(-\frac{\delta}{2\varepsilon}|u|^2\right) du \\ &= \frac{\delta^{d/2}}{\varepsilon^{d/2}(2\pi)^{d/2}} d\kappa_d \int_{r_0}^{\infty} \exp\left(-\frac{\delta}{2\varepsilon}r^2\right) r^{d-1} dr. \end{aligned}$$

We can express this integral in terms of the upper incomplete gamma function by substituting  $s = \delta r^2/2\varepsilon$  (note that  $r'(s) = 2^{-1/2}\varepsilon^{1/2}\delta^{-1/2}s^{-1/2}$ ) as

$$\int_{r_0}^{\infty} e^{-\frac{\delta}{2\varepsilon}r^2} r^{d-1} dr = \int_{\frac{\delta r_0^2}{2\varepsilon}}^{\infty} e^{-s} 2^{\frac{d}{2}-1} \varepsilon^{\frac{d}{2}} \delta^{-\frac{d}{2}} s^{\frac{d}{2}-1} ds = 2^{\frac{d}{2}-1} \varepsilon^{\frac{d}{2}} \delta^{-\frac{d}{2}} \Gamma\left(\frac{d}{2}, \frac{\delta r_0^2}{2\varepsilon}\right).$$

This leads to

$$v(\{\|x - \hat{x}\|_{\Sigma} \geq r_0\}) = \frac{d}{2} \frac{1}{\Gamma(d/2 + 1)} \Gamma\left(\frac{d}{2}, \frac{\delta r_0^2}{2\varepsilon}\right).$$

Now, using the fundamental recurrence  $\Gamma(z + 1) = z\Gamma(z)$  completes the proof.  $\square$

Now, we can prove our far range estimate.

**Proof of Proposition 3.8** Let  $x \in \mathbb{R}^d \setminus U(r_0)$ . We distinguish between two cases. First, consider the case that  $R_2(x) \geq 0$ . For  $t \geq 0$ , the estimate  $|e^{-t} - 1| \leq 1$  holds. This implies

$$\left| \exp\left(-\frac{1}{\varepsilon}R_2(x)\right) - 1 \right| \leq 1.$$

Next, consider the case that  $R_2(x) < 0$ . By Assumption 3.3, we have

$$R_2(x) = I(x) - \frac{1}{2}\|x - \hat{x}\|_{\Sigma}^2 \geq \frac{\delta - 1}{2}\|x - \hat{x}\|_{\Sigma}^2$$

for all  $x \in \mathbb{R}^d$ . For  $t \leq 0$  we have  $|e^{-t} - 1| = e^{-t} - 1 < e^{-t}$ , and thus

$$\left| \exp\left(-\frac{1}{\varepsilon}R_2(x)\right) - 1 \right| \leq \exp\left(-\frac{1}{\varepsilon}R_2(x)\right) \leq \exp\left(-\frac{\delta - 1}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2\right).$$

Together, this shows that

$$\left| \exp\left(-\frac{1}{\varepsilon}R_2(x)\right) - 1 \right| \leq \exp\left(-\frac{\delta - 1}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2\right)$$

for all  $x \in \mathbb{R}^d$ . Now it follows that

$$\left| \exp\left(-\frac{1}{\varepsilon}R_2(x)\right) - 1 \right| \exp\left(-\frac{1}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2\right) \leq \exp\left(-\frac{\delta}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2\right)$$

for all  $x \in \mathbb{R}^d$ . This yields

$$\begin{aligned} & \int_{\mathbb{R}^d \setminus U(r_0)} \left| \exp\left(-\frac{1}{\varepsilon}R_2(x)\right) - 1 \right| \mathcal{L}_{\mu, \nu}(\mathrm{d}x) \\ & \leq \frac{1}{(2\pi\varepsilon)^{d/2}\sqrt{\det \Sigma}} \int_{\mathbb{R}^d \setminus U(r_0)} \exp\left(-\frac{\delta}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2\right) \mathrm{d}x \end{aligned}$$

Now, the proposition follows from

$$\begin{aligned} & \frac{\delta^{d/2}}{(2\pi\varepsilon)^{d/2}\sqrt{\det \Sigma}} \int_{\mathbb{R}^d \setminus U(r)} \exp\left(-\frac{\delta}{2\varepsilon}\|x - \hat{x}\|_{\Sigma}^2\right) \mathrm{d}x \\ & = \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{d}{2}, \frac{\delta r^2}{2\varepsilon}\right) = 1 - \Xi_d\left(\frac{\delta r_0^2}{\varepsilon}\right), \end{aligned}$$

which in turn holds by Lemma 3.12 and the identity  $\Gamma(a, z) = \Gamma(a) - \gamma(a, z)$ .  $\square$

### 3.3 Optimal choice of the parameter

We have the following necessary optimality condition for the parameter  $r_0$  in Theorem 3.4.

**Proposition 3.13** *The optimal choice of  $r_0$  in the error bound (3.2) is either 0 or satisfies*

$$\exp\left(\frac{(1 + \varepsilon)K}{6\varepsilon}r_0^3\right) - 1 - \exp\left(\frac{1 - \delta}{2\varepsilon}r_0^2\right) = 0.$$

**Proof** The terms  $E_1$  and  $E_2$  are differentiable on  $[0, \infty)$ . Clearly, the optimal  $r_0$  is either 0 or satisfies the identity

$$E'(r_0) = E'_1(r_0) + E'_2(r_0) = 0. \tag{3.8}$$

We have that

$$E'_1(r_0) = c_d \varepsilon^{-\frac{d}{2}} f(r_0) r_0^{d-1} = c_d \varepsilon^{-\frac{d}{2}} \left( \exp\left(\frac{(1 + \varepsilon)K}{6\varepsilon}r_0^3\right) - 1 \right) \exp\left(-\frac{1}{2\varepsilon}r_0^2\right) r_0^{d-1}$$

and

$$E'_2(r_0) = -\frac{2\delta^{1-\frac{d}{2}}r_0}{\varepsilon} \mathcal{E}'_d\left(\frac{\delta r_0^2}{\varepsilon}\right) = -\frac{2}{2^{d/2}\Gamma(d/2)} \cdot \frac{r_0^{d-1}}{\varepsilon^{d/2}} \exp\left(-\frac{\delta}{2\varepsilon}r_0^2\right)$$

Identity (3.8) now corresponds to

$$\exp\left(\frac{(1+\varepsilon)K}{6\varepsilon}r_0^3\right) - 1 - \exp\left(\frac{1-\delta}{2\varepsilon}r_0^2\right) = 0$$

which yields the result. □

**Remark 3.14** The right hand side of the far range estimate (3.6) can be written as

$$c_d \varepsilon^{-\frac{d}{2}} \int_{r_0}^{\infty} \exp\left(-\frac{\delta}{2\varepsilon}r^2\right) r^{d-1} dr,$$

where  $c_d$  is defined as in Theorem 3.4. The optimal choice of  $r_0$  is therefore one for which the integrands  $f(r_0)r^{d-1}$  and  $\exp(-\delta r_0^2/2\varepsilon)r^{d-1}$  of the close and the far range estimate take the same value.

### 4 Explicit error estimate

Here, we present a non-asymptotic error estimate in terms of  $K, \delta, \varepsilon,$  and the problem dimension  $d$ . While Theorem 3.4 constitutes a non-asymptotic error estimate and is the sharpest of our three main results, it is not immediately clear how the non-Gaussianity of the likelihood and the prior distribution, as quantified by the constant  $K$  in Assumption 3.2, the noise level, and the problem dimension affect the error bound. The purpose of the following theorem is to make this influence more explicit.

**Theorem 4.1** *Suppose that  $I, \Phi,$  and  $R$  satisfy Assumptions 3.1–3.3. If  $K, \delta, \varepsilon,$  and  $d$  satisfy*

$$\frac{6\delta^{\frac{3}{2}}}{(1+\varepsilon)\varepsilon^{\frac{1}{2}}K} \geq \max \left\{ 8d^{\frac{3}{2}}, \left( 8 \ln \left( \frac{2}{C(1+\varepsilon)\varepsilon^{\frac{1}{2}}K\delta^{\frac{d}{2}}} \cdot \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2} + \frac{3}{2})} \right) \right)^{\frac{3}{2}} \right\} \tag{4.1}$$

with  $C := \sqrt{2}e/3,$  then

$$d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) \leq 2C(1+\varepsilon)\varepsilon^{\frac{1}{2}}K \frac{\Gamma(\frac{d}{2} + \frac{3}{2})}{\Gamma(\frac{d}{2})}. \tag{4.2}$$

**Remark 4.2** Condition (4.1) can be interpreted in the following way. For given  $d$  and  $\delta,$  it imposes an upper bound on the noise level  $\varepsilon^{1/2}$  and  $K,$  whereas for given  $\delta, K,$  and  $\varepsilon,$  it imposes an upper bound on the dimension  $d$ . As  $d \rightarrow \infty,$  the ratio  $\Gamma(\frac{d}{2} + \frac{3}{2})/\Gamma(\frac{d}{2})$  grows in the order of  $d^{3/2},$  see [28, pp. 67–68].



In order to prove this theorem, we introduce an exponential tail estimate for the Laplace approximation, which is a modified version of [25, Prop. 4].

**Lemma 4.3** *Let  $v = \mathcal{N}(\hat{x}, \delta^{-1}\varepsilon\Sigma)$  with  $\delta > 0$ . Then,*

$$v(\mathbb{R}^d \setminus U(r)) = 1 - \mathcal{E}_d\left(\frac{\delta r^2}{\varepsilon}\right) \leq 2 \exp\left(-\frac{\delta r^2}{8\varepsilon}\right)$$

holds for all  $r \geq 2(d\varepsilon/\delta)^{1/2}$ .

**Proof** By Lemma 3.12, we have

$$v(\mathbb{R}^d \setminus U(r_0)) = \frac{1}{\Gamma(d/2)} \Gamma\left(\frac{d}{2}, \frac{\delta r_0^2}{2\varepsilon}\right) = 1 - \mathcal{E}_d\left(\frac{\delta r_0^2}{\varepsilon}\right).$$

Let  $x \sim \mathcal{N}(\hat{x}, \delta^{-1}\varepsilon\Sigma)$ . Then  $u := \Sigma^{-1/2}(x - \hat{x}) \sim \mathcal{N}(0, \delta^{-1}\varepsilon I_d)$ . The concentration inequality for Gaussian measures yields

$$\mathbb{P}(|u| > s + \mathbb{E}[|u|]) \leq \mathbb{P}(|u| - \mathbb{E}[|u|] > s) \leq 2 \exp\left(-\frac{s^2}{2\sigma^2}\right) \quad \text{for all } s > 0,$$

where  $\sigma := \sup_{|z| \leq 1} \mathbb{E}[|(z, u)|^2]$ , see [13, Chapter 3]. Now,

$$\sigma^2 = \lambda_{\max}(\delta^{-1}\varepsilon I_d) = \delta^{-1}\varepsilon,$$

and

$$\mathbb{E}[|u|] \leq \mathbb{E}[|u|^2]^{1/2} = \delta^{-1/2}\varepsilon^{1/2} \text{tr}(I_d)^{1/2} = \delta^{-1/2}\varepsilon^{1/2}d^{1/2}.$$

By choosing  $s = \delta^{-1/2}\varepsilon^{1/2}d^{1/2}$  and using that  $r \geq 2s$ , we obtain

$$\mu^y(\mathbb{R}^d \setminus U(r)) = \mathbb{P}(|u| > r) \leq \mathbb{P}\left(|u| > \frac{r}{2} + \mathbb{E}[|u|]\right) \leq 2 \exp\left(-\frac{\delta r^2}{8\varepsilon}\right).$$

□

**Proof of Theorem 4.1** We choose

$$r_0 := \left(\frac{6\varepsilon}{(1+\varepsilon)K}\right)^{1/3}.$$

According to Theorem 3.4, we then have

$$d_{\text{TV}}(\mu^y, \mathcal{L}_{\mu^y}) \leq E_1(r_0; K, \varepsilon, d) + E_2(r_0; \delta, \varepsilon, d).$$

For all  $t \geq 0$ , the exponential function satisfies the estimate

$$\exp(t) - 1 = (1 - \exp(-t)) \exp(t) \leq t \exp(t).$$

Therefore, we have

$$\begin{aligned} E_1(r_0; K, \varepsilon, d) &= \frac{2}{\Gamma\left(\frac{d}{2}\right)} (2\varepsilon)^{-\frac{d}{2}} \int_0^{r_0} \left( \exp\left(\frac{(1+\varepsilon)K}{6\varepsilon} r^3\right) - 1 \right) \exp\left(-\frac{1}{2\varepsilon} r^2\right) r^{d-1} dr \\ &\leq \frac{2}{\Gamma\left(\frac{d}{2}\right)} (2\varepsilon)^{-\frac{d}{2}-1} \frac{(1+\varepsilon)K}{3} \int_0^{r_0} \exp\left(\frac{(1+\varepsilon)K}{6\varepsilon} r^3 - \frac{1}{2\varepsilon} r^2\right) r^{d+2} dr. \end{aligned}$$

By the choice of  $r_0$ , we have

$$\frac{(1+\varepsilon)K}{6\varepsilon} r^3 \leq 1 \quad \text{for all } r \in [0, r_0],$$

so that the integral is bounded by

$$e \int_0^{r_0} \exp\left(-\frac{1}{2\varepsilon} r^2\right) r^{d+2} dr.$$

By substituting  $s = r^2/2\varepsilon$ , we can in turn express this integral as

$$\frac{1}{2} (2\varepsilon)^{\frac{d}{2}+\frac{3}{2}} \int_0^{\frac{r_0^2}{2\varepsilon}} e^{-s} s^{\frac{d}{2}+\frac{1}{2}} ds = \frac{1}{2} (2\varepsilon)^{\frac{d}{2}+\frac{3}{2}} \gamma\left(\frac{d}{2} + \frac{3}{2}, \frac{r_0^2}{2\varepsilon}\right).$$

Now, we use the inequality  $\gamma(a, z) \leq \Gamma(a)$  to obtain that

$$E_1(r_0; K, \varepsilon, d) \leq \frac{2^{\frac{1}{2}} e}{3} (1+\varepsilon) \varepsilon^{\frac{1}{2}} K \frac{\Gamma\left(\frac{d}{2} + \frac{3}{2}\right)}{\Gamma\left(\frac{d}{2}\right)}.$$

By condition (4.1), we have

$$r_0 = \left(\frac{6\varepsilon}{(1+\varepsilon)K}\right)^{\frac{1}{3}} \geq 2 \left(\frac{d\varepsilon}{\delta}\right)^{\frac{1}{2}}.$$

Thus, we may apply Lemma 4.3, which yields

$$\begin{aligned} E_2(r_0; \delta, \varepsilon, d) &= \delta^{-\frac{d}{2}} \left(1 - \mathcal{E}\left(-\frac{\delta r_0^2}{\varepsilon}\right)\right) \leq 2\delta^{-\frac{d}{2}} \exp\left(-\frac{\delta r_0^2}{8\varepsilon}\right) \\ &= 2\delta^{-\frac{d}{2}} \exp\left(-\frac{1}{8} \left(\frac{6\delta^{\frac{3}{2}}}{(1+\varepsilon)\varepsilon^{\frac{1}{2}} K}\right)^{\frac{2}{3}}\right) \leq C(1+\varepsilon) \varepsilon^{\frac{1}{2}} K \frac{\Gamma\left(\frac{d}{2} + \frac{3}{2}\right)}{\Gamma\left(\frac{d}{2}\right)} \end{aligned}$$

by condition (4.1). Now, we obtain by summing up that

$$d_{TV}(\mu^y, \mathcal{L}_{\mu^y}) \leq 2C(1 + \varepsilon)\varepsilon^{\frac{1}{2}}K \frac{\Gamma(\frac{d}{2} + \frac{3}{2})}{\Gamma(\frac{d}{2})}.$$

□

### 4.1 Asymptotic behavior for fixed and increasing problem dimension

Now, we describe the convergence of the Laplace approximation for a sequence of nonlinear problems that satisfy Assumptions 3.1–3.3 with varying bounds  $\{K_n\}_{n \in \mathbb{N}}$  and  $\{\delta_n\}_{n \in \mathbb{N}}$ , respectively, and varying squared noise levels  $\{\varepsilon_n\}_{n \in \mathbb{N}}$ , both in case of a fixed and an increasing problem dimension. We denote the data by  $y_n$ , the prior distribution by  $R_n$ , the scaled negative log-likelihood by  $\Phi_n$ , and set  $I_n(x) = \Phi_n(x) + \varepsilon_n R_n(x)$ .

First, we consider the case that the problem dimension  $d$  remains constant.

**Corollary 4.4** (Fixed problem dimension) *Suppose that  $I_n, \Phi_n$ , and  $R_n$  satisfy Assumptions 3.1–3.3. If  $\varepsilon_n^{1/2}K_n \rightarrow 0$  and if there exist  $\underline{\delta} > 0$  and  $N_0 \in \mathbb{N}$  such that*

$$\delta_n \geq \underline{\delta} \quad \text{and} \quad \varepsilon_n \leq 1$$

for all  $n \geq N_0$ , then there exist  $C = C(d) > 0$  and  $N_1 \geq N_0$  such that

$$d_{TV}(\mu^{y_n}, \mathcal{L}_{\mu^{y_n}}) \leq C\varepsilon_n^{\frac{1}{2}}K_n$$

for all  $n \geq N_1$ .

**Proof** Since  $\{\delta_n\}_{n \in \mathbb{N}}$  is bounded from below and  $\{\varepsilon_n\}_{n \in \mathbb{N}}$  is bounded from above, the left hand side of (4.1) is bounded from below by  $C_1/\varepsilon_n^{1/2}K_n$  for some  $C_1 > 0$ . On the other hand, the right hand side of (4.1) is bounded from above by  $(8 \ln C_2/\varepsilon_n^{1/2}K_n)^{2/3}$  for large enough  $n$  and some  $C_2 > 0$ , since  $\{\delta_n\}_{n \in \mathbb{N}}$  is bounded from below and  $\{\varepsilon_n\}_{n \in \mathbb{N}}$  is bounded from below by 0. Consequently, there exists  $N_1 \geq N_0$  such that condition (4.1) holds for all  $n \geq N_1$  by the convergence  $\varepsilon_n^{1/2}K_n \rightarrow 0$  and since  $\lim_{t \rightarrow \infty} t^{-2/3} \ln t = 0$ . Now, Theorem 4.1 yields the proposition. □

**Remark 4.5** Corollary 4.4 covers two cases of particular interest: That of  $K_n \rightarrow 0$  while  $\varepsilon_n = \varepsilon$  remains constant, which yields a rate of  $K_n$ , and that of  $\varepsilon_n \rightarrow 0$  while  $K_n = K$  remains constant, which yields a rate of  $\varepsilon_n^{1/2}$ . The former case can, for example, occur if the sequence of forward mappings  $G_n$  converges pointwise towards a linear mapping, see Sect. 5. The convergence rate in the latter case, i.e., in the small noise limit, agrees with the rate established in [25, Theorem 2] if we set  $\varepsilon_n = \frac{1}{n}$ .

Now, we consider the case of an increasing problem dimension  $d \rightarrow \infty$ . To this end, we index  $K_d, \delta_d, \varepsilon_d$ , and  $R_d$  by  $d \in \mathbb{N}$ .

**Corollary 4.6** (Increasing problem dimension) *Suppose that  $I_d, \Phi_d,$  and  $R_d$  satisfy Assumptions 3.1–3.3 and that  $\varepsilon_d^{1/2} K_d \rightarrow 0$ . If there exists  $N_0 \in \mathbb{N}$  such that  $\delta_d \leq e^{-1/2}, \varepsilon_d \leq 1,$  and*

$$\frac{3}{\varepsilon_d^{1/2} K_d} \geq \left( \frac{8}{\delta_d} \ln \frac{1}{\delta_d} \right)^{3/2} d^{3/2} \tag{4.3}$$

for all  $d \geq N_0,$  then for every  $C > 2\sqrt{2}e/3,$  there exists  $N_1 \geq N_0$  such that

$$d_{TV}(\mu^{y_d}, \mathcal{L}_{\mu^{y_d}}) \leq C \varepsilon_d^{1/2} K_d d^{3/2}$$

for all  $d \geq N_1.$

**Proof** We can write condition (4.1) as

$$\left( \frac{6}{(1 + \varepsilon_d)\varepsilon_d^{1/2} K_d} \right)^{2/3} \geq \frac{4d}{\delta_d} \quad \text{and} \tag{4.4}$$

$$\left( \frac{6}{(1 + \varepsilon_d)\varepsilon_d^{1/2} K_d} \right)^{2/3} \geq \frac{8}{\delta_d} \ln \left( \frac{2}{C(1 + \varepsilon_d)\varepsilon_d^{1/2} K_d} \cdot \frac{\Gamma(\frac{d}{2})}{\Gamma(\frac{d}{2} + \frac{3}{2})} \right) + \frac{4d}{\delta_d} \ln \frac{1}{\delta_d}. \tag{4.5}$$

By [28, pp. 67–68], we have

$$\lim_{d \rightarrow \infty} \frac{\Gamma(\frac{d}{2} + \frac{3}{2})}{\Gamma(\frac{d}{2})} \left( \frac{d}{2} \right)^{-3/2} = 1, \tag{4.6}$$

so that the first term on the right hand side of (4.5) is bounded from above by

$$8e^{1/2} \ln \frac{C_1}{\varepsilon_d^{1/2} K_d d^{3/2}} \leq 8e^{1/2} \ln \frac{C_1}{\varepsilon_d^{1/2} K_d}$$

for some  $C_1 > 0,$  which in turn is bounded from above by

$$\frac{1}{2} \left( \frac{3}{\varepsilon_d^{1/2} K_d} \right)^{2/3} \leq \frac{1}{2} \left( \frac{6}{(1 + \varepsilon_d)\varepsilon_d^{1/2} K_d} \right)^{2/3} \tag{4.7}$$

for large enough  $d,$  due to the convergence  $\varepsilon_d^{1/2} K_d \rightarrow 0,$  the boundedness from above of  $\{\varepsilon_n\}_{n \in \mathbb{N}},$  and since  $\lim_{t \rightarrow \infty} t^{-2/3} \ln t = 0.$  By (4.3), the second term on the right hand side of (4.5) is bounded from above by (4.7) for all  $d \geq N_0$  as well. Due to the assumption  $\delta_d \leq e^{-1/2}$  and (4.7), condition (4.3) also ensures that (4.4) is satisfied for all  $d \geq N_0.$  Therefore, condition (4.1) is satisfied for large enough  $d.$

Now, Theorem 4.1 and (4.6) yield that for every  $C > 2\sqrt{2}e/3$  there exists  $N_1 \geq N_0$  such that

$$d_{\text{TV}}(\mu^{y_d}, \mathcal{L}_{\mu^{y_d}}) \leq C\varepsilon_d^{\frac{1}{2}} K_d d^{\frac{3}{2}}$$

for all  $d \geq N_1$ . □

### 5 Perturbed linear problems with Gaussian prior

In this section we consider the case that the forward mapping  $G$  is given by a linear mapping with a small nonlinear perturbation, i.e., that

$$G_\tau(x) = Ax + \tau F(x), \tag{5.1}$$

with  $A \in \mathbb{R}^{d \times d}$ ,  $F \in C^3(\mathbb{R}^d)$ , and  $\tau \geq 0$ . We then quantify the error of the Laplace approximation for small  $\tau$ , that is, when the nonlinearity of the forward mapping is small, and for fixed problem dimension  $d$ . In order to isolate the effect of the nonlinearity on estimate (4.2), we consider the case when not only the noise, but also the prior distribution is Gaussian. This ensures that all non-Gaussianity in the posterior distribution results from the nonlinearity of  $G_\tau$ .

We assign a prior distribution  $\mu = \mathcal{N}(m_0, \Sigma_0)$  with  $m_0 \in \mathbb{R}^d$  and symmetric, positive definite  $\Sigma_0 \in \mathbb{R}^{d \times d}$ , i.e., we set

$$\begin{aligned} R(x) &:= -\ln \left( \frac{1}{(2\pi)^{d/2} \sqrt{\det \Sigma_0}} \exp \left( -\frac{1}{2} \|x - m_0\|_{\Sigma_0}^2 \right) \right) \\ &= \frac{1}{2} \|x - m_0\|_{\Sigma_0}^2 + \frac{d}{2} \ln 2\pi + \frac{1}{2} \ln \det \Sigma_0. \end{aligned} \tag{5.2}$$

For each  $\tau \geq 0$  we denote the data by  $y_\tau$  and the scaled negative log-likelihood by  $\Phi_\tau(x) = \frac{1}{2} |G_\tau(x) - y_\tau|^2$ .

We make the following assumptions on the function  $I_\tau$  and the perturbation  $F$ . Let  $B(r) \subset \mathbb{R}^d$  denote the closed Euclidean ball with radius  $r$  around the origin.

**Assumption 5.1** We assume that there exists  $\tau_0 > 0$  such that for all  $\tau \in [0, \tau_0]$ ,

$$I_\tau(x) = \frac{1}{2} |Ax + \tau F(x) - y_\tau|^2 + \frac{\varepsilon}{2} \|x - m_0\|_{\Sigma_0}^2 + c_\tau$$

has a unique minimizer  $\hat{x}_\tau$  with  $D^2 I_\tau(\hat{x}_\tau) > 0$ . Furthermore, we assume that  $y_\tau$ ,  $\hat{x}_\tau$  and  $\Sigma_\tau := D^2 I_\tau(\hat{x}_\tau)^{-1}$  converge as  $\tau \rightarrow 0$  with  $\lim_{\tau \rightarrow 0} \Sigma_\tau > 0$  and denote their limits by  $y$ ,  $\hat{x}$ , and  $\Sigma$ , respectively.

**Assumption 5.2** There exist constants  $C_0, \dots, C_3 > 0$  and  $\tau_0 > 0$  such that

$$\|D^j F(x)\|_{\Sigma_\tau} \leq C_j, \quad j = 0, \dots, 3,$$

for all  $x \in \mathbb{R}^d$  and  $\tau \in [0, \tau_0]$ , and there exists  $M > 0$  such that

$$D^3 F \equiv 0 \quad \text{on } \mathbb{R}^d \setminus B(M).$$

The idea behind the following theorem is to make explicit how the nonlinearity of the forward mapping, as quantified by  $\tau$  and the constants  $C_0, \dots, C_3, M$  in Assumption 5.2, influences the total variation error bound of Theorem 4.1.

**Theorem 5.3** *Suppose that Assumptions 5.1 and 5.2 hold. Then, there exists  $\tau_1 \in (0, \tau_0]$  such that*

$$d_{TV}(\mu^{y_\tau}, \mathcal{L}_{\mu^{y_\tau}}) \leq C(1 + \varepsilon)\varepsilon^{\frac{1}{2}} \left( V(\tau)\tau + \frac{W}{2}\tau^2 \right)$$

for all  $\tau \in [0, \tau_1]$ , where

$$C := \frac{2}{3}\sqrt{2}e \frac{\Gamma(\frac{d}{2} + \frac{3}{2})}{\Gamma(\frac{d}{2})},$$

$$V(\tau) := C_3 (\|A\|M + |y_\tau|) + 3C_2 \left\| A \Sigma_\tau^{\frac{1}{2}} \right\|,$$

$$W := C_3 C_0 + 3C_2 C_1.$$

Moreover,  $\{V(\tau)\}_{\tau \in [0, \tau_1]}$  is bounded.

- Remark 5.4** 1. The choice of the upper bound  $\tau_1$  is made explicit in the proof of Theorem 5.3 and depends on  $d$  and  $\varepsilon$ , i.a., through  $\delta_0$  as defined in Proposition 5.6. The proof of Theorem 5.3 can be adapted to yield a result analogous to Corollary 4.6 in the case when the problem dimension  $d$  tends to  $\infty$  while the size  $\tau_d$  of the perturbation tends to 0. Then,  $\delta_{\tau_d}$  may converge to 0, and (4.3) imposes a bound on the rate at which  $\{\tau_d\}_{d \in \mathbb{N}}$  tends to 0.
2. By the boundedness and continuity of  $F, G_\tau$   $\Gamma$ -converges toward  $A$ . By the fundamental theorem of  $\Gamma$ -convergence and Assumption 5.1, this, in turn, implies that  $\hat{x}$  is the minimizer of

$$I(x) = \frac{1}{2}|Ax - y|^2 + \frac{\varepsilon}{2}\|x - m_0\|_{\Sigma_0}^2 + c$$

and that  $\Sigma = D^2 I(\hat{x})^{-1}$ .

3. Theorem 5.3 remains valid if the assumption that  $D^3 F \equiv 0$  outside of a bounded set is replaced by

$$\|D^3 F(x)\|_{\Sigma_\tau} \leq C_3 M|x|^{-1}$$

for all  $x \in \mathbb{R} \setminus \{0\}$  and  $\tau \in [0, \tau_0]$ .

In order to prove Theorem 5.3, we first show that Assumptions 3.2 and 3.3 are satisfied for small enough  $\tau$  and determine the bounds  $K_\tau$  and  $\delta_\tau$ . Then, we derive the error estimate for the perturbed linear case.

### 5.1 Verifying Assumption 3.3

We verify that Assumption 3.3 holds for small enough  $\tau$  and determine  $\delta_\tau$ . First, we estimate  $I_\tau$  from below.

**Lemma 5.5** *For all  $\tau \geq 0$  and  $x \in \mathbb{R}^d$ , we have*

$$I_\tau(x) \geq \frac{1}{2}|G_\tau(x) - G_\tau(\hat{x}_\tau)|^2 + \frac{\varepsilon}{2}\|x - \hat{x}_\tau\|_{\Sigma_0}^2 - |G_\tau(x) - G_\tau(\hat{x}_\tau) - DG_\tau(\hat{x}_\tau)(x - \hat{x}_\tau)| \cdot |G_\tau(\hat{x}_\tau) - y_\tau|.$$

**Proof** Since  $\hat{x}_\tau$  satisfies the necessary optimality condition

$$D\Phi_\tau(\hat{x}_\tau) + \varepsilon DR(\hat{x}_\tau) = DI_\tau(\hat{x}_\tau) = 0,$$

we can write  $I_\tau(x)$  as

$$I_\tau(x) = \Phi_\tau(x) - \Phi_\tau(\hat{x}_\tau) - D\Phi_\tau(\hat{x}_\tau)(x - \hat{x}_\tau) + \varepsilon\left(R(x) - R(\hat{x}_\tau) - DR(\hat{x}_\tau)(x - \hat{x}_\tau)\right)$$

for all  $x \in \mathbb{R}^d$ . For the log-likelihood, we have

$$\Phi_\tau(x) - \Phi_\tau(\hat{x}_\tau) = \frac{1}{2}|G_\tau(x) - G_\tau(\hat{x}_\tau)|^2 + (G_\tau(x) - G_\tau(\hat{x}_\tau), G_\tau(\hat{x}_\tau) - y_\tau),$$

and

$$D\Phi_\tau(\hat{x}_\tau)(x - \hat{x}_\tau) = (DG_\tau(\hat{x}_\tau)(x - \hat{x}_\tau), G_\tau(\hat{x}_\tau) - y_\tau)$$

for all  $x \in \mathbb{R}^d$ . From this, we obtain

$$\begin{aligned} &\Phi_\tau(x) - \Phi_\tau(\hat{x}_\tau) - D\Phi_\tau(\hat{x}_\tau)(x - \hat{x}_\tau) \\ &\geq \frac{1}{2}|G_\tau(x) - G_\tau(\hat{x}_\tau)|^2 - |G_\tau(x) - G_\tau(\hat{x}_\tau) - DG_\tau(\hat{x}_\tau)(x - \hat{x}_\tau)| \cdot |G_\tau(\hat{x}_\tau) - y_\tau| \end{aligned} \tag{5.3}$$

using the Cauchy–Schwarz inequality. For the log-prior density, we have

$$DR(\hat{x}_\tau)(x - \hat{x}_\tau) = (\hat{x}_\tau - m_0, x - m_0)_{\Sigma_0} - \|\hat{x}_\tau - m_0\|_{\Sigma_0}^2$$

for all  $x \in \mathbb{R}^d$ , and thus

$$R(x) - R(\hat{x}_\tau) - DR(\hat{x}_\tau)(x - \hat{x}_\tau) = \frac{1}{2}\|x - \hat{x}_\tau\|_{\Sigma_0}^2. \tag{5.4}$$

Now, adding up (5.3) and (5.4) multiplied with  $\varepsilon$  yields the proposition. □

**Proposition 5.6** *Suppose that Assumption 5.2 holds. Then there exists  $\tau_0 > 0$ , such that  $I_\tau$  satisfies*

$$I_\tau(x) \geq \frac{\delta_\tau}{2} \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2$$

for all  $x \in \mathbb{R}^d$  and  $\tau \in [0, \tau_0]$ , where

$$\delta_\tau := \gamma_1(\tau) - \gamma_2(\tau) |G_\tau(\hat{x}_\tau) - y_\tau| > 0$$

with

$$\gamma_1(\tau) := \frac{1}{\left\| \Sigma_\tau^{-\frac{1}{2}} (A^T A + \varepsilon \Sigma_0^{-1})^{-\frac{1}{2}} \right\|^2} - C_1^2 \tau^2 \quad \text{and} \quad \gamma_2(\tau) := C_2 \tau$$

for all  $\tau \in [0, \tau_0]$ . Furthermore,  $\lim_{\tau \rightarrow 0} \delta_\tau > 0$ .

**Proof** Let  $x \in \mathbb{R}^d$  be arbitrary. Then, there exists  $z_1 \in \mathbb{R}^d$  such that  $F(x) - F(\hat{x}_\tau) = DF(z_1)(x - \hat{x}_\tau)$ . Therefore,

$$|F(x) - F(\hat{x}_\tau)| \leq \|DF(z_1)\|_{\Sigma_\tau} \|x - \hat{x}_\tau\|_{\Sigma_\tau} \leq C_1 \|x - \hat{x}_\tau\|$$

by Assumption 5.2. Moreover, we have

$$\|x - \hat{x}_\tau\|_{\Sigma_\tau} \leq \left\| \Sigma_\tau^{-\frac{1}{2}} (A^T A + \varepsilon \Sigma_0^{-1})^{-\frac{1}{2}} \right\| \cdot \left| (A^T A + \varepsilon \Sigma_0^{-1})^{\frac{1}{2}} (x - \hat{x}_\tau) \right|,$$

and hence

$$\begin{aligned} & \frac{1}{2} |G_\tau(x) - G_\tau(\hat{x}_\tau)|^2 + \frac{\varepsilon}{2} \|x - \hat{x}_\tau\|_{\Sigma_0}^2 \\ & \geq \frac{1}{2} |A(x - \hat{x}_\tau)|^2 - \frac{\tau^2}{2} |F(x) - F(\hat{x}_\tau)|^2 + \frac{\varepsilon}{2} \|x - \hat{x}_\tau\|_{\Sigma_0}^2 \\ & \geq \frac{1}{2} \left( (A^T A + \varepsilon \Sigma_0^{-1}) (x - \hat{x}_\tau), x - \hat{x}_\tau \right) - \frac{\tau^2 C_1^2}{2} \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2 \\ & \geq \frac{1}{2} \gamma_1(\tau) \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2 \end{aligned}$$

for all  $\tau \leq \tau_0$ . There also exists  $z_2 \in \mathbb{R}^d$  such that

$$G_\tau(x) - G_\tau(\hat{x}_\tau) - DG_\tau(\hat{x}_\tau)(x - \hat{x}_\tau) = \frac{1}{2} D^2 G_\tau(z_2)(x - \hat{x}_\tau, x - \hat{x}_\tau),$$



and  $D^2G_\tau = \tau D^2F$ . By Assumption 5.2, we thus have

$$\begin{aligned} |G_\tau(x) - G_\tau(\hat{x}_\tau) - DG_\tau(\hat{x}_\tau)(x - \hat{x}_\tau)| &\leq \frac{\tau}{2} \|D^2F(z_2)\|_{\Sigma_\tau} \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2 \\ &\leq \frac{\tau C_2}{2} \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2 = \frac{1}{2} \gamma_2(\tau) \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2 \end{aligned}$$

for all  $\tau \leq \tau_0$ . Now, Lemma 5.5 yields that

$$I_\tau(x) \geq \frac{1}{2} \gamma_1(\tau) \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2 - \frac{1}{2} \gamma_2(\tau) \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2 |G_\tau(\hat{x}_\tau) - y_\tau| = \frac{\delta_\tau}{2} \|x - \hat{x}_\tau\|_{\Sigma_\tau}^2.$$

It remains to show that  $\lim_{\tau \rightarrow 0} \delta_\tau > 0$ . The convergence of  $y_\tau$  and  $\hat{x}_\tau$  yields

$$G_\tau(\hat{x}_\tau) - y_\tau \rightarrow A\hat{x} - y.$$

Now, it follows from  $\lim_{\tau \rightarrow 0} \gamma_2(\tau) = 0$  and the convergence of  $\Sigma_\tau$  that

$$\lim_{\tau \rightarrow 0} \delta_\tau = \lim_{\tau \rightarrow 0} \gamma_1(\tau) = \frac{1}{\left\| \Sigma^{-\frac{1}{2}} (A^T A + \varepsilon \Sigma_0^{-1})^{-\frac{1}{2}} \right\|^2} > 0.$$

□

### 5.2 Verifying Assumption 3.2

Now, we verify that Assumption 3.2 holds for small  $\tau$ . The following proposition also describes how the nonlinearity of the forward mapping translates into non-Gaussianity of the likelihood, as quantified by the constant  $K_\tau$ .

**Proposition 5.7** *Suppose that Assumption 5.2 holds. Then  $\Phi_\tau$  and  $R$  satisfy Assumption 3.2 for all  $\tau \in [0, \tau_0]$  with*

$$K_\tau := \tau \left( C_3 (\|A\|M + |y_\tau|) + 3C_2 \left\| A \Sigma_\tau^{\frac{1}{2}} \right\| \right) + \frac{\tau^2}{2} (C_3 C_0 + 3C_2 C_1).$$

**Proof** We express the scaled negative log-likelihood for all  $x \in \mathbb{R}^d$  as

$$\begin{aligned} \Phi_\tau(x) &= \frac{1}{2} |Ax + \tau F(x) - y_\tau|^2 \\ &= \frac{1}{2} |Ax - y_\tau|^2 + \tau (Ax - y_\tau, F(x)) + \frac{\tau^2}{2} |F(x)|^2 \\ &= \Phi_0(x) + \tau \Psi_1(x; \tau) + \tau^2 \Psi_2(x), \end{aligned}$$

where  $\Psi_1(x; \tau) := (Ax - y_\tau, F(x))$  and  $\Psi_2(x) := \frac{1}{2}|F(x)|^2$ . For the first term, we have  $D^3\Phi_0(x) = 0$  for all  $x \in \mathbb{R}^d$  due to the linearity of  $A$ . The third differentials of  $\Psi_1, \Psi_2$  can be stated explicitly as

$$D^3\Psi_1(x; \tau)(h_1, h_2, h_3) = (D^3F(x)(h_1, h_2, h_3), Ax - y_\tau) + (D^2F(x)(h_1, h_2), Ah_3) + (D^2F(x)(h_2, h_3), Ah_1) + (D^2F(x)(h_1, h_3), Ah_2),$$

and

$$D^3\Psi_2(x)(h_1, h_2, h_3) = \frac{1}{2}(D^3F(x)(h_1, h_2, h_3), F(x)) + \frac{1}{2}(D^2F(x)(h_1, h_2), DF(x)h_3) + \frac{1}{2}(D^2F(x)(h_2, h_3), DF(x)h_1) + \frac{1}{2}(D^2F(x)(h_1, h_3), DF(x)h_2)$$

for all  $x, h_1, h_2, h_3 \in \mathbb{R}^d$  and  $\tau \geq 0$ . Therefore, we have

$$\begin{aligned} \|D^3\Psi_1(x; \tau)\|_{\Sigma_\tau} &\leq \|D^3F(x)\|_{\Sigma_\tau}|Ax - y_\tau| + 3\|D^2F(x)\|_{\Sigma_\tau} \left\| A\Sigma_\tau^{\frac{1}{2}} \right\| \\ &\leq C_3 (\|A\|M + |y_\tau|) + 3C_2 \left\| A\Sigma_\tau^{\frac{1}{2}} \right\| \end{aligned}$$

for all  $x \in \mathbb{R}^d$  and  $\tau \leq \tau_0$  by Assumption 5.2. Moreover, we obtain

$$\begin{aligned} \|D^3\Psi_2(x)\|_{\Sigma_\tau} &\leq \frac{1}{2}\|D^3F(x)\|_{\Sigma_\tau}|F(x)| + \frac{3}{2}\|D^2F(x)\|_{\Sigma_\tau}\|DF(x)\|_{\Sigma_\tau} \\ &\leq \frac{1}{2}C_3C_0 + \frac{3}{2}C_2C_1 \end{aligned}$$

for all  $x \in \mathbb{R}^d$  and  $\tau \leq \tau_0$ . Now, it follows that

$$\begin{aligned} \|D^3\Phi_\tau(x)\|_{\Sigma_\tau} &\leq \tau\|D^3\Psi_1(x; \tau)\|_{\Sigma_\tau} + \tau^2\|D^3\Psi_2(x)\|_{\Sigma_\tau} \\ &\leq \tau \left( C_3 (\|A\|M + |y_\tau|) + 3C_2 \left\| A\Sigma_\tau^{\frac{1}{2}} \right\| \right) + \frac{\tau^2}{2} (C_3C_0 + 3C_2C_1) \end{aligned}$$

for all  $x \in \mathbb{R}^d$  and  $\tau \leq \tau_0$ . □

### 5.3 Proof of Theorem 5.3

**Proof of Theorem 5.3** First of all, we note that Assumption 3.1 holds for all  $\tau \leq \tau_0$  by definition of  $G_\tau, R$  and by Assumption 5.1. Second of all, we note that Assumption 3.3 holds for all  $\tau \leq \tau_0$  by Proposition 5.6 with  $\delta_\tau$  as defined in Proposition 5.6, and that  $\delta_0 = \lim_{\tau \rightarrow 0} \delta_\tau > 0$ . This allows us to choose  $\tau_1 \leq \tau_0$  such that  $\delta_\tau \geq \frac{1}{2}\delta_0 =: \underline{\delta}$  for all  $\tau \in [0, \tau_1]$ . Third of all, we note that Assumption 3.2 holds for all  $\tau \leq \tau_0$  by Proposition 5.7 with  $K_\tau$  as defined in Proposition 5.7, and that  $\lim_{\tau \rightarrow 0} K_\tau = 0$ .

Since  $\{\delta_\tau\}_{\tau \in [0, \tau_1]}$  is bounded from below, the left hand side of condition (4.1) from Theorem 4.1 is bounded from below by  $\kappa_1/K_\tau$  for some  $\kappa_1 > 0$ , and the right hand side of condition (4.1) is bounded from above by  $(8 \ln \kappa_2/K_\tau)^{3/2}$  for small enough  $\tau$  and some  $\kappa_2 > 0$ . Therefore, we can choose  $\tau_2 \leq \tau_1$  such that condition (4.1) is satisfied for all  $\tau \in [0, \tau_2]$ . Now, Theorem 4.1 yields that

$$d_{TV}(\mu^{y_\tau}, \mathcal{L}_{\mu^{y_\tau}}) \leq \frac{2}{3} \sqrt{2\ell}(1 + \varepsilon) \varepsilon^{\frac{1}{2}} \frac{\Gamma(\frac{d}{2} + \frac{3}{2})}{\Gamma(\frac{d}{2})} \cdot \left[ \left( C_3 (\|A\|M + |y_t|) + 3C_2 \left\| A \Sigma_\tau^{\frac{1}{2}} \right\| \right) \tau + \frac{1}{2} (C_3 C_0 + 3C_2 C_1) \tau^2 \right]$$

for all  $\tau \in [0, \tau_2]$ . By the convergence of  $y_\tau$  and  $\Sigma_\tau$ , we can, moreover, choose  $\tau_3 \leq \tau_2$  such that both  $\{|y_\tau|\}_{\tau \in [0, \tau_3]}$  and  $\{\|A \Sigma_\tau^{1/2}\|\}_{\tau \in [0, \tau_3]}$  are bounded. □

## 6 Outlook

In this paper we prove novel error estimates for the Laplace approximation when applied to nonlinear Bayesian inverse problems. Here, the error is measured in TV distance and our estimates aim to quantify effects independent of the noise level. Our central error estimate in Theorem 3.4 is of particular use for high-dimensional problems because it can be evaluated without integrating in  $\mathbb{R}^d$ . Our estimate in Theorem 4.1 makes the influence of the noise level, the nonlinearity of the forward operator, and the problem dimension explicit. Our estimate for perturbed linear problems in Theorem 5.3, in turn, specifies in more detail how the properties of the nonlinear perturbation affect the approximation error.

We point out that our central estimate diverges with an increasing dimension for a fixed noise level and forward mapping, and therefore such asymptotics does not provide any added value compared to the trivial TV upper bound of 1. This unsatisfactory observation is natural since the limiting posterior and Laplace approximation (if well-defined) are singular with respect to each other and, consequently, the TV distance is maximized. Future study is therefore needed to establish similar bounds with distances that metrize the weak convergence such as the 1-Wasserstein distance. Such effort would be aligned with recent developments in BvM theory that extend to nonparametric Bayesian inference and, in particular, Bayesian inverse problems.

**Funding** Open Access funding enabled and organized by Projekt DEAL. The work of the authors was supported by the Academy of Finland (decision 326961).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Alexanderian, A., Petra, N., Stadler, G., Ghattas, O.: A fast and scalable method for a-optimal design of experiments for infinite-dimensional Bayesian nonlinear inverse problems. *SIAM J. Sci. Comput.* **38**(1), A243–A272 (2016)
2. Beck, J., Dia, B.M., Espath, L.F., Long, Q., Tempone, R.: Fast Bayesian experimental design: Laplace-based importance sampling for the expected information gain. *Comput. Methods Appl. Mech. Eng.* **334**, 523–553 (2018)
3. Bishop, C.M.: *Pattern Recognition and Machine Learning*. Springer, Berlin (2006)
4. Chen, P., Villa, U., Ghattas, O.: Hessian-based adaptive sparse quadrature for infinite-dimensional Bayesian inverse problems. *Comput. Methods Appl. Mech. Eng.* **327**, 147–172 (2017)
5. Gibbs, A., Su, F.: On choosing and bounding probability metrics. *Int. Stat. Rev.* **70**, 419–435 (2002). <https://doi.org/10.1111/j.1751-5823.2002.tb00178.x>
6. Giordano, M., Kekkonen, H.: Bernstein-von Mises theorems and uncertainty quantification for linear inverse problems. *SIAM/ASA J. Uncertain. Quantif.* **8**(1), 342–373 (2020). <https://doi.org/10.1137/18M1226269>
7. Inglot, T., Majerski, P.: Simple upper and lower bounds for the multivariate Laplace approximation. *J. Approx. Theory* **186**, 1–11 (2014). <https://doi.org/10.1016/j.jat.2014.06.011>
8. Kaipio, J., Somersalo, E.: *Statistical and Computational Inverse Problems*, vol. 160. Springer, Berlin (2006). <https://doi.org/10.1007/b138659>
9. Kraft, C.: Some conditions for consistency and uniform consistency of statistical procedures. *Univ. Calif. Publ. Stat.* **2**, 125–141 (1955)
10. Łapiński, T.M.: Multivariate Laplace’s approximation with estimated error and application to limit theorems. *J. Approx. Theory* **248**, 105305 (2019). <https://doi.org/10.1016/j.jat.2019.105305>
11. Law, K., Stuart, A., Zygalakis, K.: *Data Assimilation. Texts in Applied Mathematics*, 1st edn. Springer, Berlin (2015). <https://doi.org/10.1007/978-3-319-20325-6>
12. Le Cam, L.: *Asymptotic Methods in Statistical Decision Theory*. Springer, Berlin (2012)
13. Ledoux, M., Talagrand, M.: *Probability in Banach Spaces*. Springer, Berlin (2002). <https://doi.org/10.1007/978-3-642-20212-4>
14. Long, Q., Scavino, M., Tempone, R., Wang, S.: Fast estimation of expected information gains for Bayesian experimental designs based on Laplace approximations. *Comput. Methods Appl. Mech. Eng.* **259**, 24–39 (2013)
15. Lu, Y.: On the Bernstein-von Mises theorem for high dimensional nonlinear Bayesian inverse problems. Preprint (2017)
16. McClure, J., Wong, R.: Error bounds for multidimensional Laplace approximation. *J. Approx. Theory* **37**(4), 372–390 (1983). [https://doi.org/10.1016/0021-9045\(83\)90044-8](https://doi.org/10.1016/0021-9045(83)90044-8)
17. Monard, F., Nickl, R., Paternain, G.P., et al.: Efficient nonparametric Bayesian inference for X-ray transforms. *Ann. Stat.* **47**(2), 1113–1147 (2019)
18. Nemes, G.: An explicit formula for the coefficients in Laplace’s method. *Constr. Approx.* **38**(3), 471–487 (2013)
19. Nickl, R.: Bernstein-von Mises theorems for statistical inverse problems I: Schrödinger equation. *J. Eur. Math. Soc.* **22**, 2697–2750 (2020)
20. Olver, F.W.J.: Error bounds for the Laplace approximation for definite integrals. *J. Approx. Theory* **1**(3), 293–313 (1968). [https://doi.org/10.1016/0021-9045\(68\)90007-5](https://doi.org/10.1016/0021-9045(68)90007-5)
21. Olver, F.W.J.: *Asymptotics and Special Functions*. Academic Press, Cambridge (1974). <https://doi.org/10.1016/C2013-0-11254-8>
22. Rue, H., Martino, S., Chopin, N.: Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* **71**(2), 319–392 (2009)
23. Ryan, E.G., Drovandi, C.C., McGree, J.M., Pettitt, A.N.: A review of modern computational algorithms for Bayesian optimal design. *Int. Stat. Rev.* **84**(1), 128–154 (2016)
24. Schillings, C., Schwab, C.: Scaling limits in computational Bayesian inversion. *ESAIM: Math. Model. Numer. Anal.* **50**(6), 1825–1856 (2016)
25. Schillings, C., Sprungk, B., Wacker, P.: On the convergence of the Laplace approximation and noise-level-robustness of Laplace-based Monte Carlo methods for Bayesian inverse problems. *Numer. Math.* **145**, 915–971 (2020). <https://doi.org/10.1007/s00211-020-01131-1>

26. Stuart, A.M.: Inverse problems: a Bayesian perspective. *Acta Numer.* **19**, 451 (2010). <https://doi.org/10.1017/S0962492910000061>
27. Sullivan, T.: Introduction to Uncertainty Quantification. *Texts in Applied Mathematics*, vol. 63, 1st edn. Springer, Berlin (2015). <https://doi.org/10.1007/978-3-319-23395-6>
28. Temme, N.M.: *Special Functions: An Introduction to the Classical Functions of Mathematical Physics*. Wiley, New York (1996). <https://doi.org/10.1002/9781118032572>
29. Van der Vaart, A.W.: *Asymptotic Statistics*, vol. 3. Cambridge University Press, Cambridge (2000). <https://doi.org/10.1017/CBO9780511802256>
30. Wong, R.: Asymptotic approximations of integrals. *Soc. Ind. Appl. Math.* (2001). <https://doi.org/10.1137/1.9780898719260>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.