arXiv:2209.14992v1 [math.ST] 29 Sep 2022

HOW GOOD IS YOUR GAUSSIAN APPROXIMATION OF THE POSTERIOR? FINITE-SAMPLE COMPUTABLE ERROR BOUNDS FOR A VARIETY OF USEFUL DIVERGENCES

BY MIKOŁAJ J. KASPRZAK[1,2,a], RYAN GIORDANO[1,b] AND TAMARA BRODERICK[1,c]

[1] *Laboratory for Information and Decision Systems, Massachusetts Institute of Technology*

[2]*Department of Mathematics, University of Luxembourg*

[a]*mikolaj.kasprzak@uni.lu*; [b]*rgiordan@mit.edu*; [c]*tbroderick@mit.edu*

The Bayesian Central Limit Theorem (BCLT) for finite-dimensional models, also known as the Bernstein – von Mises Theorem, is a primary motivation for the widely-used Laplace approximation. But currently the BCLT is expressed only in terms of total variation (TV) distance and lacks non-asymptotic bounds on the rate of convergence that are readily computable in applications. Likewise, the Laplace approximation is not equipped with non-asymptotic quality guarantees for the vast classes of posteriors for which it is asymptotically valid. To understand its quality and real-problem applicability, we need finite-sample bounds that can be computed for a given model and data set. And to understand the quality of posterior mean and variance estimates, we need bounds on divergences alternative to the TV distance. Our work provides the first closed-form, finite-sample bounds for the quality of the Laplace approximation that do not require log concavity of the posterior or an exponential-family likelihood. We bound not only the TV distance but also (A) the Wasserstein-1 distance, which controls error in a posterior mean estimate, and (B) an integral probability metric that controls the error in a posterior variance estimate. We compute exact constants in our bounds for a variety of standard models, including logistic regression, and numerically investigate the utility of our bounds. And we provide a framework for analysis of more complex models.

1. Introduction.

1.1. *Motivation.* Researchers and practitioners who use statistical inference for solving real-world problems need good point estimates and uncertainties. Bayesian inference provides a way of obtaining those through expectations calculated with respect to the posterior distribution – and in particular through the posterior mean and variance. Such expectations are, however, often intractable or costly to compute, which forces users to use approximations. An easy and fast way approach is to approximate the posterior by a suitably chosen Gaussian distribution. This is known at *the Laplace approximation* in the approximate Bayesian inference literature. As we describe below, in Subsection 1.3, it is grounded in the celebrated *Bernstein–von Mises theorem*. Laplace approximation is a commonly used tool in many communities [26, 33]. Studies have shown its appealing empirical performance, for instance in the context of Bayesian neural networks [9]. Widely applicable, computable and rigorously justified theoretical guarantees on the quality of the Laplace approximation are, however, still not available in the literature. This leaves researchers and practitioners using it unable to tell with high confidence whether their inference is robust enough for their purposes. Moreover, convergence in the Bernstein–von Mises theorem is normally expressed in

MSC2020 subject classifications: Primary 62E17, 62F15; secondary 60F05, 60E05, 60B10, 60E15.

Keywords and phrases: Bayesian inference, Laplace approximation, Bernstein–von Mises theorem, approximate inference, log-Sobolev inequality, Stein's method.

terms of the total variation distance. Likewise, the Laplace approximation is typically justified by the fact that the total variation distance between the rescaled posterior and the Gaussian vanishes in the limit. The total variation distance, however, does not control the difference of means or the difference of covariances in general. At the same time, exactly those quantities are most often reported by users of approximate Bayesian inference. Practitioners and researchers therefore need finite sample guarantees on the quality of Laplace approximation, expressed not only in terms of the total variation distance but also in terms of metrics that control the error in mean and covariance approximation.

1.2. *Our contribution.* Our work provides the first closed-form, fully computable, finite-sample bounds for the Laplace approximation that do not require the likelihood to come from an exponential family or the posterior to be log-concave . We quantify the control not only over the total variation distance but also

- the Wasserstein-1 distance, which controls the difference of means
- another integral probability metric that bounds the difference of covariances.

Our bounds are computable without access to the true parameter or integrals with respect to the posterior. They are expressed in terms of the data and work under any distribution of the data, also when the model is misspecified. In particular, our results are fully applicable to models involving generalized likelihoods and the resulting generalized posteriors (see [4, 29, 6]). Our assumptions on the generalized likelihood and the prior are standard and no stronger than the assumptions of the classical poofs of the Bernstein–von Mises Theorem, for instance Le Cam's one (see [14, Section 1.4] for details). We compute our bounds explicitly for a variety of Bayesian models, including logistic regression with Student's t prior.

Our contribution lies also in our proof techniques. In order to control the discrepancy inside a ball around the maximum likelihood estimator (MLE) or the maximum a posteriori (MAP), we use the log-Sobolev inequality or Stein's method. In order to control the discrepancy over the rest of the parameter space, we carefully bound the tail growth using standard assumptions of the Bernstein–von Mises Theorem. We believe that this approach to proving computable non-asymptotic bounds could be extended so as to cover more general statistical models satisfying the conditions of the local asymptotic normality (LAN) theory [22, Chapters 1-3]. We consider this, however, a separate problem and leave it for future work.

1.3. *Laplace approximation and the Bernstein–von Mises theorem.* The foundations of Laplace approximation date back to the work on Laplace [24] (see [25] for an English translation and [1] for an intuitive discussion). It was originally introduced as a method of approximating integrals of the form

$$\text{Int}(n) := \int_K e^{-nf(x)} dx, \quad n \in \mathbb{N},$$

where K is a subset of $\mathbb{R}^d$ and f is a real-valued function on $\mathbb{R}^d$. Suppose that $x^* \in K$ is a strict global maximizer of f on K. Heuristically, under appropriate smoothness assumptions on f, we can use Taylor's expansion to obtain:

$$\begin{aligned} f(x) \approx & f(x^*) + f'(x^*)(x - x^*) + \frac{1}{2}(x - x^*)^T f''(x^*)(x - x^*) \\ = & f(x^*) + \frac{1}{2}(x - x^*)^T f''(x^*)(x - x^*). \end{aligned}$$

We therefore have:

$$\text{Int}(n) \approx \int_K \exp\left[-nf(x^*) - \frac{n}{2}(x - x^*)^T f''(x^*)(x - x^*)\right] dx$$

As a result, heuristically, up to a constant not depending on K, Int(n) can be approximated by the integral of the density of the Gaussian measure with mean x^* and covariance matrix given by $\frac{1}{n}f''(x^*)^{-1}$. Now, suppose that $e^{-nf(x)}$ is an unnormalized posterior density, i.e. that $\Pi_n(\cdot) \propto e^{-nf(\cdot)}$ for some posterior Π_n. Writing $\mathbf{\Pi}_n$ for the posterior probability measure and $\bar{\theta}_n$ for the posterior mode (i.e. the maximum a posteriori or MAP), the above statement suggests that

$$\mathbf{\Pi}_n \approx \mathcal{N}(\bar{\theta}_n, \left(\log(\Pi_n)''(\bar{\theta}_n)\right)^{-1}), \tag{1.1}$$

where $\mathcal{N}(\mu, \Sigma)$ denotes the normal law with mean μ and covariance Σ. The computation of the mean and covariance of the above Gaussian is in the majority of cases easy numerically. It can be achieved using standard optimization schemes and does not require access to integrals with respect to the posterior, the normalizing constant of the posterior density or the true parameter. This is why Laplace approximation is a popular tool in approximate Bayesian inference.

While the above heuristic considerations may be turned into rigorous statements under certain conditions, a proper probabilistic grounding for the Laplace approximation is provided by the Bernstein–von Mises (BvM) theorem. As described in numerous classical references, including [14, Section 1.4] or [38, Section 10.2], the BvM theorem says that under mild assumptions on the likelihood and the prior, the posterior distribution converges to a Gaussian law in the following sense. Suppose that $\tilde{\boldsymbol{\theta}}_n$ is distributed according to the posterior, obtained after observing n data points. Let θ_0 be the true parameter, $\hat{\theta}_n$ be the MLE and $I(\cdot)$ be the Fisher information matrix. Let TV denote the total variation distance and let the function $\mathcal{L}(\cdot)$ return the law of its argument. Then, if the model is well-specified and certain regularity conditions are satisfied,

$$\text{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)\right), \mathcal{N}(0, I(\theta_0)^{-1})\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty, \tag{1.2}$$

where n is the number of data and the convergence occurs in probability with respect to the law of the data.

While the model being well-specified is a crucial assumption in the above statement, the authors of [23] proved its modified version, under model misspecification. The main difference is in the limiting covariance matrix. Specifically, in this context, the authors assume the model is of the form $\theta \mapsto p_\theta$ and the observations are sampled from a density p_0 that is not necessarily of the form p_{θ_0} for some θ_0. They show that under certain regularity conditions,

$$\text{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)\right), \mathcal{N}(0, V(\theta^*)^{-1})\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty, \tag{1.3}$$

where θ^* minimizes the Kullback-Leibler divergence $\theta \mapsto \int \log(p_0(x)/p_\theta(x))p_0(x)dx$ and $V(\theta^*)$ is minus the second derivative of this map, evaluated at θ^*.

Let us now denote by L_n the generalized log-likelihood. A closer look at the classical proofs, including Le Cam's one (see e.g. [14, Section 1.4]), or more recent ones, including that of [29, Appendix B], reveals that, under standard regularity conditions:

$$\text{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)\right), \mathcal{N}\left(0, \left[-\frac{L_n''(\hat{\theta}_n)}{n}\right]^{-1}\right)\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty, \tag{1.4}$$

no matter if the model is well-specified or not. It is known that under mild assumptions the MLE $\hat{\theta}_n$ and the maximum a posteriori (MAP) $\bar{\theta}_n$ get arbitrarily close to each other as the number of data n goes to infinity. Similarly, denoting by $\overline{L}_n$ the logarithm of the posterior density, $\overline{L}_n$ and L_n get arbitrarily close as n goes to infinity. It can be shown, in a similar fashion to (1.4), that under standard regularity assumptions,

$$\text{TV}\left(\mathcal{L}\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n)\right), \mathcal{N}\left(0, \left[-\frac{\overline{L}_n''(\bar{\theta}_n)}{n}\right]^{-1}\right)\right) \xrightarrow{\mathcal{P}} 0, \qquad \text{as } n \to \infty. \tag{1.5}$$

Equation (1.5) gives a rigorous justification and meaning to the Laplace approximation (1.1) and (1.4) provides its alternative version. The approximating covariance in both (1.5) and (1.4) is computable without access to the posterior normalizing constant or the true parameter.

The recent paper [29] proves almost sure versions of the statements (1.2) and (1.3) for a large collection of commonly used models. Naturally, similar almost sure convergence statements can be obtained for the approximations appearing in (1.4) and (1.5). Below, in Sections 2 and 4, we shall provide computable finite-sample bounds on the total variation distances appearing in (1.4) and (1.5). We will also prove analogous bounds on the 1-Wasserstein distance and another metric that controls the difference of covariances. Our bounds depend on the data and are computable in applications, for real data-sets. Apart from that, they can be used to quantify the convergence in probability appearing in (1.4) and (1.5), or the analogous almost sure convergence, when it occurs.

1.4. *Setup, assumptions and notation.* Our setup and assumptions are similar to those found for instance in [29]. We fix $n \in \mathbb{N}$ and study probability measures on $\mathbb{R}^d$ having Lebesgue densities of the form:

$$\Pi_n(\theta) = e^{L_n(\theta)}\pi(\theta)/z_n, \tag{1.6}$$

where $\pi : \mathbb{R}^d \to \mathbb{R}$ is a Lebesgue probability density function, $L_n : \mathbb{R}^d \to \mathbb{R}$ and $z_n \in \mathbb{R}_+$ is a suitable normalizing constant. Throughout the paper, we call π *the prior density* (or simply *the prior*), L_n *the generalized log-likelihood* and Π_n *the generalized posterior*. By

$$\overline{L}_n(\theta) := \log\left(\Pi_n(\theta)\right)$$

we denote the *generalized log-posterior* and let:

$$\hat{\theta}_n := \arg\max_{\theta \in \mathbb{R}^d} L_n(\theta), \qquad \bar{\theta}_n = \arg\max_{\theta \in \mathbb{R}^d} \overline{L}_n(\theta),$$

whenever those quantities exist. If those quantities are unique, we call $\hat{\theta}_n$ the *maximum likelihood estimator* (*MLE*) and $\bar{\theta}_n$ the *maximum a posteriori* (*MAP*). For any twice-differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, we let f' stand for its gradient and f'' for its Hessian. We shall write

$$J_n(\theta) = -\frac{L_n''(\theta)}{n}, \qquad \bar{J}_n(\theta) = -\frac{\overline{L}_n''(\theta)}{n},$$

whenever those expressions make sense (i.e. when the Hessians exist). For any three times differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, we will also write f''' for its third (Frechét) derivative, defined as the following multilinear 3-form on $\mathbb{R}^d$:

$$f'''(\theta)[u, v, w] = \sum_{i,j,k=1}^{d} \frac{\partial f}{\partial\theta_i\partial\theta_j\partial\theta_k}(\theta)u_i v_j w_k.$$

The norm $\|\cdot\|^*$ of this third derivative will be defined in the following way:

$$\|f'''(\theta)\|^* := \sup_{\|u\|\le 1, \|v\|\le 1, \|w\|\le 1} \left|f'''(\theta)[u, v, w]\right|,$$

where $\|\cdot\|$ denotes the Euclidean norm, as it will throughout the paper. We will also let $\lambda_{\min}(\hat{\theta}_n)$ be the minimal eigenvalue of $J_n(\hat{\theta}_n)$ and $\overline{\lambda}_{\min}(\bar{\theta}_n)$ be the minimal eigenvalue of $\bar{J}_n(\bar{\theta}_n)$. Throughout the paper $\|\cdot\|_{op}$ will denote the operator (i.e. spectral) norm, $\langle\cdot,\cdot\rangle$ will be the Euclidean inner product and $\tilde{\boldsymbol{\theta}}_n$ will always denote a random variable distributed according to the generalized posterior measure with density (1.6). $\mathcal{N}(\mu,\Sigma)$ will denote the normal distribution with mean μ and covariance Σ and function $\mathcal{L}(\cdot)$ will return the law of its argument. $I_{d\times d}$ will always denote the d-dimensional identity matrix.

Our bounds will be derived for two types of approximations. The first type is what we call the *MLE-centric approach*. Within this approach, $\mathcal{L}(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n))$ is approximated by $\mathcal{N}(0,J_n(\hat{\theta}_n)^{-1})$. On the other hand, what we call the *MAP-centric approach* is the approximation of $\mathcal{L}(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n))$ by $\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})$. The bounds we obtain are on the following distances:

1. The Total Variation (TV) distance, which, for two probability measures ν_1 and ν_2 on a measurable space $(\Omega,\mathcal{F})$ is defined by
$$TV(\nu_1,\nu_2):=\sup_{A\in\mathcal{F}}|\nu_1(A)-\nu_2(A)|$$
2. The Wasserstein-1 distance, which, for probability measures ν_1 and ν_2 and the set $\Gamma(\nu_1,\nu_2)$ of all couplings between them, is defined by
$$W_1(\nu_1,\nu_2):=\inf_{\gamma\in\Gamma(\nu_1,\nu_2)}\int\|x-y\|d\gamma(x,y).$$
Kantorovich duality (see, e.g. [40, Theorem 5.10]) provides an equivalent definition. Let $\|\cdot\|_L$ return the Lipschitz constant of the input. Then
$$W_1(\nu_1,\nu_2)=\sup_{\substack{f \text{ Lipschitz:}\\ \|f\|_L=1}}\left|\mathbb{E}_{\nu_1}f-\mathbb{E}_{\nu_2}f\right|.$$
3. The following integral probability metric which, for $Y_1\sim\nu_1$ and $Y_2\sim\nu_2$ is defined by
$$\sup_{v:\|v\|\le 1}\left|\mathbb{E}\left\langle v,Y_1\right\rangle^2-\mathbb{E}\left\langle v,Y_2\right\rangle^2\right|.$$

1.4.1. *Assumptions made throughout the paper.* Now we list the assumptions that we will need to prove our finite-sample bounds and define constants used therein. We will present those assumptions which will stand for both approaches described above and others, which are divided between those relevant for the MLE and MAP approach. We reiterate that the conditions we require are similar to the classical assumptions of the Bernstein–von Mises theorem, as given in [14, 29].

The first fundamental assumption that will be used throughout the paper is the following:

ASSUMPTION 1. *There exists a unique MLE $\hat{\theta}_n$. There also exists a real number $\delta>0$ such that the generalized log-likelihood L_n is three times differentiable inside $\{\theta:\|\theta-\hat{\theta}_n\|\le\delta\}$. For the same $\delta>0$ there exists a real number $M_2>0$, such that:*

$$\sup_{\|\theta-\hat{\theta}_n\|\le\delta}\frac{\|L_n'''(\theta)\|^*}{n}\le M_2. \tag{1.7}$$

Moreover, we make the following assumption on the prior:

ASSUMPTION 2. *For the same $\delta > 0$ as in Assumption 1, there exists a real number $\hat{M}_1 > 0$, such that*

$$\sup_{\theta:\|\theta-\hat{\theta}_n\|<\delta} \left|\frac{1}{\pi(\theta)}\right| \leq \hat{M}_1.$$

REMARK 1.1. Note that for Assumption 2 to be satisfied, it suffices to assume that π is continuous and positive in the δ-ball around $\hat{\theta}_n$.

1.4.2. *Additional assumptions in the MLE-centric approach.* Besides Assumptions 1 and 2, in the MLE-centric approach, we have the following assumptions:

ASSUMPTION 3. *For the same $\delta > 0$ and $M_2 > 0$ as in Assumption 1,*

$$\lambda_{\min}(\hat{\theta}_n) > \delta M_2. \tag{1.8}$$

REMARK 1.2. If $J_n(\hat{\theta}_n)$ is positive definite and Assumption 1 is satisfied then one can adjust the choice of δ so that both (1.7) and (1.8) hold. This is because decreasing the value of δ in Assumption 1 does not lead to an increase in the value of M_2. At the same time, decreasing the value of δ in Assumption 2, while keeping M_2 fixed, decreases the right-hand side of (1.8).

ASSUMPTION 4. *For the same $\delta > 0$, as in Assumption 1, there exists $\kappa > 0$, such that*

$$\sup_{\theta:\|\theta-\hat{\theta}_n\|>\delta} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \leq -\kappa.$$

REMARK 1.3. Assumption 4 ensures that any local maxima of L_n achieved outside of the δ-ball around the MLE do not get arbitrarily close to the global maximum achieved at the MLE.

ASSUMPTION 5. *For the same δ as in Assumption 1, there exists real numbers $M_1 > 0$ and $\widetilde{M_1} > 0$, such that*

$$\sup_{\theta:\|\theta-\hat{\theta}_n\|\leq\delta} \left\|\frac{\pi'(\theta)}{\pi(\theta)}\right\| \leq M_1 \qquad \textit{and} \qquad \sup_{\theta:\|\theta-\hat{\theta}_n\|\leq\delta} |\pi(\theta)| \leq \widetilde{M_1}.$$

REMARK 1.4. Note that for Assumption 5 to be satisfied, it suffices that π is continuously differentiable and positive inside the δ-ball around $\hat{\theta}_n$.

1.4.3. *Assumptions in the MAP-centric approach.* In the MAP-centric approach we keep Assumptions 1 and 2 and additionally assume the following:

ASSUMPTION 6. *There exists a unique MAP $\bar{\theta}_n$. There also exists a real number $\bar{\delta} > 0$, such that the log-prior, $\log \pi$, is three times differentiable inside $\{\theta : \|\theta - \bar{\theta}_n\| \leq \bar{\delta}\}$. Moreover, for the same $\bar{\delta}$, there exists a real number $\overline{M}_2 > 0$, such that*

$$\sup_{\theta:\|\theta-\bar{\theta}_n\|\leq\bar{\delta}} \frac{\|\overline{L}_n'''(\theta)\|^*}{n} \leq \overline{M}_2$$

ASSUMPTION 7. *For the same $\bar{\delta} > 0$ and $\overline{M}_2 > 0$ as in Assumption 6,*

$$\overline{\lambda}_{\min}(\bar{\theta}_n) > \bar{\delta}\,\overline{M}_2.$$

REMARK 1.5. Assumption 6 is very similar to Assumption 1. The difference is that we now consider a ball around the MAP rather than the MLE and we require additional differentiability of the prior density inside this ball. Assumption 7 is an analogue of Assumption 3 for the MAP-centric approach.

ASSUMPTION 8. *For the same $\bar{\delta} > 0$, as in Assumption 6, there exists $\bar{\kappa} > 0$, such that*

$$\sup_{\theta:\|\theta-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \frac{L_n(\theta)-L_n(\hat{\theta}_n)}{n} \leq -\bar{\kappa}.$$

REMARK 1.6. Assumption 8 is very similar to Assumption 4. Note that for the vast majority of commonly used parametric models and data generating distributions, $\|\hat{\theta}_n - \bar{\theta}_n\|$, which appears in the expression for the radius of the ball, will tend to 0 as $n \to \infty$, almost surely.

1.5. *Structure of the paper.* In Section 2 we present our results on the MAP-centric approach. In Section 3, we discuss the related work and compare our results to the existing non-asymptotic studies of the Bernstein–von Mises Theorem and the Laplace approximation. In Section 4 we present results analogous to those of Section 2, yet focused on the MLE-centric approach. We also present a bound in dimension one which controls the difference of expectations of general test functions, going beyond quadratic ones. In Section 5 we show how to compute our bounds for the (non-log-concave) posterior in the logistic regression model with Student's t prior. We also present plots of our bounds computed numerically in this case. Moreover, we numerically compare our control over the difference of means and the difference of variances to the ground truth for some conjugate prior models. In Section 6 we present conclusions of our work. All the proofs of the results of Sections 2 and 4 are postponed to the appendix.

2. Main results.

In this section we present our bounds on the quality of Laplace approximation in the *MAP-centric* approach, as described in Subsection 1.4. This approach is arguably the most popular one among users of Laplace approximation. The proofs of the results from this section are presented in Appendices A and B. A discussion of our bounds can be found in Subsection 2.5 below. Our bounds in the *MLE-centric* approach will be presented below, in Section 4.

2.1. *Additional notation.* Before we state the results, we introduce additional notation that will make presentation of the bounds more concise. We let

$$R_1(n,\delta) := \left(J_n(\hat{\theta}_n) + (\delta M_2/3) I_{d\times d}\right)^{-1}; \quad R_2(n,\delta) := \left(J_n(\hat{\theta}_n) - (\delta M_2/3) I_{d\times d}\right)^{-1};$$
$$\bar{R}_1(n,\bar{\delta}) := \left(\bar{J}_n(\bar{\theta}_n) + (\bar{\delta}\,\overline{M}_2/3) I_{d\times d}\right)^{-1} \quad \bar{R}_2(n,\bar{\delta}) := \left(\bar{J}_n(\bar{\theta}_n) - (\bar{\delta}\,\overline{M}_2/3) I_{d\times d}\right)^{-1}.$$

2.2. *Control over the total variation distance.* We start with a bound over the total variation distance.

THEOREM 2.1. *Suppose that Assumptions 1, 2 and 6 – 8 hold and retain the notation theoreof. Suppose that* $\max\left\{\|\hat{\theta}_n - \bar{\theta}_n\|, \sqrt{\frac{Tr\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}}\right\} < \bar{\delta}$ *and* $\sqrt{\frac{Tr[R_1(n,\delta)]}{n}} < \delta$*. Let TV* *denote the total variation distance. Then:*

$$
\begin{aligned}
&TV\left(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right), \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})\right)\\
\leq & A_1 n^{-1/2} + 2\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right] + A_2 n^{d/2} e^{-n\bar{\kappa}},
\end{aligned}
$$

where

$$
A_1 = \frac{\sqrt{3}\, Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\sqrt{\overline{M}_2}}{4\sqrt{\left(\overline{\lambda}_{\min}(\bar{\theta}_n)/\overline{M}_2-\bar{\delta}\right)\left(1-\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\overline{\lambda}_{\min}(\bar{\theta}_n)\right]\right)}};
$$

$$
A_2 = \frac{2\hat{M}_1 \left|det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{Tr\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.
$$

2.3. *Control over the Wasserstein-1 distance.* Now, we bound the 1-Wasserstein distance, which is known to control the difference of means.

THEOREM 2.2. *Suppose that Assumptions 1, 2 and 6 – 8 hold and retain the notation theoreof. Suppose that* $\max\left\{\|\hat{\theta}_n-\bar{\theta}_n\|, \sqrt{\frac{Tr\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}}\right\} < \bar{\delta}$ *and* $\sqrt{\frac{Tr[R_1(n,\delta)]}{n}} < \delta$. *Let* W_1 *denote the Wasserstein-1 distance. Then:*

$$
\begin{aligned}
&W_1\left(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\bar{\theta}_n\right)\right), \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})\right)\\
\leq & B_1 n^{-1/2} + B_3\left[B_2+\sqrt{n}\int_{\|u-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|u-\bar{\theta}_n\|\pi(u)du\right]n^{d/2}e^{-n\bar{\kappa}}\\
&+\left(\bar{\delta}\sqrt{n}+\sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}}+B_2\right)\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{Tr\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\overline{\lambda}_{\min}(\bar{\theta}_n)\right],
\end{aligned}
$$

where

$$
B_1 := \frac{\sqrt{3}\, Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{2\left(\overline{\lambda}_{\min}(\bar{\theta}_n)/\overline{M}_2-\bar{\delta}\right)\sqrt{\left(1-\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\overline{\lambda}_{\min}(\bar{\theta}_n)\right]\right)}};
$$

$$
B_2 := \frac{\left|det\left(\bar{R}_1(n,\delta)\right)\right|^{-1/2}\left|det\left(\bar{R}_2(n,\delta)\right)\right|^{1/2}\sqrt{Tr\left[\bar{R}_2(n,\delta)\right]}}{1-\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{Tr\left[\bar{R}_1(n,\delta)\right]}\right)^2\left[\left\|\bar{R}_1(n,\delta)\right\|_{op}\right]^{-1}\right]};
$$

$$
B_3 := \frac{\hat{M}_1\left|det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{Tr\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.
$$

2.4. *Control over the difference of covariances.* Finally, we upper bound an integral probability metric that lets us control the difference of covariances.

THEOREM 2.3. *Suppose that Assumptions 1, 2 and 6 – 8 hold and retain the notation theoreof. Suppose that* $\max\left\{\|\hat{\theta}_n - \bar{\theta}_n\|, \sqrt{\frac{Tr\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}}\right\} < \bar{\delta}$ *and* $\sqrt{\frac{Tr[R_1(n,\delta)]}{n}} < \delta$. *Let* $Z_n \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$. *Then:*

$$\begin{aligned}
&\sup_{v:\|v\|\leq 1}\left|\mathbb{E}\left[\left\langle v, \sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right\rangle^2\right] - \mathbb{E}\left[\langle v, Z_n\rangle^2\right]\right| \\
\leq & \frac{3Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]C_1}{4\left(\overline{\lambda}_{\min}(\bar{\theta}_n)/\overline{M}_2 - \bar{\delta}\right)^2} n^{-1} + \frac{\sqrt{3Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\,C_1}{\left(\overline{\lambda}_{\min}(\bar{\theta}_n)/\overline{M}_2 - \bar{\delta}\right)} n^{-1/2} \\
& + \left(\bar{\delta}^2 n + \frac{\sqrt{2\pi}}{\overline{\lambda}_{\min}(\bar{\theta}_n)}\right)\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{Tr\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right] \\
& + n^{d/2+1}e^{-n\bar{\kappa}}C_3\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|u-\bar{\theta}_n\|^2\pi(u)du \\
& + C_2\left\{\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{Tr\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right] + C_3 n^{d/2}e^{-n\bar{\kappa}}\right\}
\end{aligned}$$

for

$$C_1 := \frac{Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{\left(1 - \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{Tr\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right]\right)}$$

$$C_2 := \frac{\left|det\left(\bar{R}_1(n,\delta)\right)\right|^{-1/2}\left|det\left(\bar{R}_2(n,\delta)\right)\right|^{1/2} Tr\left[\bar{R}_2(n,\delta)\right]}{1 - \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{Tr\left[\bar{R}_1(n,\delta)\right]}\right)^2\left[\left\|\bar{R}_1(n,\delta)\right\|_{op}\right]^{-1}\right]}$$

$$C_3 := \frac{\hat{M}_1\left|det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.$$

2.5. *Discussion of the bounds.* We make some remarks about the bounds and their applicability is approximate inference:

REMARK 2.4. The quantities $A_1, A_2, A_3, B_1, B_2, B_3, C_1, C_2, C_3$ appearing in the above bounds depend on n but, for models and data generating distributions that satisfy the assumptions of the Bernstein–von Mises theorem (as in [14]), they are bounded in n. In particular, they are almost surely bounded in n, as long as the constants $M_2, \overline{M}_2, \hat{M}_1$ are bounded from above and $\bar{\kappa}$ is bounded from below by a positive number and the data are i.i.d. (without any other assumption on their distribution). Therefore, our bounds are to be expected to vanish as $n \to \infty$ at the rate of $\frac{1}{\sqrt{n}}$ for the majority of common modelling setups. We reiterate that our bounds are fully non-asymptotic and computable.

REMARK 2.5. In the typical applications, the bounds in Theorems 2.1 - 2.3 depend on the data. One can assume the data come from a certain distribution, should that be of interest. It is then straightforward to use our bounds in order to quantify the speed of almost sure convergence or convergence in probability of the distances between the prior and the Gaussian.

In order to do this, one only needs to control the speed of the relevant mode of convergence of our bounds, which, in most cases, should be achievable using standard results, similar to those quantifying the rate of convergence in the law of large numbers.

REMARK 2.6. Assumption saying that $\max\left\{\|\hat{\theta}_n - \bar{\theta}_n\|, \sqrt{\frac{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}}\right\} < \bar{\delta}$ and $\sqrt{\frac{\mathrm{Tr}[R_1(n,\delta)]}{n}} < \delta$ appearing in the Theorems above is an assumption on the size of n and the choice of $\bar{\delta}$. Indeed, as long as the MLE and MAP converge to the same limit (which is true in the majority of commonly used modelling setups), we expect $\max\left\{\|\hat{\theta}_n - \bar{\theta}_n\|, \sqrt{\frac{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{n}}\right\}$ to go to zero as $n \to \infty$. We similarly expect $\sqrt{\frac{\mathrm{Tr}[R_1(n,\delta)]}{n}}$ to go to zero. Moreover $\sqrt{\frac{\mathrm{Tr}[R_1(n,\delta)]}{n}} < \delta$ will be satisfied if $\sqrt{\frac{\mathrm{Tr}\left(J_n(\hat{\theta}_n)^{-1}\right)}{n}} < \delta$, which might be an easier condition to check.

REMARK 2.7. Our bound on the total variational distance provides quality guarantees on the approximate computation of posterior credible sets. Indeed, suppose, for instance, that one is interested in finding a value b_α, such that

$$\mathbb{P}\left(\|\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\| \leq \frac{b_\alpha}{\sqrt{n}}\right) \geq 1 - \alpha,$$

for a fixed value α. Let $A(n)$ denote the value of our upper bound in Theorem 2.1. If n is sufficiently large and $A(n)$ is smaller that α, then one could choose $b_\alpha = \tilde{b}_\alpha$, such that

$$\mathbb{P}\left(\|\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})\| \leq \tilde{b}_\alpha\right) = 1 - \alpha + A(n).$$

Our bound implies that:

$$\left|\mathbb{P}\left(\|\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\| \leq \frac{\tilde{b}_\alpha}{\sqrt{n}}\right) - \mathbb{P}\left(\|\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})\| \leq \tilde{b}_\alpha\right)\right| \leq A(n)$$

and so

$$\mathbb{P}\left(\|\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\| \leq \frac{\tilde{b}_\alpha}{\sqrt{n}}\right) \geq 1 - \alpha.$$

REMARK 2.8. Our bound on the Wasserstein 1-distance controls the difference of means in the Laplace approximation, in the following way. The upper bound in Theorem 2.2 controls $\sqrt{n}\|\mathbb{E}[\tilde{\boldsymbol{\theta}}_n] - \bar{\theta}_n\|$. In order to obtain an upper bound on $\|\mathbb{E}[\tilde{\boldsymbol{\theta}}_n] - \bar{\theta}_n\|$, one simply needs to divide our bound from Theorem 2.2 by $\sqrt{n}$.

REMARK 2.9. Theorem 2.3 together with Theorem 2.2 let us control the difference of covariances. Suppose, for instance, that we are interested in the operator norm of the difference of the posterior covariance matrix and the covariance matrix of the Gaussian Laplace approximation. Let $B(n)$ denote the value of our bound from Theorem 2.2 and let $C(n)$ be the value of our bound from Theorem 2.3. Then, for $Z_n \sim \mathcal{N}\left(0, \bar{J}_n(\bar{\theta}_n)^{-1}\right)$

$$\begin{aligned}&\left\|\mathrm{Cov}(\tilde{\boldsymbol{\theta}}_n) - \frac{\bar{J}_n(\bar{\theta}_n)^{-1}}{n}\right\|_{op}\\ =&\frac{1}{n}\sup_{v:\|v\|\leq 1}\left|\mathbb{E}\left\langle v, \sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \mathbb{E}\tilde{\boldsymbol{\theta}}_n\right)\right\rangle^2 - \mathbb{E}\left\langle v, Z_n\right\rangle^2\right|\end{aligned}$$

TABLE 1
Very brief summary of related results. See the text for full details.

	Present work	[32]	[37]	[21]	[10]	[36]	[42]
True parameter not needed	Y	N	Y	Y	Y	Y	N
Global log-concavity not needed	Y	Y	N	N	N	N	Y
Generic priors / likelihoods	Y	N	N	Y	Y	N	N
Explicit bounds (no order notation)	Y	Y	Y	Y	N	Y	Y
Controls TV distance	Y	Y	Y	N	Y	Y	Y
Controls means and variances	Y	Y	Y	Y	N	N	N

$$\leq \frac{1}{n} \sup_{v:\|v\|\leq 1} \left| \mathbb{E} \left\langle v, \sqrt{n} \left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right\rangle^2 - \mathbb{E} \left\langle v, Z_n \right\rangle^2 \right|$$
$$+ \frac{1}{n} \sup_{v:\|v\|\leq 1} \left| \mathbb{E} \left\langle v, \sqrt{n} \left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right\rangle^2 - \mathbb{E} \left\langle v, \sqrt{n} \left(\tilde{\boldsymbol{\theta}}_n - \mathbb{E}\tilde{\boldsymbol{\theta}}_n \right) \right\rangle^2 \right|.$$

Now, note that

$$\begin{aligned}
&\sup_{v:\|v\|\leq 1} \left| \mathbb{E} \left\langle v, \sqrt{n} \left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right\rangle^2 - \mathbb{E} \left\langle v, \sqrt{n} \left(\tilde{\boldsymbol{\theta}}_n - \mathbb{E}\tilde{\boldsymbol{\theta}}_n \right) \right\rangle^2 \right| \\
&= \sup_{v:\|v\|\leq 1} \left| \mathbb{E} \left[\left\langle v, \sqrt{n} \left(\mathbb{E}\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right\rangle \left\langle v, \sqrt{n} \left(2\tilde{\boldsymbol{\theta}}_n - \mathbb{E}\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right] \right\rangle \right| \\
&= \sup_{v:\|v\|\leq 1} \left| \left\langle v, \sqrt{n} \left(\mathbb{E}\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right\rangle \left\langle v, \sqrt{n} \left(2\mathbb{E}\tilde{\boldsymbol{\theta}}_n - \mathbb{E}\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right\rangle \right| \\
&\leq n \|\mathbb{E}\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\|^2 \\
&\leq B(n)^2.
\end{aligned}$$

Therefore,

$$\left\| \mathrm{Cov}(\tilde{\boldsymbol{\theta}}_n) - \frac{\bar{J}_n(\bar{\theta}_n)^{-1}}{n} \right\|_{op} \leq \frac{1}{n} \left(B(n)^2 + C(n) \right).$$

3. Related work.

A number of recent papers have studied the non-asymptotic properties of the Laplace approximation and the Bernstein von-Mises theorem for Bayesian posteriors. In this section we contrast the present work with these analyses in terms of the computability of the bounds, strength of the results, and restrictiveness of the assumptions. A brief summary of the differences can be found in table 1.

A non-asymptotic analysis of the Bernstein–von Mises (BvM) Theorem has previously been performed in [32]. The aim and focus of that analysis is however significantly different from that of the present paper. In [32] the authors consider semiparametric inference and prove results whose purpose it is to provide insight into the *critical dimension* of the parameter for which the BvM Theorem (of the form (1.2) or (1.3)) holds. The authors concentrate on determining how large the dimension of the parameter may be in relation to the sample size while ensuring the BvM Theorem stays true. In contrast, we work with parametric inference and our goal is to provide computable guarantees on the quality of Laplace approximation, of the form of (1.4) or (1.5) and analogous forms for a variety of divergences. We aim for our results to be available to practitioners who approximate an intractable posterior by a Gaussian law and wish to know how good this approximation is. Apart from those key distinctions, there are other important differences between the two papers:

- While the authors of [32] only prove their bounds for the non-informative and the Gaussian prior, our results hold for a variety of priors satisfying mild assumptions.
- Computing the bounds in [32] requires the knowledge of the true parameter, which is the object of inference. Our bounds, on the other hand, are computable without access to it and without access to any statistics of the posterior. This makes our bounds readily applicable and computable in real-life applications, in which the true parameter is unknown and the posterior often intractable.
- The results of [32] provide control over the total variation distance and the difference of means and the norm of the difference of variances. Our bounds also provide control over all those quantities and, in addition, control the difference of expectations of any Lipschitz function (via the 1-Wasserstein distance). Moreover, as we show in Section 4, for one-dimensional problems, we control the difference of expectations of a much larger class of functions.

Similarly, the accuracy of Bernstein–von Mises Gaussian approximation (analogous to (1.2) and (1.3)) in non-parametric models has been studied in [37]. The key assumptions used there are log-concavity of the expectation of the likelihood (or generalized likelihood) with respect to the law of the data and a Gaussian prior. The authors also consider the case of a Gaussian likelihood combined with a more general yet still log-concave prior. We make no such assumptions in the present paper. In fact, Section 5 contains examples of commonly used models involving non-log-concave priors and posteriors, for which our bounds are explicitly computable. The reason we can avoid imposing the assumption of log-concavity is that all we need to control the tail behaviour of the posterior is the assumption of the strict optimality of the MLE or MAP (Assumptions 4 or 8). This is much weaker than assuming log-concavity or strong unimodality of the posterior. Moreover, the authors of [37] compute their bounds explicitly only under further assumptions controlling the third and the fourth derivative of the log-likelihood. In comparison, our setup and techniques only require control of the third derivative.

Furthermore, the recent preprints [21, Section 6.1], [10] and [36] have offered ways of obtaining guarantees on the quality of Laplace approximation under log-concavity of the posterior. In [21, Proposition 6.1] the authors assume that the posterior is strongly log-concave and obtain a computable bound on the 1- and 2-Wasserstein distances between the posterior and the approximating Gaussian. Their bounds thus control the difference of means and the difference of convariances. In [21, Proposition 6.2] the authors relax the assumption on the posterior to weak log-concavity. However, the bound they obtain serves as an indication of the asymptotic rate of convergence and is not computable in practice, for finite data. The main differences between [21, Section 6.1] and our paper are the following:

- Our bounds hold and are fully computable for general posteriors satisfying assumptions analogous to the classical assumptions of the Bernstein von-Mises theorem (see e.g. [29, Section 4] for a recent reference or [14, Section 1.4] for a more classical one). In particular, we do not require (weak or strong) log-concavity of the posterior. Indeed, we compute our bounds explicitly for several examples of commonly used non-log-concave posteriors in Section 5. This is in contrast to [21, Section 6.1].
- Apart from controlling the difference of means and variances, we control additionally the total variation distance, which yields guarantees on an approximate computation of credible sets. This is not provided by [21, Section 6.1].

In [10], the author assumes weak, yet strict log-concavity of the posterior. As their main result [10, Theorem 5], they obtain a bound on the Kullback-Leibler (KL) divergence between the posterior and the Gaussian approximation. In their bound, only the leading terms are computable, while the higher order terms are presented using the big-O notation and not computable in practice. The main differences to our work are the following:

- Our results do not assume log-concavity of the posterior, as explained above and illustrated by the examples in Section 5.
- All of the terms in our bounds are explicitly computable and thus our bounds and fully usable in practice.
- The KL divergence upper-bounded in [10, Theorem 5] is known not to control the difference of means or the difference of variances in general (see [20, Propositions 3.1 and 3.2]). We, in contrast, control the difference of means, via the 1-Wasserstein distance, and the difference of variances, via another suitably chosen integral probability metric.

Moreover, in [36] the author assumes a Gaussian prior and a log-concave likelihood, which yields a strongly log-concave posterior. They relax this assumption only in a very specified case of a nonlinear inverse problem with a certain "warm start condition". Their bounds are on the total variation distance. They also provide a bound on the difference of means - yet this bound involves generic abstract constants and it is not clear to us whether it is computable in applications. Our bounds differ from those derived in [36] in the following ways:

- The array of priors our bounds are available for is much wider than just the Gaussian family. All we require is differentiability, boundedness and boundedness away from zero of the prior density in a small neighbourhood around the MAP or the MLE.
- The array of likelihoods our bounds are available for includes a wide array of commonly used non log-concave ones. As a result, our results work without assuming (weak or strong) log-concavity of the posterior.
- We derive bounds on the Wasserstein-1 distance and a metric controlling the difference of convariances, in addition to bounds on the total variation distance. Our bound on the difference of means is explicit and computable and so are all the other bounds in our paper.

Besides, the recent work [42] derives a Berry-Esseen-type bound on the total variation distance between the posterior and the approximating normal in the approximately linear regression model. The focus and aim of the paper is to control the coverage errors of credible rectangles. The suitable bound obtained by the authors requires access to the true parameter value. Because the focus of our work is on the quality of the Laplace approximation, we apply techniques that yield bounds computable without access to the (unknown) true parameter and valid for a wide variety of models going beyond (approximate) linear regression. Moreover, as already mentioned, we control not only the total variation distance but also other metrics that are important for applications.

In addition, we mention the concurrent work [13] which has provided bounds on the Gaussian approximation of the posterior for i.i.d. data coming from a regular k-parameter exponential family. The results of the present paper do not make any assumption about the true distribution of the data and cover generalized likelihoods not coming from exponential families.

Finally, it is worth noting that bounds similar to ours are not widely available for approximate Bayesian inference techniques in general. Indeed, they are not available at all for the popular variational inference methods [5, 41]. In the area of Markov Chain Monte Carlo methods, progress has recently been made on deriving convergence guarantees for the Unadjusted Langevin Algorithm under different sets of assumptions on the tail growth of the target distribution [7, 2, 11]. However, the popular Metropolis-adjusted Langevin Algorithm and Hamiltonian Monte Carlo are not equipped with such guarantees beyond the case of log-concave targets. Flexible and computable post-hoc checks measuring a discrepancy between the empirical distribution of a sample and the target distribution are given by graph and kernel Stein discrepancies. Graph Stein discrepancies [16, 15] can be computed by solving a linear program. They metrize weak convergence and control the difference of means for distantly-dissipative targets. They are however not known to control the difference of variances. Graph diffusion Stein discrepancies [15] have been shown to posess the same properties under a slightly weaker (yet technical) assumption of the underlying diffusion having a sufficiently rapid Wasserstein decay rate. The fast and popular kernel Stein discrepancies [8, 27, 30, 17] metrize weak convergence for certain choices of kernels and for distantly-dissipative targets. They are, however, currently not known to control the difference of means or the difference of variances. In comparison, we derive control over the rate of weak convergence and the differences of means and variances under interpretable assumptions, which do not require any particular tail behavior of the target posterior.

4. Further results. In this section we present some further results on the quality of Laplace approximation in the *MLE-centric* approach. The proofs of Theorems 4.1-4.3 can be found in Appendices A and C. The proof of Theorem 4.4 can be found in Appendices A and D.

4.1. *Additional notation.* We first introduce additional notation:

$$R_1(n,\delta) := \left(J_n(\hat{\theta}_n) + (\delta M_2/3) I_{d\times d}\right)^{-1}; \quad R_2(n,\delta) := \left(J_n(\hat{\theta}_n) - (\delta M_2/3) I_{d\times d}\right)^{-1};$$

4.2. *Control over the TV distance in the MLE-centric approach.* We start by controlling the total variation distance.

THEOREM 4.1. *Suppose that Assumptions 1-5 hold. Suppose that* $\sqrt{\frac{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}{n}} < \delta$ *and let TV denote the total variation distance. We have the following upper bound:*

$$\begin{aligned}
&TV\left(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right), \mathcal{N}(0, J_n(\hat{\theta}_n)^{-1})\right)\\
\leq & D_1 n^{-1/2} + D_2 n^{d/2} e^{-n\kappa} + 2\exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right],
\end{aligned}$$

where

$$\begin{aligned}
D_1 :=& \frac{\sqrt{3}\, Tr\left[J_n(\hat{\theta}_n)^{-1}\right]\sqrt{\widetilde{M}_1 \hat{M}_1 M_2}}{4\sqrt{\left(\frac{\lambda_{\min}(\hat{\theta}_n)}{M_2} - \delta\right)\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right]\right)}}\\
&+ \frac{M_1\sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{\lambda_{\min}(\hat{\theta}_n) - \delta M_2}};\\
D_2 :=& \frac{2\hat{M}_1 \left|det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left\{1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right\}}.
\end{aligned}$$

4.3. *Control over the* 1*-Wasserstein distance in the MLE-centric approach.*

THEOREM 4.2. *Suppose that Assumptions 1-5 hold. Suppose moreover that* $\sqrt{\frac{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}{n}} < \delta$ *and let* W_1 *denote the 1-Wasserstein distance. We have the following upper bound:*

$$
\begin{aligned}
&W_1\left(\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right), \mathcal{N}(0, J_n(\hat{\theta}_n)^{-1})\right) \\
&\leq E_1 n^{-1/2} + E_3\left[E_2 + \int_{\|u\|>\delta} \|u\| \pi(u + \hat{\theta}_n) du\right] n^{d/2} e^{-n\kappa} \\
&\quad + \left(\delta\sqrt{n} + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}} + E_2\right) \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left(J_n(\hat{\theta}_n)^{-1}\right)}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right],
\end{aligned}
$$

where

$$
\begin{aligned}
E_1 :=& \frac{\sqrt{3}\, Tr\left[J_n(\hat{\theta}_n)^{-1}\right] \widetilde{M}_1 \hat{M}_1}{2\left(\lambda_{\min}(\hat{\theta}_n)/M_2 - \delta\right)\sqrt{\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right]\right)}} \\
&+ \frac{M_1 \widetilde{M}_1 \hat{M}_1}{\lambda_{\min}(\hat{\theta}_n) - \delta M_2}; \\
E_2 :=& \frac{\hat{M}_1 \widetilde{M}_1 \left|det\left(R_1(n,\delta)\right)\right|^{-1/2} \left|det\left(R_2(n,\delta)\right)\right|^{1/2} \sqrt{Tr\left[R_2(n,\delta)\right]}}{1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]}; \\
E_3 :=& \frac{\hat{M}_1 \left|det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.
\end{aligned}
$$

4.4. *Control over the difference of covariances in the MLE-centric approach.*

THEOREM 4.3. *Suppose that Assumptions 1-5 hold. Suppose that* $\sqrt{\frac{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}{n}} < \delta$*. Let* $Z_n \sim \mathcal{N}(0, J_n(\hat{\theta}_n)^{-1})$*. We have that*

$$
\begin{aligned}
&\sup_{v:\|v\|\leq 1}\left|\mathbb{E}\left[\left\langle v, \sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right\rangle^2\right] - \mathbb{E}\left[\langle v, Z_n\rangle^2\right]\right| \\
&\leq (F_1)^2 n^{-1} + F_1 F_3 n^{-1/2} + \frac{\widetilde{M}_1 \left|det(R_2(n,\delta))\right|^{1/2} Tr\left[R_2(n,\delta)\right]}{(2\pi)^{d/2}} (F_2)^2 n^{d/2} e^{-n\kappa} \\
&\quad + \left(\delta^2 n + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}} + \widetilde{M}_1 \left|det(R_2(n,\delta))\right|^{1/2} Tr\left[R_2(n,\delta)\right] F_2\right) \\
&\qquad\qquad \cdot \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{Tr\left(J_n(\hat{\theta}_n)^{-1}\right)}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right] \\
&\quad + \frac{F_2}{(2\pi)^{d/2}}\left[\int_{\|u\|>\delta} \|u\|^2 \pi(u + \hat{\theta}_n) du\right] n^{d/2+1} e^{-n\kappa},
\end{aligned}
$$

where

$$F_1 := \frac{\sqrt{3}\, Tr\left[J_n(\hat{\theta}_n)^{-1}\right] \widetilde{M}_1 \hat{M}_1}{2\left(\lambda_{\min}(\hat{\theta}_n)/M_2 - \delta\right)\sqrt{\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right]\right)}}$$

$$+ \frac{M_1 \widetilde{M}_1 \hat{M}_1}{\lambda_{\min}(\hat{\theta}_n) - \delta M_2};$$

$$F_2 := \frac{\hat{M}_1 \left|det\left(R_1(n,\delta)\right)\right|^{-1/2}}{1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{Tr\left[R_1(n,\delta)\right]}\right)^2 \left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]};$$

$$F_3 := \frac{2\sqrt{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}}{\sqrt{1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{Tr\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right]}}.$$

Finally, we have the following bound, which works for pretty arbitrary test functions, but only in dimension one:

THEOREM 4.4. *Assume that we study a* ***univariate posterior****, i.e. that* $d = 1$*. Let* $\sigma_n^2 := J_n(\hat{\theta}_n)^{-1}$*. Suppose that* $Z_n \sim \mathcal{N}(0, \sigma_n^2)$*. Then, for any function* $g : \mathbb{R} \to \mathbb{R}$*, which is integrable with respect to the posterior and with respect to* $\mathcal{N}(0, \sigma_n^2)$*,*

$$\left|\mathbb{E}\left[g\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)\right)\right] - \mathbb{E}\left[g(Z_n)\right]\right|$$

$$\leq \frac{2\widetilde{M}_1 \hat{M}_1}{\sqrt{2\pi\sigma_n^2}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |ug(u)| \left[\left(M_1 + \frac{3}{\delta}\right) e^{-C_n^{(1)} u^2} - \frac{3}{\delta} e^{-C_n^{(2)} u^2}\right] du \cdot n^{-1/2}$$

$$+ \frac{2\sqrt{C_n^{(4)}}\left(\widetilde{M}_1 \hat{M}_1\right)^2 \left(M_1 + \frac{3}{\delta}\right) \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |g(u)| e^{-C_n^{(2)} u^2} du}{C_n^{(1)} \pi \sqrt{\sigma_n^2}\left(1 - 2e^{-\delta^2 n C_n^{(4)}}\right)} \left(\frac{M_1 + \frac{3}{\delta}}{C_n^{(1)}} - \frac{3}{\delta C_n^{(2)}}\right) \cdot n^{-1/2}$$

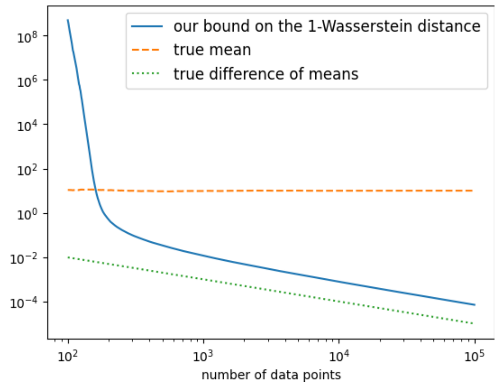
$$+ \left|\int_{|u|>\delta\sqrt{n}} g(u) \frac{e^{-u^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} du\right| + \frac{\hat{M}_1 \sqrt{C_n^{(4)}} \int_{|u|>\delta} |g(u\sqrt{n})| \pi(u + \hat{\theta}_n) du}{\sqrt{\pi}\left\{1 - 2\exp\left[-C_n^{(4)} \delta^2 n\right]\right\}} \cdot n^{1/2} e^{-n\kappa}$$

$$+ \frac{\hat{M}_1 \widetilde{M}_1 \sqrt{C_n^{(4)}} \int_{|t| \leq \delta\sqrt{n}} |g(t)| e^{-C_n^{(2)} t^2} dt}{\sqrt{\pi}\left\{1 - 2\exp\left[-C_n^{(4)} \delta^2 n\right]\right\}} \left[2e^{-\delta^2 n/(2\sigma_n^2)} + \frac{\hat{M}_1 \sqrt{C_n^{(4)}} n^{1/2} e^{-n\kappa}}{\sqrt{\pi}\left(1 - 2\exp\left[-C_n^{(4)} \delta^2 n\right]\right)}\right],$$

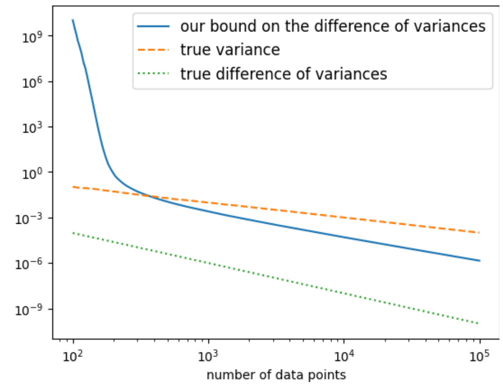
for

$$C_n^{(1)} = \frac{1}{2\sigma_n^2} - \frac{\delta M_2}{3} > 0; \qquad C_n^{(2)} = \left(\frac{1}{2\sigma_n^2} - \frac{\delta M_2}{6}\right) > 0;$$

$$C_n^{(3)} = \left(\frac{1}{2\sigma_n^2} + \frac{\delta M_2}{3}\right) > 0; \qquad C_n^{(4)} = \left(\frac{1}{2\sigma_n^2} + \frac{\delta M_2}{6}\right) > 0.$$



(a) Posterior mean



(b) Posterior variance

Fig 1: Poisson likelihood with gamma prior and exponential data (MAP-centric approach)

REMARK 4.5. The proof of Theorem 4.4 can easily be modified in order to yield analogous bounds on the quality of approximation in the MAP-centric approach, as described in Section 1.4 and presented in Section 2.

5. Example applications.

Now, we present some examples of our bounds computed for different models.

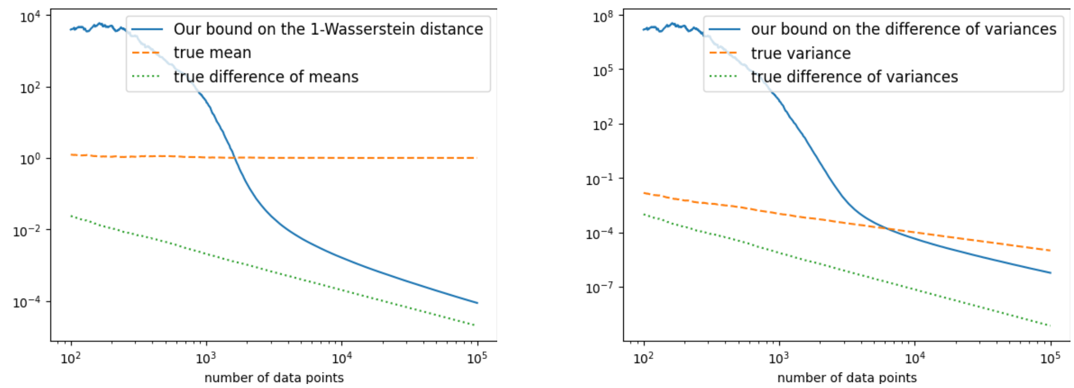
5.1. *Our bounds work under misspecification: Poisson likelihood with gamma prior and exponential data.* First, we look at a one-dimensional conjugate model, for which we can compare our bounds on the difference of means and the difference of variances to the ground truth. We consider a Poisson likelihood and a gamma prior with shape equal to 0.1 and rate equal to 3. Our data are generated from the exponential distribution with mean 10. Figures 1a and 1b show good control over the difference of means and the difference of variances provided by our bounds (in the MAP-centric approach) for sample sizes between around 200. More detail on how one can compute the constants appearing in the bounds can be found in Appendix E.

5.2. *Our bounds work for non log-concave posteriors: Weibull likelihood with inverse-gamma prior.* Now, we consider another conjugate model and compare our bounds to the ground truth. In this case, the posterior is not log-concave. In our experiment, we set the shape of the Weibull to $\frac{1}{2}$ and we make inference about the scale. The prior is inverse-gamma with shape equal to 3 and scale equal to 10. The data are Weibull with shape $1/2$ and scale 1. Figures 2a and 2b demonstrate that our bounds on the difference of means and the difference of variances (in the MAP centric approach) get close to the true difference of means and the true difference of variances for sample sizes in low thousands. More detail on how one can compute the constants appearing in the bounds can be found in Appendix E.

5.3. *Our bounds work for multivariate heavy-tailed posteriors: logistic regression with Student's t prior.*

5.3.1. *Setup.* Suppose $X_1, \ldots, X_n \in \mathbb{R}^d$ and $Y_1, \ldots, Y_n \in \{-1, 1\}$. We will study the following log-likelihood:

$$L_n(\theta) := L_n\left(\theta | \left(Y_i\right)_{i=1}^n, \left(X_i\right)_{i=1}^n\right) = -\sum_{i=1}^n \log\left(1 + e^{-X_i^T \theta Y_i}\right). \tag{5.1}$$



(a) Posterior mean

(b) Posterior variance

Fig 2: Weibull likelihood with inverse-gamma prior (MAP-centric approach)

For a covariance matrix Σ, a vector $\mu \in \mathbb{R}^d$ and hyperparameter $\nu > 0$, we consider a d-dimensional Student's t prior on θ, given by

$$\pi(\theta) = \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}} \left[1 + \frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right]^{-(\nu+d)/2}. \tag{5.2}$$

Combining the log-likelihood given by (5.1) with the prior given by (5.2) yields a posterior that is known not to be log-concave. Moreover, for certain data sets the posterior is heavy tailed. An easy one-dimensional example of such a data-set is one for which $X_iY_i > 0$ for all i. Let us first focus on calculating the constants appearing in the bounds of Section 2.

5.3.2. *Calculating $J_n(\hat{\theta}_n)$ and $\bar{J}_n(\bar{\theta}_n)$.* Note that:

$$J_n(\hat{\theta}_n) = \frac{1}{n}\sum_{k=1}^{n} \frac{e^{X_k^T\hat{\theta}_nY_k}}{(1+e^{X_k^T\hat{\theta}_nY_k})^2} X_k\,(X_k)^T\,;$$

$$\begin{aligned}\bar{J}_n(\bar{\theta}_n) =& \frac{1}{n}\sum_{k=1}^{n} \frac{e^{X_k^T\bar{\theta}_nY_k}}{(1+e^{X_k^T\bar{\theta}_nY_k})^2} X_k\,(X_k)^T \\ &+ \frac{\nu+d}{2\nu n}\left[1+\frac{1}{\nu}(\bar{\theta}_n-\mu)^T\Sigma^{-1}(\bar{\theta}_n-\mu)\right]^{-1}\left(\Sigma^{-1}+\operatorname{diag}\left(\Sigma^{-1}\right)\right) \\ &- \frac{\nu+d}{2\nu^2 n}\left[1+\frac{1}{\nu}(\bar{\theta}_n-\mu)^T\Sigma^{-1}(\bar{\theta}_n-\mu)\right]^{-2} \\ &\qquad \cdot\left[\left(\Sigma^{-1}+\operatorname{diag}\left(\Sigma^{-1}\right)\right)(\bar{\theta}_n-\mu)\right]\left[\left(\Sigma^{-1}+\operatorname{diag}\left(\Sigma^{-1}\right)\right)(\bar{\theta}_n-\mu)\right]^T.\end{aligned}$$

5.3.3. *Calculating M_1, $\widetilde{M}_1$ and $\hat{M}_1$.* Note that

$$\begin{aligned}\pi'(\theta) =& \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}}\left(-\frac{\nu+d}{2}\right)\left[1+\frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right]^{-(\nu+d)/2-1} \\ &\cdot\left[\frac{1}{\nu}\left(\Sigma^{-1}+\operatorname{diag}\left(\Sigma^{-1}_{1,1},\ldots,\Sigma^{-1}_{d,d}\right)\right)\right](\theta-\mu).\end{aligned} \tag{5.3}$$

It follows from the expressions (5.2) and (5.3) that

$$\sup_{\theta:\|\theta-\hat{\theta}\|<\delta}|\pi(\theta)| \leq \frac{\Gamma((\nu+d)/2)}{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}} =: \widetilde{M}_1;$$

$$\sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\frac{1}{|\pi(\theta)|} \leq \sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\frac{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}}{\Gamma((\nu+d)/2)}\left[1+\frac{1}{\nu\lambda_{\min}(\Sigma)}\|\theta-\mu\|^2\right]^{(\nu+d)/2}$$

$$\leq \frac{\Gamma(\nu/2)\nu^{d/2}\pi^{d/2}|\Sigma|^{1/2}}{\Gamma((\nu+d)/2)}\left[1+\frac{2\delta^2+2\|\hat{\theta}-\mu\|^2}{\nu\lambda_{\min}(\Sigma)}\right]^{(\nu+d)/2} =: \hat{M}_1;$$

$$\sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\frac{\|\pi'(\theta)\|}{|\pi(\theta)|} = \sup_{\theta:\|\theta-\hat{\theta}\|<\delta}\left(\frac{\nu+d}{2}\right)\frac{\left\|\frac{1}{\nu}\left(\Sigma^{-1}+\mathrm{diag}\left(\Sigma^{-1}_{1,1},\ldots,\Sigma^{-1}_{d,d}\right)\right)(\theta-\mu)\right\|}{\left[1+\frac{1}{\nu}(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)\right]}$$

$$\leq \left(\frac{\nu+d}{2}\right)\frac{\left(\delta+\|\hat{\theta}-\mu\|\right)\left\|\Sigma^{-1}+\mathrm{diag}\left(\Sigma^{-1}_{1,1},\ldots,\Sigma^{-1}_{d,d}\right)\right\|}{\nu}$$

$$\leq \frac{(\nu+d)\left(\delta+\|\hat{\theta}-\mu\|\right)}{\nu\lambda_{\min}(\Sigma)} =: M_1.$$

5.3.4. *Calculating M_2 and $\overline{M}_2$.* Note that, for all $\theta\in\mathbb{R}^d$,

$$L_n'''(\theta)[u_1,u_2,u_3] = \sum_{i=1}^n Y_i^3 \sum_{j,k,l=1}^d \frac{e^{X_i^T\theta Y_i}\left(e^{X_i^T\theta Y_i}-1\right)}{\left(1+e^{X_i^T\theta Y_i}\right)^3} X_i^{(j)}X_i^{(k)}X_i^{(l)}u_1^{(j)}u_1^{(k)}u_1^{(l)}$$

and therefore, for all $\theta\in\mathbb{R}^d$,

$$\frac{1}{n}\|L_n'''(\theta)\| \leq \frac{1}{n}\sum_{k=1}^n\|X_k\|^3\frac{e^{X_k^T\theta Y_k}\left|e^{X_k^T\theta Y_k}-1\right|}{\left(1+e^{X_k^T\theta Y_k}\right)^3} \leq \frac{1}{6\sqrt{3}\,n}\sum_{k=1}^n\|X_k\|^3 =: M_2. \tag{5.4}$$

Now, a straightforward calculation reveals that, for $\|u\|\leq 1, \|v\|\leq 1, \|w\|\leq 1$ and any $\theta\in\mathbb{R}^d$, and $\bar{\delta}\leq 1$,

$$\sup_{\|\theta-\bar{\theta}_n\|\leq\bar{\delta}}\left|\sum_{i,j,k=1}^d\left(\frac{\partial^3}{\partial\theta_j\partial\theta_i\partial\theta_k}\log\pi(\theta)\right)u_iv_jw_k\right|$$

$$\leq \sup_{\|\theta-\bar{\theta}_n\|\leq\bar{\delta}}\left\{\frac{3(\nu+d)}{[\nu+(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)]^2}\left\|\Sigma^{-1}+\mathrm{diag}(\Sigma^{-1})\right\|_{op}^2\|\theta-\mu\|\right.$$

$$\left.+\frac{2(\nu+d)}{[\nu+(\theta-\mu)^T\Sigma^{-1}(\theta-\mu)]^3}\left\|\Sigma^{-1}+\mathrm{diag}(\Sigma^{-1})\right\|_{op}^3\|\theta-\mu\|^3\right\}$$

$$\leq \frac{3(\nu+d)}{\nu^2}\left\|\Sigma^{-1}+\mathrm{diag}(\Sigma^{-1})\right\|_{op}^2\left(1+\|\bar{\theta}_n-\mu\|\right)$$

$$+\frac{2(\nu+d)}{\nu^2}\left\|\Sigma^{-1}+\mathrm{diag}(\Sigma^{-1})\right\|_{op}^3\left(1+\|\bar{\theta}_n-\mu\|\right)^3.$$

Combining this with (5.4), we obtain

$$\frac{1}{n}\left\|\overline{L}_n'''(\bar\theta_n)\right\| \leq \frac{1}{6\sqrt{3}\,n}\sum_{k=1}^{n}\|X_k\|^3 + \frac{3(\nu+d)}{\nu^2\,n}\left\|\Sigma^{-1}+\text{diag}(\Sigma^{-1})\right\|_{op}^2\left(1+\|\bar\theta_n-\mu\|\right)$$

$$+\frac{2(\nu+d)}{\nu^3\,n}\left\|\Sigma^{-1}+\text{diag}(\Sigma^{-1})\right\|_{op}^3\left(1+\|\bar\theta_n-\mu\|\right)^3 =: \overline{M}_2. \tag{5.5}$$

5.3.5. *Calculating κ and $\bar\kappa$.* Note that L_n is strictly concave. Therefore,

$$\begin{aligned}\sup_{\theta:\|\theta-\hat\theta_n\|>\delta}\frac{L_n(\theta)-L_n(\hat\theta)}{n} &\leq \sup_{\theta:\|\theta-\hat\theta_n\|=\delta}\frac{L_n(\theta)-L_n(\hat\theta)}{n}\\ &\leq \sup_{\theta:\|\theta-\hat\theta_n\|=\delta}\left\{-\frac{1}{2}\left(\theta-\hat\theta_n\right)^T J_n(\hat\theta_n)\left(\theta-\hat\theta_n\right)\right\}+\frac{M_2\delta^3}{2}\\ &\leq -\frac{1}{2}\lambda_{\min}(\hat\theta_n)\delta^2+\frac{M_2\delta^3}{2} =: -\kappa.\end{aligned}$$

Since M_2 in (5.4) provides a uniform bound on $\frac{1}{n}\|L_n'''(\theta)\|$ over $\theta\in\mathbb{R}^d$, a similar calculation shows:

$$\begin{aligned}&\sup_{\theta:\|\theta-\hat\theta_n\|>\bar\delta-\|\hat\theta_n-\bar\theta_n\|}\frac{L_n(\theta)-L_n(\hat\theta_n)}{n}\\ &\leq -\frac{1}{2}\lambda_{\min}(\hat\theta_n)\left(\bar\delta-\|\hat\theta_n-\bar\theta_n\|\right)^2+\frac{M_2\left(\bar\delta-\|\hat\theta_n-\bar\theta_n\|\right)^3}{2} =: -\bar\kappa.\end{aligned}$$

5.3.6. *Finding appropriate values of δ and $\bar\delta$.* In order to apply our results in the MLE-centric approach, we need

$$\sqrt{\frac{\text{Tr}\left[J_n(\hat\theta_n)^{-1}\right]}{n}} < \delta < \frac{\lambda_{\min}(\hat\theta_n)}{M_2}. \tag{5.6}$$

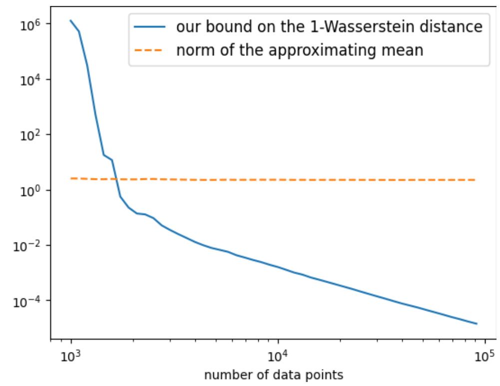
This assumption also ensures that κ from Subsection 5.3.5 is positive. In the MAP-centric approach also require

$$\max\left\{\|\hat\theta_n-\bar\theta_n\|,\sqrt{\frac{\text{Tr}\left(\bar J_n(\bar\theta_n)^{-1}\right)}{n}}\right\} < \bar\delta < \frac{\overline{\lambda}_{\min}(\bar\theta_n)}{\overline{M}_2},$$

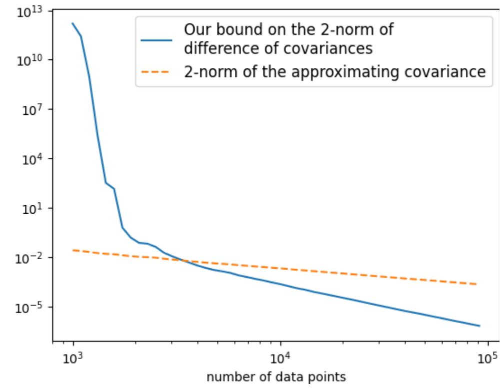
which also ensures that $\bar\kappa$ from Subsection 5.3.5 is positive. Such choices of $\bar\delta$ and δ will be available for sufficiently large n. In order to choose the appropriate concrete values of $\bar\delta$ and δ one can run a numerical optimization scheme.

5.3.7. *Summing up.* The bounds from Sections 2 and 4 may be computed using $M_2, \overline{M}_2, \kappa, \bar\kappa$ derived above. The MLE $\hat\theta_n$ and $\bar\theta_n$ can be easily obtained numerically, using built-in R or Python packages for global optimization. As there is a certain degree of choice for δ and $\bar\delta$, the choice therof may also be optimized numerically, using the same packages.

In order to improve on the bounds, one can also run a numerical optimizer in order to derive tighter values of M_2 and $\overline{M}_2$ than those we derive analytically here. A robust (yet slow) approach to doing this is via grid search, combined with the mean value theorem and a bound



(a) Posterior mean



(b) Posterior covariance

Fig 3: 5-dimensional logistic regression with t prior (MAP-centric approach)

on the fourth derivative of the log-likelihood (or log-posterior). Another option is to run a faster built-in global optimizer to derive the maximum of the third derivative inside a ball around the MLE (or MAP). What is gained in terms of speed is lost in terms of robustness. Nevertheless, this might be a useful approach for users of our bounds. Figures 3a and 3b present our bounds obtained using this approach. The experiments were performed for the 5-dimensional logistic regression with Student's t prior with mean zero and identity convariance matrix. The data we used came from logistic regression with parameter $(1, 1, 1, 1, 1)$. Figures 3a and 3b demonstrate that our bounds on the 1-Wasserstein distance and the 2-norm of the difference of variances go well below the approximate values of the mean and 2-norm of the covariance, respectively, for reasonably moderate sample sizes. Therefore, our bounds go well below the true values of the mean and the norm of the covariance, for moderate sample sizes. This indicates that they are applicable for practitioners who wish to assess how confident they should be in their mean and variance estimates.

6. Conclusions and future work. We provide bounds on the quality of the Laplace approximation which are computable and hold under the standard assumptions of the Bernstein – von Mises Theorem. A crucial question is one about the tightness of our bounds. Our proof technique relies mainly on using the log-Sobolev inequality inside a ball around the MLE or MAP. It would be useful to investigate whether using the log-Sobolev inequality inside another convex region around the MLE/MAP could produce tighter bounds. A crucial question would also be whether this alternative convex region would still be such that for many commonly used Bayesian models the posterior satisfies the log-Sobolev inequality inside it. Another interesting question is whether, for multivariate posteriors, we could derive bounds on more general integral probability metrics, in a way similar to our univariate Theorem 4.4. This is to a certain extent a question about the applicability of Stein's method in dimension greater than one for measures truncated to a bounded convex set. So far we have struggled to find enough theory that would support it but we hope such theory will be provided in the future.

Acknowledgements. The authors are grateful to Krishna Balasubramanian, Xiao Fang, Max Fathi, Jonathan Huggins, Tin D. Nguyen, Jeff Miller, Daniel Paulin, Gesine Reinert, Håvard Rue, Adil Salim and Yvik Swan for interesting discussions. This work is part of project **Stein-ML** that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement **No 101024264**. Moreover, this work was supported in part by an NSF Career Award and an ONR Early Career Grant.

REFERENCES

[1] BACH, F. (2021). Approximating integrals with Laplace's method. https://francisbach.com/laplace-method/. Accessed: 2022-08-29.

[2] BALASUBRAMANIAN, K., CHEWI, S., ERDOGDU, M. A., SALIM, A. and ZHANG, S. (2022). Towards a Theory of Non-Log-Concave Sampling:First-Order Stationarity Guarantees for Langevin Monte Carlo. In *Proceedings of Thirty Fifth Conference on Learning Theory* (P.-L. LOH and M. RAGINSKY, eds.). *Proceedings of Machine Learning Research* **178** 2896–2923. PMLR.

[3] BISHOP, C. M. (2006). *Pattern Recognition and Machine Learning*. *Information Science and Statistics*. Springer New York, NY.

[4] BISSIRI, P. G., HOLMES, C. C. and WALKER, S. G. (2016). A general framework for updating belief distributions. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **78** 1103–1130.

[5] BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational Inference: A Review for Statisticians. *Journal of the American Statistical Association* **112** 859-877.

[6] CHERNOZHUKOV, V. and HONG, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics* **115** 293-346.

[7] CHEWI, S., ERDOGDU, M. A., LI, M., SHEN, R. and ZHANG, S. (2022). Analysis of Langevin Monte Carlo from Poincare to Log-Sobolev. In *Proceedings of Thirty Fifth Conference on Learning Theory* (P.-L. LOH and M. RAGINSKY, eds.). *Proceedings of Machine Learning Research* **178** 1–2. PMLR.

[8] CHWIALKOWSKI, K., STRATHMANN, H. and GRETTON, A. (2016). A Kernel Test of Goodness of Fit. In *Proceedings of The 33rd International Conference on Machine Learning* (M. F. BALCAN and K. Q. WEINBERGER, eds.). *Proceedings of Machine Learning Research* **48** 2606–2615. PMLR, New York, New York, USA.

[9] DAXBERGER, E., KRISTIADI, A., IMMER, A., ESCHENHAGEN, R., BAUER, M. and HENNIG, P. (2021). Laplace Redux - Effortless Bayesian Deep Learning. In *Advances in Neural Information Processing Systems* (M. RANZATO, A. BEYGELZIMER, Y. DAUPHIN, P. S. LIANG and J. W. VAUGHAN, eds.) **34** 20089–20103. Curran Associates, Inc.

[10] DEHAENE, G. (2019). A deterministic and computable Bernstein-von Mises theorem. arXiv:1904.02505v2.

[11] ERDOGDU, M. A., HOSSEINZADEH, R. and ZHANG, S. (2022). Convergence of Langevin Monte Carlo in Chi-Squared and Rényi Divergence. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics* (G. CAMPS-VALLS, F. J. R. RUIZ and I. VALERA, eds.). *Proceedings of Machine Learning Research* **151** 8151–8175. PMLR.

[12] ERNST, M., REINERT, G. and SWAN, Y. (2020). First-order covariance inequalities via Stein's method. *Bernoulli* **26** 2051 – 2081.

[13] FISCHER, A., GAUNT, R. E., REINERT, G. and SWAN, Y. (2022). Normal approximation for the posterior in exponential families. arXiv:2209.08806v1.

[14] GHOSH, J. K. and RAMAMOORTHI, R. V. (2003). *Bayesian Nonparametrics*. Springer Science+Business Media.

[15] GORHAM, J., DUNCAN, A. B., VOLLMER, S. J. and MACKEY, L. (2019). Measuring sample quality with diffusions. *The Annals of Applied Probability* **29** 2884 – 2928.

[16] GORHAM, J. and MACKEY, L. (2015). Measuring Sample Quality with Stein's Method. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*. *NIPS'15* 226–234. MIT Press, Cambridge, MA, USA.

[17] GORHAM, J. and MACKEY, L. (2017). Measuring Sample Quality with Kernels. In *Proceedings of the 34th International Conference on Machine Learning* (D. PRECUP and Y. W. TEH, eds.). *Proceedings of Machine Learning Research* **70** 1292–1301. PMLR.

[18] GOZLAN, N. (2009). A characterization of dimension free concentration in terms of transportation inequalities. *The Annals of Probability* **37** 2480 – 2498.

[19] HSU, D., KAKADE, S. and ZHANG, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability* **17** 1 – 6.

[20] HUGGINS, J., KASPRZAK, M., CAMPBELL, T. and BRODERICK, T. (2020). Validated Variational Inference via Practical Posterior Error Bounds. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics* (S. CHIAPPA and R. CALANDRA, eds.). *Proceedings of Machine Learning Research* **108** 1792–1802. PMLR.

[21] HUGGINS, J. H., KASPRZAK, M., CAMPBELL, T. and BRODERICK, T. (2018). Practical Bounds on the Error of Bayesian Posterior Approximations: a Nonasymptotic Approach. arXiv:1809.09505v2.
[22] IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. *Stochastic Modelling and Applied Probability*. Springer New York, NY.
[23] KLEIJN, B. J. K. and VAN DER VAART, A. W. (2012). The Bernstein-Von-Mises theorem under misspecification. *Electronic Journal of Statistics* **6** 354 – 381.
[24] LAPLACE, P. S. (1774). Mémoire sur la probabilité des causes par les événements. In *Mémoires de l'Académie royale des sciences de Paris (Savants étrangers)* 621–656.
[25] LAPLACE, P. S. (1986). Memoir on the Probability of the Causes of Events. *Statistical Science* **1** 364–378.
[26] LINDGREN, F. and RUE, H. (2015). Bayesian Spatial Modelling with R-INLA. *Journal of Statistical Software* **63** 1–25.
[27] LIU, Q., LEE, J. and JORDAN, M. (2016). A Kernelized Stein Discrepancy for Goodness-of-fit Tests. In *Proceedings of The 33rd International Conference on Machine Learning* (M. F. BALCAN and K. Q. WEINBERGER, eds.). *Proceedings of Machine Learning Research* **48** 276–284. PMLR, New York, New York, USA.
[28] MASSART, P. (2007). Concentration inequalities and model selection. In *Ecole d'Eté de Probabilités de Saint-Flour XXXIII - 2003*, (J. Picard, ed.). *Lecture Notes in Mathematics* Springer Berlin, Heidelberg.
[29] MILLER, J. W. (2021). Asymptotic Normality, Concentration, and Coverage of Generalized Posteriors. *Journal of Machine Learning Research* **22** 1–53.
[30] OATES, C. J., GIROLAMI, M. and CHOPIN, N. (2017). Control functionals for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **79** 695-718.
[31] OTTO, F. and VILLANI, C. (2000). Generalization of an Inequality by Talagrand and Links with the Logarithmic Sobolev Inequality. *Journal of Functional Analysis* **173** 361-400.
[32] PANOV, M. and SPOKOINY, V. (2015). Finite Sample Bernstein – von Mises Theorem for Semiparametric Problems. *Bayesian Analysis* **10** 665 – 710.
[33] RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71** 319-392.
[34] SAUMARD, A. and WELLNER, J. A. (2014). Log-concavity and strong log-concavity: A review. *Statistics Surveys* **8** 45 – 114.
[35] SCHLICHTING, A. (2019). Poincaré and Log-Sobolev Inequalities for Mixtures. *Entropy* **21**.
[36] SPOKOINY, V. (2022). Dimension free non-asymptotic bounds on the accuracy of high dimensional Laplace approximation. arXiv:2204.11038v3.
[37] SPOKOINY, V. and PANOV, M. (2020). Accuracy of Gaussian approximation in nonparametric Bernsteiv – von Mises Theorem. arXiv:1910.06028v5.
[38] VAN DER VAART, A. W. (1998). *Asymptotic statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.
[39] VERSHYNIN, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.
[40] VILLANI, C. (2009). *Optimal Transport Old and New*. *Grundlehren der mathematischen Wissenschaften*. Springer Berlin, Heidelberg.
[41] WAINWRIGHT, M. J. and JORDAN, M. I. (2008). Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends® in Machine Learning* **1** 1-305.
[42] YANO, K. and KATO, K. (2020). On frequentist coverage errors of Bayesian credible sets in moderately high dimensions. *Bernoulli* **26** 616 – 641.

APPENDIX A: INTRODUCTORY ARGUMENTS

A.1. Introduction. In order to prove the results of this paper, we let $g : \mathbb{R}^d \to \mathbb{R}$ and consider:

$$D_g^{MLE} := \left| \mathbb{E}\left[g\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n) \right) \right] - \mathbb{E}_{Z_n \sim \mathcal{N}(0, J_n(\hat{\theta}_n))^{-1})} \left[g(Z_n) \right] \right| ; \tag{A.1}$$

$$D_g^{MAP} := \left| \mathbb{E}\left[g\left(\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n) \right) \right] - \mathbb{E}_{\bar{Z}_n \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n))^{-1})} \left[g(\bar{Z}_n) \right] \right| . \tag{A.2}$$

The following lemma will be useful in the sequel:

LEMMA A.1. *Let $Z \sim \mathcal{N}(0,\Sigma)$. Then, for any $t>0$,*

$$\mathbb{P}\left[\|Z\| - \sqrt{Tr(\Sigma)} \geq \sqrt{2\|\Sigma\|_{op}\, t}\right] \leq e^{-t}; \tag{A.3}$$

$$\mathbb{P}\left[\|Z\|^2 - Tr(\Sigma) \geq 2\sqrt{Tr\left(\Sigma^2\right) t} + 2\|\Sigma\|_{op}\, t\right] \leq e^{-t}. \tag{A.4}$$

PROOF. The proof follows an argument similar to the one of [39, page 135]. Specifically, (A.4) comes from [19, Proposition 1]. In order to prove (A.3), we note that

$$\begin{aligned}\mathbb{P}\left[\|Z\| - \sqrt{\mathrm{Tr}(\Sigma)} \geq \sqrt{2\|\Sigma\|_{op}\, t}\right] =& \mathbb{P}\left[\|Z\|^2 - \mathrm{Tr}(\Sigma) \geq 2\|\Sigma\|_{op}\, t + 2\sqrt{2\mathrm{Tr}\left(\Sigma\right)\|\Sigma\|_{op}\, t}\right] \\ \leq& \mathbb{P}\left[\|Z\|^2 - \mathrm{Tr}(\Sigma) \geq 2\sqrt{\mathrm{Tr}\left(\Sigma^2\right) t} + 2\|\Sigma\|_{op}\, t\right] \leq e^{-t}.\end{aligned}$$

□

A.2. Initial decomposition of the distances D_g^{MLE} and D_g^{MAP}. Now, let

$$h_g^{MLE}(u) := g(u) - \frac{n^{-d/2}}{C_n^{MLE}} \int_{\|t\|<\delta\sqrt{n}} g(t)\Pi_n(n^{-1/2}t + \hat{\theta}_n)dt,$$

$$h_g^{MAP}(u) := g(u) - \frac{n^{-d/2}}{C_n^{MAP}} \int_{\|t\|<\bar{\delta}\sqrt{n}} g(t)\Pi_n(n^{-1/2}t + \bar{\theta}_n)dt,$$

for

$$C_n^{MLE} := n^{-d/2} \int_{\|u\|\leq\delta\sqrt{n}} \Pi_n(n^{-1/2}u + \hat{\theta}_n)du \tag{A.5}$$

$$C_n^{MAP} := n^{-d/2} \int_{\|u\|\leq\bar{\delta}\sqrt{n}} \Pi_n(n^{-1/2}u + \bar{\theta}_n)du.$$

Note that, for

$$F_n^{MLE} := \int_{\|u\|\leq\delta\sqrt{n}} \frac{\sqrt{|\det J_n(\hat{\theta}_n)|}e^{-u^T J_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du, \tag{A.6}$$

we have

$$\begin{aligned} &D_g^{MLE} \\ =& \left| \int_{\mathbb{R}^d} h_g^{MLE}(u) \frac{\sqrt{|\det J_n(\hat{\theta}_n)|}e^{-u^T J_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du \right. \\ & \left. \qquad - n^{-d/2} \int_{\|u\|>\delta\sqrt{n}} h_g^{MLE}(u)\Pi_n(n^{-1/2}u + \hat{\theta}_n)du \right| \\ \leq& \frac{1}{F_n^{MLE}} \left| \int_{\|u\|\leq\delta\sqrt{n}} h_g^{MLE}(u) \frac{\sqrt{|\det J_n(\hat{\theta}_n)|}e^{-u^T J_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} du \right| \\ &+ \left| \int_{\|u\|>\delta\sqrt{n}} h_g^{MLE}(u) \left[\frac{\sqrt{|\det J_n(\hat{\theta}_n)|}e^{-u^T J_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}} - n^{-d/2}\Pi_n(n^{-1/2}u + \hat{\theta}_n) \right] du \right| \end{aligned}$$

$$= \left| \int_{\|u\|\le\delta\sqrt{n}} g(u) \frac{\sqrt{|\det J_n(\hat\theta_n)|} e^{-u^T J_n(\hat\theta_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}} du \right.$$

$$\left. - \frac{n^{-d/2}}{C_n^{MLE}} \int_{\|u\|\le\delta\sqrt{n}} g(u)\Pi_n(n^{-1/2}u+\hat\theta_n)du \right|$$

$$+ \left| \int_{\|u\|>\delta\sqrt{n}} h_g^{MLE}(u) \left[\frac{\sqrt{|\det J_n(\hat\theta_n)|} e^{-u^T J_n(\hat\theta_n)u/2}}{(2\pi)^{d/2}} - n^{-d/2}\,\Pi_n(n^{-1/2}u+\hat\theta_n) \right] du \right|$$

(A.7)

$$=: I_1^{MLE} + I_2^{MLE}.$$

In a similar manner, for

$$F_n^{MAP} := \int_{\|u\|\le\bar\delta\sqrt{n}} \frac{\sqrt{|\det \bar J_n(\bar\theta_n)|} e^{-u^T \bar J_n(\bar\theta_n)u/2}}{(2\pi)^{d/2}} du, \tag{A.8}$$

we have

$$D_g^{MAP}$$

$$\le \left| \int_{\|u\|\le\bar\delta\sqrt{n}} g(u) \frac{\sqrt{|\det \bar J_n(\bar\theta_n)|} e^{-u^T \bar J_n(\bar\theta_n)u/2}}{F_n^{MAP}(2\pi)^{d/2}} du \right.$$

$$\left. - \frac{n^{-d/2}}{C_n^{MAP}} \int_{\|u\|\le\bar\delta\sqrt{n}} g(u)\Pi_n(n^{-1/2}u+\bar\theta_n)du \right|$$

$$+ \left| \int_{\|u\|>\bar\delta\sqrt{n}} h_g^{MAP}(u) \left[\frac{\sqrt{|\det \bar J_n(\bar\theta_n)|} e^{-u^T \bar J_n(\bar\theta_n)u/2}}{(2\pi)^{d/2}} - n^{-d/2}\,\Pi_n(n^{-1/2}u+\bar\theta_n) \right] du \right|$$

(A.9)

$$=: I_1^{MAP} + I_2^{MAP}.$$

A.3. Controlling term I_2^{MLE}.

Note that

$$I_2^{MLE} \le \left| \int_{\|u\|>\delta\sqrt{n}} g(u) \left(n^{-d/2}\Pi_n(n^{-1/2}u+\hat\theta_n) - \frac{\sqrt{\left|\det J_n(\hat\theta_n)\right|} e^{-u^T J_n(\hat\theta_n)u/2}}{(2\pi)^{d/2}} \right) du \right|$$

$$+ \frac{n^{-d/2}}{C_n^{MLE}} \int_{\|t\|\le\delta\sqrt{n}} |g(t)|\Pi_n(n^{-1/2}t+\hat\theta_n)dt$$

$$\cdot \left| \int_{\|u\|>\delta\sqrt{n}} \left(n^{-d/2}\Pi_n(n^{-1/2}u+\hat\theta_n) - \frac{\sqrt{\left|\det J_n(\hat\theta_n)\right|} e^{-u^T J_n(\hat\theta_n)u/2}}{(2\pi)^{d/2}} \right) du \right|$$

$$=: I_{2,1}^{MLE} + I_{2,2}^{MLE}.$$

Now, for $\|t\| < \delta\sqrt{n}$, Assumption 1 implies that

$$\left| L_n(n^{-1/2}t + \hat{\theta}_n) - L_n(\hat{\theta}_n) + \frac{1}{2}t^T I(\hat{\theta}_n)t \right| \le \frac{1}{6}n^{-1/2}M_2\|t\|^3 \le \frac{\delta M_2}{6}\|t\|^2. \tag{A.10}$$

Therefore, under Assumptions 1, 2 and 4, and using the notation of Section 2.1,

$$\begin{aligned}
&\int_{\|u\|>\delta\sqrt{n}} |g(u)| n^{-d/2}\Pi_n(n^{-1/2}u + \hat{\theta}_n)du \\
&= \frac{\int_{\|u\|>\delta\sqrt{n}} |g(u)|\pi(n^{-1/2}u + \hat{\theta}_n)e^{L_n(n^{-1/2}u+\hat{\theta}_n)}du}{\int_{\mathbb{R}^d} \pi(n^{-1/2}t + \hat{\theta}_n)e^{L_n(n^{-1/2}t+\hat{\theta}_n)}dt} \\
&\le \frac{\int_{\|u\|>\delta\sqrt{n}} |g(u)|\pi(n^{-1/2}u + \hat{\theta}_n)e^{L_n(n^{-1/2}u+\hat{\theta}_n)}du}{\int_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t + \hat{\theta}_n)e^{L_n(n^{-1/2}t+\hat{\theta}_n)}dt} \\
&= \frac{\int_{\|u\|>\delta\sqrt{n}} |g(u)|\pi(n^{-1/2}u + \hat{\theta}_n)e^{L_n(n^{-1/2}u+\hat{\theta}_n)-L_n(\hat{\theta}_n)}du}{\int_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t + \hat{\theta}_n)e^{L_n(n^{-1/2}t+\hat{\theta}_n)-L_n(\hat{\theta}_n)}dt} \\
&\overset{(A.10)}{\le} \frac{e^{-n\kappa}\int_{\|u\|>\delta\sqrt{n}} |g(u)|\pi(n^{-1/2}u + \hat{\theta}_n)du}{\int_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t + \hat{\theta}_n)e^{-t^T(J_n(\hat{\theta}_n)+(\delta M_2/3)I_{d\times d})t/2}dt} \\
&\le \frac{n^{d/2}e^{-n\kappa}\hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|u\|>\delta} |g(u\sqrt{n})|\pi(u + \hat{\theta}_n)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)},
\end{aligned} \tag{A.11}$$

if $n > \frac{\mathrm{Tr}[R_1(n,\delta)]}{\delta^2}$, where the last inequality follows from Lemma A.1. Therefore, if $n > \frac{\mathrm{Tr}[R_1(n,\delta)]}{\delta^2}$ and if Assumptions 1, 2 and 4 are satisfied,

$$\begin{aligned}
I_{2,1}^{MLE} \le &\left| \int_{\|u\|>\delta\sqrt{n}} g(u)\frac{\sqrt{\left|\det J_n(\hat{\theta}_n)\right|}e^{-u^T J_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}}du \right| \\
&+ \frac{n^{d/2}e^{-n\kappa}\hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|u\|>\delta} |g(u\sqrt{n})|\pi(u + \hat{\theta}_n)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.
\end{aligned} \tag{A.12}$$

Now, under Assumptions 1 – 5,

$$\begin{aligned}
&\frac{n^{-d/2}}{C_n^{MLE}}\int_{\|t\|\le\delta\sqrt{n}} |g(t)|\Pi_n(n^{-1/2}t + \hat{\theta}_n)dt \\
&\le \frac{\hat{M}_1\widetilde{M}_1\int_{\|t\|\le\delta\sqrt{n}} |g(t)|e^{L_n(n^{-1/2}t+L_n)-L_n(\hat{\theta}_n)}dt}{\int_{\|u\|\le\delta\sqrt{n}} e^{L_n(n^{-1/2}u+\hat{\theta}_n)-L_n(\hat{\theta}_n)}du} \\
&\le \frac{\hat{M}_1\widetilde{M}_1\int_{\|t\|\le\delta\sqrt{n}} |g(t)|e^{-\frac{1}{2}t^T\left(J_n(\hat{\theta}_n)-\frac{M_2\delta}{3}I_{d\times d}\right)t}dt}{\int_{\|u\|\le\delta\sqrt{n}} e^{-u^T(J_n(\hat{\theta}_n)+(\delta M_2/3)I_{d\times d})u/2}du}
\end{aligned}$$

$$\leq \frac{\hat{M}_1 \widetilde{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2} \int_{\|t\|\leq \delta\sqrt{n}} |g(t)| e^{-\frac{1}{2}t^T\left(J_n(\hat{\theta}_n) - \frac{M_2\delta}{3} I_{d\times d}\right)t} dt}{(2\pi)^{d/2}\left\{1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right\}}, \tag{A.13}$$

where the last inequality follows from Lemma A.1. A bound on $I_{2,2}$ can be obtained by combining (A.13) with (A.12) applied to $g = 1$. Indeed, we thus obtain:

$$\begin{aligned} I_{2,2}^{MLE} \leq & \frac{\hat{M}_1 \widetilde{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2} \int_{\|u\|\leq \delta\sqrt{n}} |g(u)| e^{-\frac{1}{2}u^T\left(J_n(\hat{\theta}_n) - \frac{M_2\delta}{3} I_{d\times d}\right)u} du}{(2\pi)^{d/2}\left\{1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right\}} \\ & \cdot \left\{ \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \left[\left\|J_n(\hat{\theta}_n)^{-1}\right\|_{op}\right]^{-1}\right] \right. \\ & \left. + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)} \right\}, \end{aligned} \tag{A.14}$$

where we used Lemma A.1. A bound on I_2^{MLE} is obtained by adding together the bounds on $I_{2,1}^{MLE}$ (A.12) and $I_{2,2}^{MLE}$(A.14).

REMARK A.2. Note that, for g, such that $|g| \leq U$, for some $U > 0$, we have that

$$\frac{n^{-d/2}}{C_n^{MLE}} \int_{\|t\|\leq \delta\sqrt{n}} |g(t)| \Pi_n(n^{-1/2}t + \hat{\theta}_n) dt \leq U.$$

Therefore, for $|g| \leq U$, the same argument as above yields a simpler bound:

$$\begin{aligned} I_{2,2}^{MLE} \leq & U \left\{ \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \left[\left\|J_n(\hat{\theta}_n)^{-1}\right\|_{op}\right]^{-1}\right] \right. \\ & \left. + \frac{n^{d/2} e^{-n\kappa} \hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)} \right\}. \end{aligned}$$

A.4. Controlling term I_2^{MAP}. Note that

$$\begin{aligned} I_2^{MAP} \leq & \left| \int_{\|u\| > \bar{\delta}\sqrt{n}} g(u) \left(n^{-d/2} \Pi_n(n^{-1/2}u + \bar{\theta}_n) - \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{(2\pi)^{d/2}} \right) du \right| \\ & + \frac{n^{-d/2}}{C_n^{MAP}} \int_{\|t\|\leq \bar{\delta}\sqrt{n}} |g(t)| \Pi_n(n^{-1/2}t + \bar{\theta}_n) dt \\ & \cdot \left| \int_{\|u\| > \bar{\delta}\sqrt{n}} \left(n^{-d/2} \Pi_n(n^{-1/2}u + \bar{\theta}_n) - \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n) u/2}}{(2\pi)^{d/2}} \right) du \right| \\ =: & I_{2,1}^{MAP} + I_{2,2}^{MAP}. \end{aligned}$$

Note that, using Assumptions 1, 2 and 8, a calculation similar to (A.11), yields

$$\begin{aligned}
&\int_{\|u\|>\bar{\delta}\sqrt{n}} |g(u)| n^{-d/2}\Pi_n(n^{-1/2}u+\bar{\theta}_n)du \\
&= \int_{\|v-\bar{\theta}_n\|>\bar{\delta}} \left|g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \Pi_n(v)dv \\
&\leq \frac{\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left|g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)e^{L_n(v)}dv}{\int_{\mathbb{R}^d}\pi(t)e^{L_n(t)}dt} \\
&\leq \frac{\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left|g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)e^{L_n(v)-L_n(\hat{\theta}_n)}dv}{\int_{\|t-\hat{\theta}_n\|\leq\delta}\pi(t)e^{L_n(t)-L_n(\hat{\theta}_n)}dt} \\
&\leq \frac{n^{d/2}e^{-n\bar{\kappa}}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left|g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)dv}{\int_{\|t\|\leq\delta\sqrt{n}}\pi(n^{-1/2}t+\hat{\theta}_n)e^{-t^T(J_n(\hat{\theta}_n)+(\delta M_2/3)I_{d\times d})t/2}dt} \\
&\leq \frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left|g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)},
\end{aligned}$$

if $n > \frac{\mathrm{Tr}[R_1(n,\delta)]}{\delta^2}$, where the last inequality follows from Lemma A.1. Therefore, under assumptions 1, 2 and 8 and if $n > \frac{\mathrm{Tr}[R_1(n,\delta)]}{\delta^2}$,

$$\begin{aligned}
I_{2,1}^{MAP} &\leq \left|\int_{\|u\|>\bar{\delta}\sqrt{n}} g(u)\frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}}du\right| \\
&+ \frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \left|g\left(\sqrt{n}(v-\bar{\theta}_n)\right)\right| \pi(v)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}. \qquad \text{(A.15)}
\end{aligned}$$

Now, under Assumptions 1, 2 and 6 – 8,

$$\begin{aligned}
&\frac{n^{-d/2}}{C_n^{MAP}}\int_{\|t\|\leq\bar{\delta}\sqrt{n}} |g(t)|\Pi_n(n^{-1/2}t+\bar{\theta}_n)dt \\
&\leq \frac{\int_{\|t\|\leq\bar{\delta}\sqrt{n}}|g(t)|e^{\overline{L}_n(n^{-1/2}t+\bar{\theta}_n)-\overline{L}_n(\bar{\theta}_n)}dt}{\int_{\|u\|\leq\bar{\delta}\sqrt{n}}e^{\overline{L}_n(n^{-1/2}u+\bar{\theta}_n)-\overline{L}_n(\bar{\theta}_n)}du} \\
&\leq \frac{\int_{\|t\|\leq\bar{\delta}\sqrt{n}}|g(t)|e^{-\frac{1}{2}t^T\left(\bar{J}_n(\bar{\theta}_n)-\frac{\overline{M}_2\bar{\delta}}{3}I_{d\times d}\right)t}dt}{\int_{\|u\|\leq\bar{\delta}\sqrt{n}}e^{-u^T(\bar{J}_n(\bar{\theta}_n)+(\bar{\delta}\overline{M}_2/3)I_{d\times d})u/2}du} \\
&\leq \frac{\left|\det\left(\overline{R}_1(n,\bar{\delta})\right)\right|^{-1/2}\int_{\|u\|\leq\bar{\delta}\sqrt{n}}|g(u)|e^{-\frac{1}{2}u^T\left(\bar{J}_n(\bar{\theta}_n)-\frac{\overline{M}_2\bar{\delta}}{3}I_{d\times d}\right)u}du}{(2\pi)^{d/2}\left\{1-\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left[\overline{R}_1(n,\bar{\delta})\right]}\right)^2\left[\left\|\overline{R}_1(n,\bar{\delta})\right\|_{op}\right]^{-1}\right]\right\}}, \qquad \text{(A.16)}
\end{aligned}$$

where the last inequality follows from Lemma A.1. A bound on $I_{2,2}^{MAP}$ can be obtained by combining (A.16) with (A.15) applied to $g=1$. Indeed, we obtain:

$$I_{2,2}^{MAP} \leq \frac{\left|\det\left(\overline{R}_1(n,\bar{\delta})\right)\right|^{-1/2}\int_{\|u\|\leq\bar{\delta}\sqrt{n}}|g(u)|e^{-\frac{1}{2}u^T\left(\bar{J}_n(\bar{\theta}_n)-\frac{\overline{M}_2\bar{\delta}}{3}I_{d\times d}\right)u}du}{(2\pi)^{d/2}\left\{1-\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\operatorname{Tr}\left[\overline{R}_1(n,\bar{\delta})\right]}\right)^2\left[\left\|\overline{R}_1(n,\bar{\delta})\right\|_{op}\right]^{-1}\right]\right\}}$$

$$\cdot\left\{\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\left[\left\|\bar{J}_n(\bar{\theta}_n)^{-1}\right\|_{op}\right]^{-1}\right]\right.$$

$$\left.+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)}\right\}, \tag{A.17}$$

where we applied Lemma A.1.

REMARK A.3. As in Remark A.2, our bound gets simpler if $|g|\leq U$, for some $U>0$. In that case, instead of (A.17), we can write:

$$I_{2,2}^{MAP} \leq U\left\{\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\left[\left\|\bar{J}_n(\bar{\theta}_n)^{-1}\right\|_{op}\right]^{-1}\right]\right.$$

$$\left.+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)}\right\}.$$

APPENDIX B: PROOFS OF THEOREMS 2.1-2.3

Throughout this section we adopt the notation of Section A. In all the proofs below, we wish to control the quantity D_g^{MAP} of (A.2) for all functions g which satisfy certain prescribed criteria. In the proof of Theorem 2.1, we look at funcitons g which are indicators of measurable sets, in the proof of Theorem 2.2 we look at 1-Lipschitz funcitons g and in the proof of Theorem 2.3 at those which are of the form $g(x)=\langle v,x\rangle^2$ for some $v\in\mathbb{R}^d$ with $\|v\|=1$. In order to prove Theorems 2.1 – 2.3, we will bound terms I_2^{MAP} and I_1^{MAP} of (A.9) separately.

B.1. Proof of Theorem 2.1.

B.1.1. *Controlling term I_2^{MAP}.* We wish to obtain a uniform bound on I_2^{MAP} for all functions g which are indicators of measurable sets. Every indicator function is upper-bounded by one, so we can use Remark A.3 to obtain:

$$I_{2,2}^{MAP} \leq \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\left[\left\|\bar{J}_n(\bar{\theta}_n)^{-1}\right\|_{op}\right]^{-1}\right]$$

$$+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)}. \tag{B.1}$$

Similarly, since $|g| \le 1$, we can use (A.15) and Lemma A.1 to obtain:

$$I_{2,1}^{MAP} \le \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\left[\left\|\bar{J}_n(\bar{\theta}_n)^{-1}\right\|_{op}\right]^{-1}\right]$$

$$+\frac{n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)}. \tag{B.2}$$

From (B.1) and (B.2), it follows that

$$I_2^{MAP} \le 2\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\left[\left\|\bar{J}_n(\bar{\theta}_n)^{-1}\right\|_{op}\right]^{-1}\right]$$

$$+\frac{2n^{d/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)}. \tag{B.3}$$

B.1.2. *Controlling term I_1^{MAP} using the log-Sobolev inequality.* For any probability measure μ, let $[\mu]_{B_0(\bar{\delta}\sqrt{n})}$ denote its restriction (truncation) to the ball of radius $\bar{\delta}\sqrt{n}$ around 0. Let $\mathrm{KL}\left(\cdot\|\cdot\right)$ denote the Kullback-Leibler divergence (i.e. the relative entropy, see e.g. [3, Section 1.6.1]).

Note that, by Assumption 6, for t such that $\|t\| < \sqrt{n}\bar{\delta}$, we have

$$-n^{-1}\overline{L}_n''(\bar{\theta}_n + n^{-1/2}t) \succeq \bar{J}_n(\bar{\theta}_n) - \bar{\delta}\overline{M}_2 I_{d\times d} \succeq \left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\overline{M}_2\right) I_{d\times d}.$$

This means that, inside the convex set $\{t \in \mathbb{R}^d : \|t\| < \sqrt{n}\bar{\delta}\}$, the density of $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n)$ is $\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\overline{M}_2\right)$-strongly log-concave (see e.g. [34]). Let F_n^{MAP} be given by (A.8). Using the **Bakry-Emery criterion**, we have that $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}$ satisfies the **log-Sobolev inequality** (see, e.g. [35, Appendix A] for a summary of all those results). By combining the log-Sobolev inequality with Pinsker's inequality (see, e.g. [28, Theorem 2.16]) we obtain that, for all functions g, which are indicators of measurable sets,

$$\begin{aligned}
I_1^{MAP} &\le \mathrm{TV}\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}, \left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})}\right)\\
&\le \sqrt{\frac{1}{2}\mathrm{KL}\left(\left[\mathcal{N}(0,\bar{J}_n(\bar{\theta}_n)^{-1})\right]_{B_0(\bar{\delta}\sqrt{n})} \middle\| \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n\right)\right)\right]_{B_0(\bar{\delta}\sqrt{n})}\right)}\\
&\le \frac{1}{2\sqrt{\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\,\overline{M}_2}}\\
&\quad\cdot\sqrt{\int_{\|u\|\le\bar{\delta}\sqrt{n}} \frac{\sqrt{|\det \bar{J}_n(\bar{\theta}_n)|}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{F_n^{MAP}(2\pi)^{d/2}}\left\|\bar{J}_n(\bar{\theta}_n)u + \frac{\overline{L}_n'(n^{-1/2}u+\bar{\theta}_n)}{\sqrt{n}}\right\|^2 du}\\
&\overset{\text{Taylor}}{\le} \frac{\overline{M}_2}{4\sqrt{n\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\,\overline{M}_2\right)}}\sqrt{\int_{\|u\|\le\bar{\delta}\sqrt{n}} \frac{|\det \bar{J}_n(\bar{\theta}_n)|^{1/2}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{F_n^{MAP}(2\pi)^{d/2}}\|u\|^4\, du}
\end{aligned}$$

(B.4)

$$\leq \frac{\sqrt{3}\,\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]\overline{M}_2}{4\sqrt{n\left(\bar{\lambda}_{\min}(\bar{\theta}_n)-\bar{\delta}\overline{M}_2\right)\left(1-\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}\right)^2\bar{\lambda}_{\min}(\bar{\theta}_n)\right]\right)}},$$

as long as $n > \frac{\mathrm{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}{\bar{\delta}^2}$, where the last inequality follows from Lemma A.1.

B.1.3. *Conclusion.* The result now follows from adding together bounds (B.3) and (B.4).

B.2. Proof of Theorem 2.2.

B.2.1. *Controlling term I_2^{MAP}.* Now we wish to control I_2^{MAP} uniformly over all functions g which are 1-Lipschitz. Let us fix a function g that is 1-Lipschitz and WLOG set $g(0)=0$. In that case $|g(u)| \leq \|u\|$ and, using the notation of Section A and equation (A.15),

$$I_{2,1}^{MAP} \leq \int_{\|u\|>\bar{\delta}\sqrt{n}} \|u\| \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}}du$$
$$+\frac{n^{d/2+1/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|v-\bar{\theta}_n\|\pi(v)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.$$

Now, for $U \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$, and assuming that $n > \frac{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{\bar{\delta}^2}$,

$$\int_{\|u\|>\bar{\delta}\sqrt{n}} \|u\| \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|}e^{-u^T\bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}}du$$

(B.5)
$$= \int_0^\infty \mathbb{P}\left[\|U\|\mathbb{1}_{[\|U\|>\bar{\delta}\sqrt{n}]} > t\right]dt$$
$$\leq \int_0^\infty \mathbb{P}\left[\|U\| > \max(t, \bar{\delta}\sqrt{n})\right]dt$$
$$\leq \int_{\bar{\delta}\sqrt{n}}^\infty \exp\left[-\frac{1}{2}\left(t-\sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\bar{\lambda}_{\min}(\bar{\theta}_n)\right]dt$$
$$+\bar{\delta}\sqrt{n}\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\bar{\lambda}_{\min}(\bar{\theta}_n)\right]$$

(B.6)
$$\leq \left(\bar{\delta}\sqrt{n}+\sqrt{\frac{2\pi}{\bar{\lambda}_{\min}(\bar{\theta}_n)}}\right)\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\bar{\lambda}_{\min}(\bar{\theta}_n)\right].$$

This means that

$$I_{2,1}^{MAP} \leq \left(\bar{\delta}\sqrt{n}+\sqrt{\frac{2\pi}{\bar{\lambda}_{\min}(\bar{\theta}_n)}}\right)\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\mathrm{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2\bar{\lambda}_{\min}(\bar{\theta}_n)\right]$$

(B.7)
$$+\frac{n^{d/2+1/2}e^{-n\bar{\kappa}}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|}\|v-\bar{\theta}_n\|\pi(v)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.$$

Now, using (A.17), we have

$$I_{2,2}^{MAP} \le \frac{\left|\det\left(\overline{R}_1(n,\bar\delta)\right)\right|^{-1/2}\left|\det\left(\overline{R}_2(n,\bar\delta)\right)\right|^{1/2}\sqrt{\mathrm{Tr}\left[\overline{R}_2(n,\bar\delta)\right]}}{1-\exp\left[-\frac{1}{2}\left(\bar\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[\overline{R}_1(n,\bar\delta)\right]}\right)^2\left[\|\overline{R}_1(n,\bar\delta)\|_{op}\right]^{-1}\right]}$$

$$\cdot\left\{\exp\left[-\frac{1}{2}\left(\bar\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[\bar J_n(\bar\theta_n)^{-1}\right]}\right)^2\overline{\lambda}_{\min}(\bar\theta_n)\right]\right.$$

$$\left. + \frac{n^{d/2}e^{-n\bar\kappa}\hat M_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}\right\}. \tag{B.8}$$

Adding together bounds (B.7) and (B.8) now yields a bound on I_2^{MAP}.

B.2.2. *Controlling term I_1^{MAP} using the log-Sobolev inequality and the transportation-information inequality.* As in Subsection B.1.2, we shall use the log-Sobolev inequality for the measure $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol\theta}_n-\bar\theta_n\right)\right)\right]_{B_0(\bar\delta\sqrt{n})}$. A consequence of the log-Sobolev inequality is that we can apply the transportation-information inequality for $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol\theta}_n-\bar\theta_n\right)\right)\right]_{B_0(\bar\delta\sqrt{n})}$ (see [31, Theorem 1] or [18]), which lets us upper bound the 1- and 2-Wasserstein distances by a constant times the Fisher divergence. Let $W_2(\cdot,\cdot)$ denote the 2-Wasserstein distance and $W_1(\cdot,\cdot)$ denote the 1-Wasserstein distance. We have that

$$\begin{aligned}
I_1^{MAP} \le & W_1\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol\theta}_n-\bar\theta_n\right)\right)\right]_{B_0(\bar\delta\sqrt{n})},\left[\mathcal{N}(0,\bar J_n(\bar\theta_n)^{-1})\right]_{B_0(\bar\delta\sqrt{n})}\right)\\
\le & W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol\theta}_n-\bar\theta_n\right)\right)\right]_{B_0(\bar\delta\sqrt{n})},\left[\mathcal{N}(0,\bar J_n(\bar\theta_n)^{-1})\right]_{B_0(\bar\delta\sqrt{n})}\right)\\
\le & \frac{1}{\overline{\lambda}_{\min}(\bar\theta_n)-\bar\delta\,\overline{M}_2}\\
&\cdot\sqrt{\int_{\|u\|\le\bar\delta\sqrt{n}}\frac{\sqrt{|\det\bar J_n(\bar\theta_n)|}e^{-u^T\bar J_n(\bar\theta_n)u/2}}{F_n^{MAP}(2\pi)^{d/2}}\left\|\bar J_n(\bar\theta_n)u+\frac{\overline{L}_n'(n^{-1/2}u+\bar\theta_n)}{\sqrt{n}}\right\|^2du}\\
\overset{\text{Taylor}}{\le} & \frac{\overline{M}_2}{2\sqrt{n}\left(\overline{\lambda}_{\min}(\bar\theta_n)-\bar\delta\,\overline{M}_2\right)}\sqrt{\int_{\|u\|\le\bar\delta\sqrt{n}}\frac{|\det\bar J_n(\bar\theta_n)|^{1/2}e^{-u^T\bar J_n(\bar\theta_n)u/2}}{F_n^{MAP}(2\pi)^{d/2}}\|u\|^4du}\\
\le & \frac{\sqrt{3}\,\mathrm{Tr}\left[\bar J_n(\bar\theta_n)^{-1}\right]\overline{M}_2}{2\left(\overline{\lambda}_{\min}(\bar\theta_n)-\bar\delta\overline{M}_2\right)\sqrt{n\left(1-\exp\left[-\frac{1}{2}\left(\bar\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[\bar J_n(\bar\theta_n)^{-1}\right]}\right)^2\overline{\lambda}_{\min}(\bar\theta_n)\right]\right)}},
\end{aligned}\tag{B.9}$$

as long as $n>\frac{\mathrm{Tr}\left[\bar J_n(\bar\theta_n)^{-1}\right]}{\bar\delta^2}$, where the last iequality follows from Lemma A.1.

B.2.3. *Conclusion.* The result now follows from adding together bounds (B.7), (B.8) and (B.9).

B.3. Proof of Theorem 2.3.

B.3.1. *Controlling term I_2^{MAP}.* Now we wish to control I_2^{MAP} uniformly over all functions g which are of the form $g(u) = \langle v, u\rangle^2$ for some $v \in \mathbb{R}^d$ with $\|v\| = 1$. Such functions satisfy the following property: $|g(u)| \leq \|u\|^2$. Using the notation of Section A and equation (A.15), we therefore have that:

$$I_{2,1}^{MAP} \leq \int_{\|u\|>\bar{\delta}\sqrt{n}} \|u\|^2 \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}} du + \frac{n^{d/2+1} e^{-n\bar{\kappa}} \hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2} \int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \|v-\bar{\theta}_n\|^2 \pi(v) du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.$$

Now, for $U \sim \mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1})$, and assuming that $n > \frac{\operatorname{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}{\bar{\delta}^2}$,

$$\begin{aligned}
&\int_{\|u\|>\bar{\delta}\sqrt{n}} \|u\|^2 \frac{\sqrt{\left|\det \bar{J}_n(\bar{\theta}_n)\right|} e^{-u^T \bar{J}_n(\bar{\theta}_n)u/2}}{(2\pi)^{d/2}} du \\
&= \int_0^\infty \mathbb{P}\left[\|U\|^2 \mathbb{1}_{[\|U\|>\bar{\delta}\sqrt{n}]} > t\right] dt \qquad \text{(B.10)}\\
&\leq \int_0^\infty \mathbb{P}\left[\|U\| > \max(\sqrt{t}, \bar{\delta}\sqrt{n})\right] dt \\
&\leq \int_{\bar{\delta}^2 n}^\infty \exp\left[-\frac{1}{2}\left(\sqrt{t} - \sqrt{\operatorname{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right] dt \\
&\quad + \bar{\delta}^2 n \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{\operatorname{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right] \\
&\leq \left(\bar{\delta}^2 n + \sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}}\right) \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{\operatorname{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right]. \qquad \text{(B.11)}
\end{aligned}$$

where we have used Lemma A.1. This means that

$$I_{2,1}^{MAP} \leq \left(\bar{\delta}^2 n + \sqrt{\frac{2\pi}{\overline{\lambda}_{\min}(\bar{\theta}_n)}}\right) \exp\left[-\frac{1}{2}\left(\bar{\delta}^2\sqrt{n} - \sqrt{\operatorname{Tr}\left(\bar{J}_n(\bar{\theta}_n)^{-1}\right)}\right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n)\right]$$

$$+ \frac{n^{d/2+1/2} e^{-n\bar{\kappa}} \hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2} \int_{\|v-\hat{\theta}_n\|>\bar{\delta}-\|\hat{\theta}_n-\bar{\theta}_n\|} \|v-\bar{\theta}_n\|^2 \pi(v) du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}. \qquad \text{(B.12)}$$

Now, using (A.17), we have

$$I_{2,2}^{MAP} \leq \frac{\left|\det\left(\overline{R}_1(n,\bar{\delta})\right)\right|^{-1/2} \left|\det\left(\overline{R}_2(n,\bar{\delta})\right)\right|^{1/2} \operatorname{Tr}\left[\overline{R}_2(n,\bar{\delta})\right]}{1-\exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n}-\sqrt{\operatorname{Tr}\left[\overline{R}_1(n,\bar{\delta})\right]}\right)^2\left[\|\overline{R}_1(n,\bar{\delta})\|_{op}\right]^{-1}\right]}$$

$$\cdot \left\{ \exp\left[-\frac{1}{2} \left(\bar{\delta}\sqrt{n} - \sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]} \right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n) \right] \right.$$

(B.13)

$$\left. + \frac{n^{d/2} e^{-n\bar{\kappa}} \hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2} \left(1 - \exp\left[-\frac{1}{2} \left(\delta\sqrt{n} - \sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]} \right)^2 \left[\|R_1(n,\delta)\|_{op} \right]^{-1} \right] \right)} \right\}.$$

A bound on I_2^{MAP} now follows from adding up the bounds (B.12) and (B.13).

B.3.2. *Controlling term I_1^{MAP} using the log-Sobolev inequality and the transportation-information inequality.* Note that calculation (B.9) yields that

$$W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right) \right]_{B_0(\bar{\delta}\sqrt{n})}, \left[\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1}) \right]_{B_0(\bar{\delta}\sqrt{n})} \right)$$

(B.14)

$$\leq \frac{\sqrt{3}\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right] \overline{M}_2}{2\left(\overline{\lambda}_{\min}(\bar{\theta}_n) - \bar{\delta}\overline{M}_2 \right) \sqrt{n\left(1 - \exp\left[-\frac{1}{2}\left(\bar{\delta}\sqrt{n} - \sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]} \right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n) \right] \right)}}.$$

Now, let us fix two random vectors: $X \sim \left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right) \right]_{B_0(\bar{\delta}\sqrt{n})}$ and $Y \sim \left[\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1}) \right]_{B_0(\bar{\delta}\sqrt{n})}$ and a vector v such that $\|v\| = 1$. Let γ denote the set of all couplings between $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right) \right]_{B_0(\bar{\delta}\sqrt{n})}$ and $\left[\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1}) \right]_{B_0(\bar{\delta}\sqrt{n})}$. Let $(\tilde{X}, \tilde{Y}) \in \gamma$ be such that

$$\inf_{(Z_1,Z_2)\in\gamma} \mathbb{E}\left[\|Z_1 - Z_2\|^2\right] = \mathbb{E}\left[\|\tilde{X} - \tilde{Y}\|^2\right].$$

It follows that:

$$\begin{aligned}
\mathbb{E}\left[\langle v, X\rangle^2\right] - \mathbb{E}\left[\langle v, Y\rangle^2\right] =& \mathbb{E}\left[\left\langle v, \tilde{X} - \tilde{Y} \right\rangle \left\langle v, \tilde{X} + \tilde{Y} \right\rangle \right] \\
=& \mathbb{E}\left[\left\langle v, \tilde{X} - \tilde{Y} \right\rangle^2 \right] + 2\mathbb{E}\left[\left\langle v, \tilde{X} - \tilde{Y} \right\rangle \left\langle v, \tilde{Y} \right\rangle \right] \\
\leq& \mathbb{E}\left[\left\| \tilde{X} - \tilde{Y} \right\|^2 \right] + 2\sqrt{\mathbb{E}\left[\left\| \tilde{X} - \tilde{Y} \right\|^2 \right]} \sqrt{\mathbb{E}\left[\|Y\|^2\right]}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
I_1^{MAP} \leq& W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right) \right]_{B_0(\bar{\delta}\sqrt{n})}, \left[\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1}) \right]_{B_0(\bar{\delta}\sqrt{n})} \right)^2 \\
&+ 2W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \bar{\theta}_n \right) \right) \right]_{B_0(\bar{\delta}\sqrt{n})}, \left[\mathcal{N}(0, \bar{J}_n(\bar{\theta}_n)^{-1}) \right]_{B_0(\bar{\delta}\sqrt{n})} \right)
\end{aligned}$$

(B.15)

$$\cdot \frac{\sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]}}{\sqrt{\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\operatorname{Tr}\left[\bar{J}_n(\bar{\theta}_n)^{-1}\right]} \right)^2 \overline{\lambda}_{\min}(\bar{\theta}_n) \right] \right)}},$$

if $n > \frac{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}{\delta^2}$ and the final bound on I_1^{MAP} may be obtained by using (B.14).

B.3.3. *Conclusion.* The result now follows by combining (B.15) with (B.14) and then summing together with (B.12) and (B.13).

APPENDIX C: PROOFS OF THEOREMS 4.1 – 4.3

Throughout this section we adopt the notation of Section A. In all the proofs below, we wish to control the quantity D_g^{MLE} of (A.1) for all functions g which satisfy certain prescribed criteria. In the proof of Theorem 2.1, we look at functions g which are indicators of measurable sets, in the proof of Theorem 2.2 we look at 1-Lipschitz functions g and in the proof of Theorem 2.3 at those which are of the form $g(x) = \langle v, x\rangle^2$ for some $v \in \mathbb{R}^d$ with $\|v\| = 1$. In order to prove Theorems 4.1 – 4.3, we will bound terms I_2^{MLE} and I_1^{MAP} of (A.7) separately.

C.1. Proof of Theorem 4.1.

C.1.1. *Controlling term I_2^{MLE}.* We wish to obtain a uniform bound on I_2^{MLE} for all functions g which are indicators of measurable sets. Every indicator function is upper-bounded by one, so we can use Remark A.2 to obtain:

$$I_{2,2}^{MLE} \leq \exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right] + \frac{n^{d/2}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)}.$$

Similarly, since $|g| \leq 1$, we can use (A.12) and Lemma A.1 to obtain

$$I_{2,1}^{MLE} \leq \exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right] + \frac{n^{d/2}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)},$$

which implies that

$$I_2^{MLE} \leq 2\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right] + 2\frac{n^{d/2}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\left\|R_1(n,\delta)\right\|_{op}\right]^{-1}\right]\right)}. \tag{C.1}$$

C.1.2. *Controlling term I_1^{MLE} using the log-Sobolev inequality.* We shall proceed as we did in Subsection B.1.2. For any probability measure μ, let $[\mu]_{B_0(\delta\sqrt{n})}$ denote its restriction (truncation) to the ball of radius $\delta\sqrt{n}$ around 0. Let $\mathrm{KL}\left(\cdot\|\cdot\right)$ denote the Kullback-Leibler divergence (i.e. the relative entropy, see e.g. [3, Section 1.6.1]).

Note that, by Assumption 1, for t such that $\|t\| < \sqrt{n}\delta$, we have

$$-n^{-1}L_n''(\hat{\theta}_n + n^{-1/2}t) \succeq J_n(\hat{\theta}_n) - \delta M_2 I_{d\times d} \succeq \left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right) I_{d\times d}.$$

This means that, inside the convex set $\{t \in \mathbb{R}^d : \|t\| < \sqrt{n}\delta\}$, the density of $\sqrt{n}(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n)$ is $\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)$-strongly log-concave (see e.g. [34]). Also, note that

$$\frac{\sup_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t + \hat{\theta}_n)}{\inf_{\|t\|\le\delta\sqrt{n}} \pi(n^{-1/2}t + \hat{\theta}_n)} \le \widetilde{M}_1 \hat{M}_1.$$

Let F_n^{MLE} be given by (A.6). Using the **Bakry-Emery criterion and the Holley-Stroock perturbation principle**, we therefore have that $\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})}$ satisfies the **log-Sobolev ienquality** with constant $\widetilde{M}_1 \hat{M}_1 \left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)$ (see, e.g. [35, Appendix A] for a summary of all those results). By combining the log-Sobolev inequality with Pinsker's inequality (see, e.g. [28, Theorem 2.16]) we obtain that, for all functions g, which are indicators of measurable sets,

$$\begin{aligned}
I_1^{MLE} \le & \mathrm{TV}\left(\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})}, \left[\mathcal{N}(0, J_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right) \\
\le & \sqrt{\frac{1}{2}\mathrm{KL}\left(\left[\mathcal{N}(0, J_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})} \middle\| \left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})}\right)} \\
\le & \frac{\sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{\lambda_{\min}(\hat{\theta}_n) - \delta M_2}} \\
& \cdot \sqrt{\int_{\|u\|\le\delta\sqrt{n}} \frac{\sqrt{|\det J_n(\hat{\theta}_n)|} e^{-u^T J_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}} \left\| J_n(\hat{\theta}_n)u + \frac{L_n'(n^{-1/2}u + \hat{\theta}_n)}{\sqrt{n}} \right\|^2 du} \\
& + \frac{\sqrt{\widetilde{M}_1 \hat{M}_1}}{2\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}} \\
& \quad \cdot \sqrt{\int_{\|u\|\le\delta\sqrt{n}} \frac{\sqrt{|\det J_n(\hat{\theta}_n)|} e^{-u^T J_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}} \left\| \frac{\pi'(n^{-1/2}u + \hat{\theta}_n)}{\pi(n^{-1/2}u + \hat{\theta}_n)} \right\|^2 du} \\
\overset{\text{Taylor}}{\le} & \frac{\sqrt{\widetilde{M}_1 \hat{M}_1} M_2}{4\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}} \sqrt{\int_{\|u\|\le\delta\sqrt{n}} \frac{\sqrt{|\det J_n(\hat{\theta}_n)|} e^{-u^T J_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}} \|u\|^4 du}
\end{aligned}$$

$$+\frac{M_1\sqrt{\widetilde{M}_1\hat{M}_1}}{2\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)}}$$

$$\le\frac{\sqrt{3}\,\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]\sqrt{\widetilde{M}_1\hat{M}_1}\,M_2}{4\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2\lambda_{\min}(\hat{\theta}_n)\right]\right)}}$$

(C.2)

$$+\frac{M_1\sqrt{\widetilde{M}_1\hat{M}_1}}{2\sqrt{n\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)}},$$

as long as $n>\frac{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}{\delta^2}$, where we have used Lemma A.1.

C.1.3. *Conclusion.* The result now follows by adding together the bounds (C.1) and (C.2).

C.2. Proof of Theorem 4.2.

C.2.1. *Controlling term I_2^{MLE}.* Now we wish to control I_2^{MLE} uniformly over all functions g which are 1-Lipschitz and WLOG set $g(0)=0$. It follows that $|g(u)|\le\|u\|$ and, using the notation of Section A and equation (A.12),

$$I_{2,1}^{MLE}\le\int_{\|u\|>\delta\sqrt{n}}\|u\|\frac{\sqrt{\left|\det J_n(\hat{\theta}_n)\right|}e^{-u^TJ_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}}du$$
$$+\frac{n^{d/2+1/2}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|u\|>\delta}\|u\|\pi(u+\hat{\theta}_n)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.$$

A calculation similar to (B.6) reveals that

$$\int_{\|u\|>\delta\sqrt{n}}\|u\|\frac{\sqrt{\left|\det J_n(\hat{\theta}_n)\right|}e^{-u^TJ_n(\hat{\theta}_n)u/2}}{(2\pi)^{d/2}}du$$
$$\le\left(\delta\sqrt{n}+\sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right)\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left(J_n(\hat{\theta}_n)^{-1}\right)}\right)^2\lambda_{\min}(\hat{\theta}_n)\right]$$

and so

$$I_{2,1}^{MLE}\le\left(\delta\sqrt{n}+\sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right)\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\mathrm{Tr}\left(J_n(\hat{\theta}_n)^{-1}\right)}\right)^2\lambda_{\min}(\hat{\theta}_n)\right]$$

$$+\frac{n^{d/2+1/2}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|u\|>\delta}\|u\|\pi(u+\hat{\theta}_n)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\text{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}. \tag{C.3}$$

Now, using (A.14), we obtain

$$I_{2,2}^{MLE}\leq\frac{\hat{M}_1\widetilde{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\left|\det\left(R_2(n,\delta)\right)\right|^{1/2}\sqrt{\text{Tr}\left[R_2(n,\delta)\right]}}{1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\text{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]}$$

$$\cdot\left\{\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\text{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2\lambda_{\min}(\hat{\theta}_n)\right]\right.$$

$$\left.+\frac{n^{d/2}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\text{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}\right\}. \tag{C.4}$$

Adding together bounds (C.3) and (C.4) yields a bound on I_2^{MLE}.

C.2.2. *Controlling term I_1^{MLE} using the log-Sobolev inequality and the transportation-information inequality.* As in Subsection C.1.2, we shall use the log-Sobolev inequality for the measure $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}$. A consequence of the log-Sobolev inequality is that we can apply the transportation-information inequality for $\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})}$ (see [31, Theorem 1] or [18]), which lets us upper bound the 1- and 2-Wasserstein distances by a constant times the Fisher divergence. Let $W_2(\cdot,\cdot)$ denote the 2-Wasserstein distance and $W_1(\cdot,\cdot)$ denote the 1-Wasserstein distance. We have that

$$\begin{aligned}
I_1^{MLE}\leq& W_1\left(\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,J_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right)\\
\leq& W_2\left(\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,J_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right)\\
\leq&\frac{\widetilde{M}_1\hat{M}_1}{\lambda_{\min}(\hat{\theta}_n)-\delta M_2}\\
&\cdot\sqrt{\int_{\|u\|\leq\delta\sqrt{n}}\frac{\sqrt{|\det J_n(\hat{\theta}_n)|}e^{-u^TJ_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}}\left\|J_n(\hat{\theta}_n)u+\frac{L_n'(n^{-1/2}u+\hat{\theta}_n)}{\sqrt{n}}\right\|^2du}\\
&+\frac{\widetilde{M}_1\hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)}\\
&\quad\cdot\sqrt{\int_{\|u\|\leq\delta\sqrt{n}}\frac{\sqrt{|\det J_n(\hat{\theta}_n)|}e^{-u^TJ_n(\hat{\theta}_n)u/2}}{F_n^{MLE}(2\pi)^{d/2}}\left\|\frac{\pi'(n^{-1/2}u+\hat{\theta}_n)}{\pi(n^{-1/2}u+\hat{\theta}_n)}\right\|^2du}
\end{aligned}$$

$$\overset{\text{Taylor}}{\leq} \frac{\widetilde{M}_1 \hat{M}_1 M_2}{2\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)} \sqrt{\int_{\|u\| \leq \delta\sqrt{n}} \frac{\sqrt{|\det J_n(\hat{\theta}_n)|} e^{-u^T J_n(\hat{\theta}_n) u/2}}{F_n^{MLE}(2\pi)^{d/2}} \|u\|^4 \, du}$$

$$+ \frac{M_1 \widetilde{M}_1 \hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}$$

$$\leq \frac{\sqrt{3}\,\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right] \widetilde{M}_1 \hat{M}_1 \, M_2}{2\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right) \sqrt{n\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right]\right)}}$$

(C.5)

$$+ \frac{M_1 \widetilde{M}_1 \hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n) - \delta M_2\right)}.$$

C.2.3. *Conclusion.* The result now follows from adding together the bounds (C.3), (C.4) and (C.5).

C.3. Proof of Theorem 4.3.

C.3.1. *Controlling term I_2^{MLE}.* Now we want to control I_2^{MLE} uniformly over all functions g which are of the form $g(u) = \langle v, u \rangle$, for some $v \in \mathbb{R}^d$ with $\|v\| = 1$. For such functions we have that $|g(u)| \leq \|u\|^2$. Using the notation of Section A and equation (A.12), we have that

$$I_{2,1}^{MLE} \leq \int_{\|u\| > \delta\sqrt{n}} \|u\|^2 \frac{\sqrt{\left|\det J_n(\hat{\theta}_n)\right|} e^{-u^T J_n(\hat{\theta}_n) u/2}}{(2\pi)^{d/2}} du$$

$$+ \frac{n^{d/2+1} e^{-n\kappa} \hat{M}_1 \left|\det\left(R_1(n,\delta)\right)\right|^{-1/2} \int_{\|u\| > \delta} \|u\|^2 \pi(u + \hat{\theta}_n) du}{(2\pi)^{d/2}\left(1 - \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left[R_1(n,\delta)\right]}\right)^2 \left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}.$$

A calculation similar to (B.11) reveals that

$$\int_{\|u\| > \delta\sqrt{n}} \|u\|^2 \frac{\sqrt{\left|\det J_n(\hat{\theta}_n)\right|} e^{-u^T J_n(\hat{\theta}_n) u/2}}{(2\pi)^{d/2}} du$$

$$\leq \left(\delta^2 n + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right) \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left(J_n(\hat{\theta}_n)^{-1}\right)}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right]$$

and so

$$I_{2,1}^{MLE} \leq \left(\delta^2 n + \sqrt{\frac{2\pi}{\lambda_{\min}(\hat{\theta}_n)}}\right) \exp\left[-\frac{1}{2}\left(\delta\sqrt{n} - \sqrt{\mathrm{Tr}\left(J_n(\hat{\theta}_n)^{-1}\right)}\right)^2 \lambda_{\min}(\hat{\theta}_n)\right]$$

$$+\frac{n^{d/2+1}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\int_{\|u\|>\delta}\|u\|^2\pi(u+\hat{\theta}_n)du}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}. \tag{C.6}$$

Now, using (A.14), we obtain

$$\begin{aligned}
I_{2,2}^{MLE} \leq& \frac{\hat{M}_1\widetilde{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}\left|\det\left(R_2(n,\delta)\right)\right|^{1/2}\operatorname{Tr}\left[R_2(n,\delta)\right]}{1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]} \\
&\cdot\left\{\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2\lambda_{\min}(\hat{\theta}_n)\right]\right. \\
&\left.+\frac{n^{d/2}e^{-n\kappa}\hat{M}_1\left|\det\left(R_1(n,\delta)\right)\right|^{-1/2}}{(2\pi)^{d/2}\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[R_1(n,\delta)\right]}\right)^2\left[\|R_1(n,\delta)\|_{op}\right]^{-1}\right]\right)}\right\}.
\end{aligned} \tag{C.7}$$

Adding together bounds (C.6) and (C.7) yields a bound on I_2^{MLE}.

C.3.2. *Controlling term I_1^{MLE} using the log-Sobolev inequality and the transportation-entropy inequality.* Note that calculation (C.5) yields that

$$\begin{aligned}
&W_2\left(\left[\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,J_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right) \\
\leq& \frac{\sqrt{3}\operatorname{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]\widetilde{M}_1\hat{M}_1M_2}{2\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)\sqrt{n\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2\lambda_{\min}(\hat{\theta}_n)\right]\right)}} \\
&+\frac{M_1\widetilde{M}_1\hat{M}_1}{\sqrt{n}\left(\lambda_{\min}(\hat{\theta}_n)-\delta M_2\right)}.
\end{aligned} \tag{C.8}$$

An argument similar to the one that led to (B.15) yields:

$$\begin{aligned}
I_1^{MLE} \leq& W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,J_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right)^2 \\
&+2W_2\left(\left[\mathcal{L}\left(\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n-\hat{\theta}_n\right)\right)\right]_{B_0(\delta\sqrt{n})},\left[\mathcal{N}(0,J_n(\hat{\theta}_n)^{-1})\right]_{B_0(\delta\sqrt{n})}\right) \\
&\cdot\frac{\sqrt{\operatorname{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}}{\sqrt{\left(1-\exp\left[-\frac{1}{2}\left(\delta\sqrt{n}-\sqrt{\operatorname{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}\right)^2\lambda_{\min}(\hat{\theta}_n)\right]\right)}},
\end{aligned} \tag{C.9}$$

and the final bound on I_1^{MLE} follows from (C.8).

C.3.3. *Conclusion.* The result now follows from combining (C.8) and (C.9) and adding together with (C.6) and (C.7).

APPENDIX D: PROOF OF THEOREM 4.4

In this section, we concentrate on the univariate context (i.e. on $d = 1$). We shall apply Stein's method, in the framework described in [12, Section 2.1]. Before we do that, however, let us recall that we want to upper-bound the quantity D_g^{MLE} given by (A.1) for all functions g for which all the two expectations in (A.1) exist. Recall the definition of C_n^{MLE} from (A.5) and let:

$$h(t) = h_g^{MLE}(t) = g(t) - \frac{n^{-1/2}}{C_n} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} g(u)\Pi_n(n^{-1/2}u + \hat{\theta}_n)du. \tag{D.1}$$

We can repeat the calculation leading to(A.7), without dividing the first term after the first inequality by F_n^{MLE}. We then obtain:

$$D_g^{MLE} \leq \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} h(u) \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} \right| + \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} h(u) \left[\frac{e^{-u^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} - n^{-1/2}\Pi_n(n^{-1/2}u + \hat{\theta}_n) \right] \right|$$
$$=: \tilde{I}_1 + \tilde{I}_2.$$

We will bound $\tilde{I}_1$ and $\tilde{I}_2$ separately.

D.1. Controlling term $\tilde{I}_2$. Note that $\tilde{I}_2$ is the same as I_2^{MLE} defined by (A.7), for $d = 1$. We will use the calculations leading to (A.12) and (A.13). Instead of using Lemma A.1, we will, however, apply the standard one-dimensional Gaussian concentration inequality, which says that, for $Z_n \sim \mathcal{N}\left(0, \sigma_n^2\right)$,

$$\mathbb{P}\left[|Z_n| > \delta\sqrt{n}\right] \leq 2e^{-\delta^2 n/(2\sigma_n^2)}.$$

We obtain

$$\begin{aligned}
\tilde{I}_2 \leq& \left| \int_{|u|>\delta\sqrt{n}} g(u) \frac{e^{-u^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} du \right| \\
&+ \frac{n^{1/2}e^{-n\kappa}\hat{M}_1 \left(\frac{1}{\sigma_n^2} + \frac{\delta M_2}{3}\right)^{1/2} \int_{|u|>\delta} |g(u\sqrt{n})| \pi(u + \hat{\theta}_n) du}{\sqrt{2\pi}\left\{1 - 2\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{M_2\delta}{3}\right)\delta^2 n\right]\right\}} \\
&+ \frac{\hat{M}_1\widetilde{M}_1 \left(\frac{1}{\sigma_n^2} + \frac{\delta M_2}{3}\right)^{1/2} \int_{|t|\leq\delta\sqrt{n}} |g(t)| e^{-\frac{1}{2}\left(\frac{1}{\sigma_n^2} - \frac{M_2\delta}{3}\right)t^2} dt}{\sqrt{2\pi}\left\{1 - 2\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{M_2\delta}{3}\right)\delta^2 n\right]\right\}} \\
&\cdot \left\{ 2e^{-\delta^2 n/(2\sigma_n^2)} + \frac{n^{1/2}e^{-n\kappa}\hat{M}_1 \left(\frac{1}{\sigma_n^2} + \frac{\delta M_2}{3}\right)^{1/2}}{\sqrt{2\pi}\left\{1 - 2\exp\left[-\frac{1}{2}\left(\frac{1}{\sigma_n^2} + \frac{M_2\delta}{3}\right)\delta^2 n\right]\right\}} \right\}.
\end{aligned} \tag{D.2}$$

D.2. Controlling term $\tilde{I}_1$ using Stein's method. As mentioned above, in this section we will use Stein's method in the framework on [12, Section 2.1]. Note that, by integration by parts, for all continuous functions $f : [-\delta\sqrt{n}, \delta\sqrt{n}] \to \mathbb{R}$ which are differentiable on $(-\delta\sqrt{n}, \delta\sqrt{n})$ and satisfy $f(-\delta\sqrt{n}) = f(\delta\sqrt{n})$, we have

$$\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} f'(t) \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt = \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} t f(t) \frac{e^{-t^2/(2\sigma_n^2)}}{\sigma_n^2\sqrt{2\pi\sigma_n^2}} dt \tag{D.3}$$

Now, for our function h, given by (D.1), let

(D.4) $$f(t) := \begin{cases} \frac{1}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)} \int_{-\delta\sqrt{n}}^{t} h(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du, & \text{if } t \in (-\delta\sqrt{n}, \delta\sqrt{n}) \\ 0, & \text{otherwise.} \end{cases}$$

Note that f is continuous on $[-\delta\sqrt{n}, \delta\sqrt{n}]$, differentiable on $(-\delta\sqrt{n}, \delta\sqrt{n})$ and $f(-\delta\sqrt{n}) = f(\delta\sqrt{n}) = 0$. Moreover, on $(-\delta\sqrt{n}, \delta\sqrt{n})$, f solves the Stein equation associated to the distribution of $\sqrt{n}\left(\tilde{\boldsymbol{\theta}}_n - \hat{\theta}_n\right)$ for test function h, as described in [12, Section 2.1]. In other words,

(D.5) $$h(t) = f'(t) + f(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}\log \Pi_n(n^{-1/2}t+\hat{\theta}_n)\right), \qquad t \in \left(-\delta\sqrt{n}, \delta\sqrt{n}\right).$$

Now, by Taylor's theorem, we obtain that, for some $c \in (0,1)$,

$$\begin{aligned}
&\tilde{I}_1 \\
&\overset{(D.5)}{=} \left| \int_{-\delta/\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) + f(t)\left(\frac{\mathrm{d}}{\mathrm{d}t} L_n(n^{-1/2}t+\hat{\theta}_n) + \frac{\mathrm{d}}{\mathrm{d}t}\log\pi(n^{-1/2}t+\hat{\theta}_n)\right)\right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\le \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) + f(t)\left(\frac{L_n'(\hat{\theta}_n)}{\sqrt{n}} + \frac{tL_n''(\hat{\theta}_n)}{n} + \frac{t^2}{2n^{3/2}} L_n'''(\hat{\theta}_n + cn^{-1/2}t)\right)\right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\quad + \frac{1}{\sqrt{n}} \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} f(t) \frac{\pi'(n^{-1/2}t+\hat{\theta}_n)}{\pi(n^{-1/2}t+\hat{\theta}_n)} \cdot \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&= \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) + f(t)\left(-\frac{t}{\sigma_n^2} + \frac{t^2}{2n^{3/2}} L_n'''(\hat{\theta}_n + cn^{-1/2}t)\right)\right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\quad + \frac{1}{\sqrt{n}} \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} f(t) \frac{\pi'(n^{-1/2}t+\hat{\theta}_n)}{\pi(n^{-1/2}t+\hat{\theta}_n)} \cdot \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| \\
&\le \left| \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \left[f'(t) - \frac{tf(t)}{\sigma_n^2}\right] \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \right| + \frac{M_2}{2\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t^2 f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \\
&\quad + \frac{M_1}{\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \\
&\overset{(D.3)}{=} \frac{M_2}{2\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |t^2 f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt + \frac{M_1}{\sqrt{n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |f(t)| \frac{e^{-t^2/(2\sigma_n^2)}}{\sqrt{2\pi\sigma_n^2}} dt \\
&\overset{(D.4)}{=} \frac{M_2}{2\sqrt{2\pi\sigma_n^2}\, n} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \frac{t^2 e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)} \left| \int_{-\delta\sqrt{n}}^{t} h(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du \right| dt \\
&\quad + \frac{M_1}{\sqrt{2\pi\sigma_n^2}\, n} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} \frac{e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)} \left| \int_{-\delta\sqrt{n}}^{t} h(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du \right| dt \\
&= \frac{M_2}{2\sqrt{2\pi\sigma_n^2}\, n} \int_{-\delta\sqrt{n}}^{0} \frac{t^2 e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)} \left| \int_{-\delta\sqrt{n}}^{t} h(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du \right| dt \\
&\quad + \frac{M_2}{2\sqrt{2\pi\sigma_n^2}\, n} \int_{0}^{\delta\sqrt{n}} \frac{t^2 e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)} \left| \int_{t}^{\delta\sqrt{n}} h(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du \right| dt
\end{aligned}$$

$$
\begin{aligned}
&+\frac{M_1}{\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}\frac{e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)}\left|\int_{-\delta\sqrt{n}}^{t}h(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du\right|dt\\
&+\frac{M_1}{\sqrt{2\pi\sigma_n^2\, n}}\int_{0}^{\delta\sqrt{n}}\frac{e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)}\left|\int_{t}^{\delta\sqrt{n}}h(u)\Pi_n(n^{-1/2}u+\hat{\theta}_n)du\right|dt\\
=:&\tilde{I}_{1,1}+\tilde{I}_{1,2}+\tilde{I}_{1,3}+\tilde{I}_{1,4}.
\end{aligned}
$$

Now, note that, for some $c_1, c_2 \in (0,1)$,

$$
\begin{aligned}
\tilde{I}_{1,1}\leq&\frac{M_2}{2\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}\frac{t^2e^{-t^2/(2\sigma_n^2)}}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)}\int_{-\delta\sqrt{n}}^{t}|h(u)|\Pi_n(n^{-1/2}u+\hat{\theta}_n)\,du\,dt\\
=&\frac{M_2}{2\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|h(u)|\int_{u}^{0}t^2e^{-t^2/(2\sigma_n^2)}\frac{\Pi_n(n^{-1/2}u+\hat{\theta}_n)}{\Pi_n(n^{-1/2}t+\hat{\theta}_n)}\,dt\,du\\
\leq&\frac{\widetilde{M}_1\hat{M}_1M_2}{2\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|h(u)|\int_{u}^{0}t^2e^{-t^2/(2\sigma_n^2)}\exp\Big[L_n(\hat{\theta}_n)+\frac{u}{\sqrt{n}}L_n'(\hat{\theta}_n)+\frac{u^2}{2n}L_n''(\hat{\theta}_n)\\
&\qquad\qquad+\frac{u^3}{6n^{3/2}}L_n'''(\hat{\theta}_n+c_1n^{-1/2}u)\Big]\\
&\cdot\exp\Big[-L_n(\hat{\theta}_n)-\frac{t}{\sqrt{n}}L_n'(\hat{\theta}_n)-\frac{t^2}{2n}L_n''(\hat{\theta}_n)-\frac{t^3}{6n^{3/2}}L_n'''(\hat{\theta}_n+c_2n^{-1/2}t)\Big]dt\,du\\
\leq&\frac{\widetilde{M}_1\hat{M}_1M_2}{2\sqrt{2\pi\sigma_n^2 n}}\int_{-\delta\sqrt{n}}^{0}|h(u)|e^{-u^2/(2\sigma_n^2)}e^{\delta M_2u^2/6}\int_{u}^{0}t^2e^{\delta M_2t^2/6}\,dt\,du\\
\leq&\frac{3\widetilde{M}_1\hat{M}_1}{\delta\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|uh(u)|\left(e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}-e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)u^2}\right)du.
\end{aligned}
$$

By a similar argument,

$$
\begin{aligned}
\tilde{I}_{1,2}\leq&\frac{3\widetilde{M}_1\hat{M}_1}{\delta\sqrt{2\pi\sigma_n^2\, n}}\int_{0}^{\delta\sqrt{n}}|uh(u)|\left(e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}-e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)u^2}\right)du.\\
\tilde{I}_{1,3}\leq&\frac{\widetilde{M}_1\hat{M}_1M_1}{\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{0}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}du.\\
\tilde{I}_{1,4}\leq&\frac{\widetilde{M}_1\hat{M}_1M_1}{\sqrt{2\pi\sigma_n^2\, n}}\int_{0}^{\delta\sqrt{n}}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}du.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
\tilde{I}_1\leq&\frac{2\widetilde{M}_1\hat{M}_1\left(M_1+\frac{3}{\delta}\right)}{\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{3}\right)u^2}du\\
&-\frac{6\widetilde{M}_1\hat{M}_1}{\delta\sqrt{2\pi\sigma_n^2\, n}}\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|uh(u)|e^{-\left(\frac{1}{2\sigma_n^2}-\frac{\delta M_2}{6}\right)u^2}du. \qquad \text{(D.6)}
\end{aligned}
$$

Now, by Taylor's expansion:

$$
\frac{n^{-1/2}}{C_n^{MLE}}\left|\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}g(t)\Pi_n(n^{-1/2}t+\hat{\theta}_n)dt\right|\leq\frac{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}|g(t)|\Pi_n(n^{-1/2}t+\hat{\theta}_n)dt}{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}}\Pi_n(n^{-1/2}u+\hat{\theta}_n)du}
$$

$$\leq \widetilde{M}_1 \hat{M}_1 \frac{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |g(t)| e^{-(1/(2\sigma_n^2)-\delta M_2/6)t^2} dt}{\int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} e^{-(1/(2\sigma_n^2)+\delta M_2/6)u^2} du}$$

$$\leq \widetilde{M}_1 \hat{M}_1 \frac{\sqrt{\frac{1}{2\sigma_n^2} + \frac{\delta M_2}{6}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |g(t)| e^{-(1/(2\sigma_n^2)-\delta M_2/6)t^2} dt}{\sqrt{2\pi}\left(1 - 2e^{-\delta^2 n(1/(2\sigma_n^2)+\delta M_2/6)}\right)}. \tag{D.7}$$

Equations (D.6) and (D.7), together with a standard expression for the normal first absolute moment now yield that

$$\tilde{I}_1 \leq \frac{2\widetilde{M}_1 \hat{M}_1}{\sqrt{2\pi\sigma_n^2 n}} \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |ug(u)| \left[\left(M_1 + \frac{3}{\delta}\right) e^{-\left(\frac{1}{2\sigma_n^2} - \frac{\delta M_2}{3}\right)u^2} - \frac{3}{\delta} e^{-\left(\frac{1}{2\sigma_n^2} - \frac{\delta M_2}{6}\right)u^2}\right] du$$
$$+ \frac{2\sqrt{\frac{1}{2\sigma_n^2} + \frac{\delta M_2}{6}} \left(\widetilde{M}_1 \hat{M}_1\right)^2 \left(M_1 + \frac{3}{\delta}\right) \int_{-\delta\sqrt{n}}^{\delta\sqrt{n}} |g(u)| e^{-(1/(2\sigma_n^2)-\delta M_2/6)u^2} du}{\left(\frac{1}{2\sigma_n^2} - \frac{\delta M_2}{3}\right) \pi \sqrt{\sigma_n^2} \left(1 - 2e^{-\delta^2 n(1/(2\sigma_n^2)+\delta M_2/6)}\right) \sqrt{n}}$$
$$\cdot \left(\frac{M_1 + \frac{3}{\delta}}{\frac{1}{2\sigma_n^2} - \frac{\delta M_2}{3}} - \frac{3}{\delta\left(\frac{1}{2\sigma_n^2} - \frac{\delta M_2}{6}\right)} \right). \tag{D.8}$$

D.3. Conclusion. The final bound now follows from (D.2) and (D.8).

APPENDIX E: MORE DETAIL ON THE EXAMPLES

E.1. Calculations for Example 5.1.

E.1.1. *The MLE-centric approach.* Let $X_1, \ldots, X_n \geq 0$ be our data and assume that their sum is positive. We have

$$L_n(\theta) = -n\theta + (\log\theta)\left(\sum_{i=1}^n X_i\right) - \sum_{i=1}^n \log(X_i!)$$

The MLE is given by $\hat{\theta}_n = \overline{X}_n := \frac{1}{n}\sum_{i=1}^n X_i$. Also,

$$L_n'(\theta) = -n + \frac{n\overline{X}_n}{\theta}, \quad L_n''(\theta) = -\frac{n\overline{X}_n}{\theta^2}, \quad L_n'''(\theta) = \frac{2n\overline{X}_n}{\theta^3}.$$

We have $\sigma_n^2 := J_n(\hat{\theta}_n)^{-1} = \hat{\theta}_n^2 / \overline{X}_n = \left|\overline{X}_n\right|$.

Now, for $c \in (0,1)$ and $\delta = c\hat{\theta}_n$, we have that for $\theta \in (\hat{\theta}_n - \delta, \hat{\theta}_n + \delta) = \left(\overline{X}_n - c\overline{X}_n, \overline{X}_n + c\overline{X}_n\right)$,

$$\frac{|L_n'''(\theta)|}{n} = \frac{2\overline{X}_n}{|\theta|^3} \leq \frac{2}{(1-c)^3\left(\overline{X}_n\right)^2} =: M_2$$

Moreover, for θ, such that $\left|\theta - \hat{\theta}_n\right| > \delta = c\overline{X}_n$, i.e. for $\theta > \overline{X}_n + c\overline{X}_n$ or $\theta < \overline{X}_n - c\overline{X}_n$,

$$\frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n} \leq \max\left\{\frac{L_n((1+c)\overline{X}_n) - L_n(\overline{X}_n)}{n}, \frac{L_n((1-c)\overline{X}_n) - L_n(\overline{X}_n)}{n}\right\}$$
$$\leq \overline{X}_n \cdot \max\{\log(1+c) - c, \log(1-c) + c\} = [\log(1+c) - c]\,\overline{X}_n =: -\kappa.$$

We also need to make sure that $J_n(\hat{\theta}_n) > \delta M_2$, i.e. that $\frac{1}{\overline{X}_n} > \frac{2c}{(1-c)^3\overline{X}_n}$, which is true for all $0 < c \leq 0.229$.

Now, the gamma prior with shape α and rate β satisfies

$$\pi'(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left[(\alpha-1)\theta^{\alpha-2}e^{-\beta\theta} - \beta\theta^{\alpha-1}e^{-\beta\theta}\right].$$

Note that, for $\alpha < 1$,

$$\sup_{\theta\in((1-c)\overline{X}_n,(1+c)\overline{X}_n)} |\pi'(\theta)| \leq \frac{\beta^\alpha}{\Gamma(\alpha)} e^{-\beta(1-c)\overline{X}_n}\overline{X}_n^{\alpha-2}(1-c)^{\alpha-2}\left[(1-\alpha)+\beta(1-c)\overline{X}_n\right]$$

$$\sup_{\theta\in((1-c)\overline{X}_n,(1+c)\overline{X}_n)} |\pi(\theta)| \leq \frac{\beta^\alpha}{\Gamma(\alpha)}(1-c)^{\alpha-1}\overline{X}_n^{\alpha-1}e^{-\beta(1-c)\overline{X}_n}$$

$$\sup_{\theta\in((1-c)\overline{X}_n,(1+c)\overline{X}_n)} \frac{1}{|\pi(\theta)|} \leq \frac{\Gamma(\alpha)}{\beta^\alpha}(1+c)^{1-\alpha}\overline{X}_n^{1-\alpha}e^{\beta(1+c)\overline{X}_n}.$$

Finally, the bounds in the MLE-centric approach are computed under the assumption $\sqrt{\frac{\mathrm{Tr}\left[J_n(\hat{\theta}_n)^{-1}\right]}{n}} < \delta$. In our case, it says $\sqrt{\frac{\overline{X}_n}{n}} < c\overline{X}_n$, i.e. that $c > \frac{1}{\sqrt{n\overline{X}_n}}$. Therefore, assuming $\frac{1}{\sqrt{n\overline{X}_n}} < 0.299$, letting $c \in \left(\frac{1}{\sqrt{n\overline{X}_n}}, 0.229\right]$, and assuming the shape α of the gamma prior is smaller than one, we can set

a) $\delta = c\overline{X}_n$
b) $M_1 = (1-c)^{\alpha-2}(1+c)^{1-\alpha}\overline{X}_n^{-1}e^{2\beta c\overline{X}_n}\left[(1-\alpha)+\beta(1-c)\overline{X}_n\right]$
c) $\widetilde{M}_1 = \frac{\beta^\alpha}{\Gamma(\alpha)}(1-c)^{\alpha-1}\overline{X}_n^{\alpha-1}e^{-\beta(1-c)\overline{X}_n}$
d) $\hat{M}_1 = \frac{\Gamma(\alpha)}{\beta^\alpha}(1+c)^{1-\alpha}\overline{X}_n^{1-\alpha}e^{\beta(1+c)\overline{X}_n}$
e) $M_2 = \frac{2}{(1-c)^3(\overline{X}_n)^2}$
f) $J_n(\hat{\theta}_n) = \overline{X}_n^{-1}$
g) $\kappa = [c - \log(1+c)]\,\overline{X}_n$.

The concrete choice of c may be optimized numerically.

E.1.2. *The MAP-centric approach.* Let us still assume $\alpha < 1$. We note that

$$\overline{L}_n(\theta) = -n\theta + n(\log\theta)\overline{X}_n - \sum_{i=1}^n \log(X_i!) + \alpha\log(\beta) - \log(\Gamma(\alpha)) + (\alpha-1)\log\theta - \beta\theta;$$

$$\overline{L}_n'(\theta) = -n + \frac{n\overline{X}_n}{\theta} + \frac{(\alpha-1)}{\theta} - \beta = 0 \quad \text{iff} \quad \theta = \bar{\theta}_n := \frac{n\overline{X}_n + (\alpha-1)}{(n+\beta)};$$

$$\overline{L}_n''(\theta) = -\frac{n\overline{X}_n + \alpha - 1}{\theta^2} \quad \text{and so} \quad \bar{J}_n(\bar{\theta}_n) = \frac{(n+\beta)^2}{n(n\overline{X}_n + \alpha - 1)};$$

$$\overline{L}_n'''(\theta) = 2\frac{n\overline{X}_n + \alpha - 1}{\theta^3}$$

and so for $\bar{c} \in (0,1), \bar{\delta} = \bar{c}\bar{\theta}_n$ and $\theta \in (\bar{\theta}_n - \bar{\delta}, \bar{\theta}_n + \bar{\delta}) = \left((1-\bar{c})\bar{\theta}_n, (1+\bar{c})\bar{\theta}_n\right)$, we have that

$$\frac{1}{n}|\overline{L}_n'''(\theta)| \leq \frac{2(n+\beta)^3}{n(n\overline{X}_n + \alpha - 1)^2(1-\bar{c})^3} =: \overline{M}_2.$$

Now, we require that $\bar{J}_n(\bar{\theta}_n) > \bar{\delta}\,\overline{M}_2$, which means that

$$\frac{(n+\beta)^2}{n(n\overline{X}_n + \alpha - 1)} > \frac{2\bar{c}(n+\beta)^2}{n(n\overline{X}_n + \alpha - 1)(1-\bar{c})^3}, \quad \text{which holds for} \quad \bar{c} \in (0, 0.229], \text{ if } \quad n\overline{X}_n > 1 - \alpha.$$

We'll also want to make sure that $\bar{\delta} = \bar{c}\frac{n\overline{X}_n+(\alpha-1)}{n+\beta} \geq \|\bar{\theta}_n - \hat{\theta}_n\| = \frac{\beta\overline{X}_n+1-\alpha}{n+\beta}$. Assuming that $n\overline{X}_n > 1-\alpha$, this translates to $\bar{c} \geq \frac{1}{n} \cdot \frac{\beta\overline{X}_n+1-\alpha}{\overline{X}_n+(\alpha-1)/n}$. Moreover, we require that $\bar{\delta} = \bar{c}\frac{n\overline{X}_n+(\alpha-1)}{n+\beta} > \sqrt{\frac{1}{n\bar{J}_n(\bar{\theta}_n)}} = \sqrt{\frac{n\overline{X}_n+\alpha-1}{(n+\beta)^2}}$, which is equivalent to saying that $\bar{c} > \sqrt{\frac{1}{n\overline{X}_n+\alpha-1}}$.

Finally, the value of $\bar{\kappa}$ may be obtained in the following way:

$$\bar{\kappa} := -\max\Bigg\{ -\bar{\delta} + \|\bar{\theta}_n - \hat{\theta}_n\| + \hat{\theta}_n \log\left(\frac{\hat{\theta}_n + \bar{\delta} - \|\bar{\theta}_n - \hat{\theta}_n\|}{\hat{\theta}_n}\right), \\ \bar{\delta} - \|\bar{\theta}_n - \hat{\theta}_n\| + \hat{\theta}_n \log\left(\frac{\hat{\theta}_n - \bar{\delta} + \|\bar{\theta}_n - \hat{\theta}_n\|}{\hat{\theta}_n}\right)\Bigg\}$$

$$= -\overline{X}_n\left\{\frac{\beta - \bar{c}n + (1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta} + \log\left(1 + \frac{\bar{c}n - \beta - (1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta}\right)\right\}.$$

Therefore, in addition to the values we listed at the end of Subsection E.1.2, we have the following. We assume that $\alpha < 1$, $n\overline{X}_n > 1-\alpha$ and $\max\left\{\sqrt{\frac{1}{n\overline{X}_n+\alpha-1}}, \frac{1}{n} \cdot \frac{\beta\overline{X}_n+1-\alpha}{\overline{X}_n+(\alpha-1)/n}\right\} < 0.229$. We let $\bar{c} \in \left(\frac{1}{n} \cdot \frac{\beta\overline{X}_n+1-\alpha}{\overline{X}_n+(\alpha-1)/n}, 0.229\right]$. Then

i) $\bar{\delta} = \bar{c}\frac{n\overline{X}_n+(\alpha-1)}{(n+\beta)}$
ii) $\bar{J}_n(\bar{\theta}_n) = \frac{(n+\beta)^2}{n(n\overline{X}_n+\alpha-1)}$
iii) $\overline{M}_2 = \frac{2\cdot(n+\beta)^3}{(1-\bar{c})^3 n(n\overline{X}_n+\alpha-1)^2}$
iv) $\bar{\kappa} = -\overline{X}_n\left\{\frac{\beta-\bar{c}n+(1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta} + \log\left(1 + \frac{\bar{c}n-\beta-(1-\alpha)(1+\bar{c})/\overline{X}_n}{n+\beta}\right)\right\}$.

E.2. Calculations for Example 5.2: the MAP-centric approach. Let k be the shape of the Weibull and let $X_1, \ldots, X_n \geq 0$ be our data. Our log-likelihood is given by:

$$L_n(\theta) = n\left[\log(k) - \log(\theta)\right] + (k-1)\sum_{i=1}^{n}\log(X_i) - \frac{\sum_{i=1}^n X_i^k}{\theta}$$

E.2.1. *Calculating $\hat{\theta}_n$ and $J_n(\hat{\theta}_n)$.* Now

$$L_n'(\theta) = -\frac{n}{\theta} + \frac{\sum_{i=1}^n X_i^k}{\theta^2}, \quad L_n''(\theta) = \frac{n}{\theta^2} - \frac{2\sum_{i=1}^n X_i^k}{\theta^3}$$

The MLE is $\hat{\theta}_n = \frac{\sum_{i=1}^n X_i^k}{n} =: \overline{X^k}(n)$. We have that

$$J_n(\hat{\theta}_n) = -\frac{1}{\hat{\theta}_n^2} + \frac{2}{\hat{\theta}_n^2} = \frac{1}{\hat{\theta}_n^2}.$$

E.2.2. *Calculating M_2.* Now, note that

$$L_n'''(\theta) = \frac{6\sum_{i=1}^n X_i^k - 2n\theta}{\theta^4}, \quad L_n^{(4)}(\theta) = \frac{6n\theta - 24\sum_{i=1}^n X_i^k}{\theta^5}, \quad L_n^{(5)}(\theta) = \frac{120\sum_{i=1}^n X_i^k - 24n\theta}{\theta^6}.$$

Therefore, for $0 < \theta < 2\hat{\theta}_n$, L_n''' is decreasing and positive. This means that, if we let $0 < \delta < \hat{\theta}_n$, then

$$\sup_{\theta \in (\hat{\theta}_n - \delta, \hat{\theta}_n + \delta)} \frac{|L_n'''(\theta)|}{n} \leq \frac{L_n'''(\hat{\theta}_n - \delta)}{n} =: M_2.$$

E.2.3. *Calculating $\hat{M}_1$.* Now, for a given shape $\alpha > 0$ and scale $\beta > 0$,

$$\pi(\theta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \left(\frac{1}{\theta}\right)^{\alpha+1} \exp\left(-\beta/\theta\right), \quad \pi'(\theta) = \frac{\beta^\alpha \exp\left(-\beta/\theta\right)}{\Gamma(\alpha)\theta^{\alpha+2}} \left(-\alpha - 1 + \frac{\beta}{\theta}\right)$$

and it follows that, for $\theta > \frac{\beta}{\alpha+1}$, $\frac{1}{\pi(\theta)}$ is increasing and it is decreasing otherwise. Therefore:

$$\sup_{\theta \in (\hat{\theta}_n - \hat{\delta}, \hat{\theta}_n + \delta)} \left|\frac{1}{\pi(\theta)}\right| \leq \max\left\{\frac{1}{\pi(\hat{\theta}_n - \delta)}, \frac{1}{\pi(\hat{\theta}_n + \delta)}\right\} =: \hat{M}_1.$$

E.2.4. *Calculating $\bar{\theta}_n$ and $\bar{J}_n(\bar{\theta}_n)$.* Now

$$\overline{L}_n(\theta) = n\left[\log(k) - \log(\theta)\right] + (k-1)\sum_{i=1}^n \log(X_i) - \frac{\sum_{i=1}^n X_i^k}{\theta} - (\alpha+1)\log(\theta) - \frac{\beta}{\theta}.$$

Therefore,

$$\overline{L}_n(\theta)' = -\frac{n}{\theta} + \frac{\sum_{i=1}^n X_i^k}{\theta^2} - \frac{\alpha+1}{\theta} + \frac{\beta}{\theta^2}$$

$$\overline{L}_n(\theta)'' = \frac{n}{\theta^2} - \frac{2\sum_{i=1}^n X_i^k}{\theta^3} + \frac{\alpha+1}{\theta^2} - \frac{2\beta}{\theta^3}$$

and the MAP $\bar{\theta}_n$ is given by

$$(n+\alpha+1)\bar{\theta}_n = \beta + \sum_{i=1}^n X_i^k \quad \Leftrightarrow \quad \bar{\theta}_n = \frac{\beta + \sum_{i=1}^n X_i^k}{n+\alpha+1}.$$

Moreover,

$$\bar{J}_n(\bar{\theta}_n) = -\frac{1}{\bar{\theta}_n^2} + \frac{2\sum_{i=1}^n X_i^k}{n\bar{\theta}_n^3} - \frac{\alpha+1}{n\bar{\theta}_n^2} + \frac{2\beta}{n\bar{\theta}_n^3}.$$

E.2.5. *Calculating $\bar{M}_2$.* Now, note that

$$\overline{L}_n'''(\theta) = -\frac{2n}{\theta^3} + \frac{6\sum_{i=1}^n X_i^k}{\theta^4} - \frac{2(\alpha+1)}{\theta^3} + \frac{6\beta}{\theta^4}, \quad \overline{L}_n^{(4)}(\theta) = \frac{6n}{\theta^4} - \frac{24\sum_{i=1}^n X_i^k}{\theta^5} + \frac{6(\alpha+1)}{\theta^4} - \frac{24\beta}{\theta^5}.$$

Therefore $\overline{L}_n'''$ is increasing if and only if

$$6(n+\alpha+1)\theta > 24\left(\beta + \sum_{i=1}^n X_i^k\right) \quad \Leftrightarrow \theta > 4\bar{\theta}_n$$

This means that, for $\bar{\delta} \in \left(0, \bar{\theta}_n\right)$ and $\theta \in (\bar{\theta}_n - \bar{\delta}, \bar{\theta}_n + \bar{\delta})$, $\overline{L}_n'''$ is decreasing and

$$\sup_{|\theta - \bar{\theta}_n| < \bar{\delta}} \frac{|\overline{L}_n'''(\theta)|}{n} \leq \frac{|\overline{L}_n'''(\bar{\theta}_n - \bar{\delta})|}{n} =: \overline{M}_2.$$

E.2.6. *Calculating $\bar{\kappa}$.* Now, note that, for $\theta < \hat{\theta}_n$, L_n, is increasing and otherwise it's decreasing. This means that,

$$\sup_{|\theta - \hat{\theta}_n| > \bar{\delta} - |\hat{\theta}_n - \bar{\theta}_n|} \frac{L_n(\theta) - L_n(\hat{\theta}_n)}{n}$$

$$\leq \max\left\{\frac{L_n(\hat{\theta}_n - \bar{\delta} + |\hat{\theta}_n - \bar{\theta}_n|) - L_n(\hat{\theta}_n)}{n}, \frac{L_n(\hat{\theta}_n + \bar{\delta} - |\hat{\theta}_n - \bar{\theta}_n|) - L_n(\hat{\theta}_n)}{n}\right\} =: -\bar{\kappa}.$$

E.2.7. *Constraints on $\bar{\delta}$ and δ.* Finally, we need the derive the constraints on $\bar{\delta}$. We first assume that $0 < \bar{\delta} < \bar{\theta}_n$. We also suppose that $\bar{\delta} > \max\left(\|\bar{\theta}_n - \hat{\theta}_n\|, \frac{1}{\sqrt{n\bar{J}_n(\bar{\theta}_n)}}\right)$ and conditions on how large n needs to be for this to hold are easy to obtain numerically. We also require $\bar{\delta} < \frac{\lambda(\bar{\theta}_n)}{\overline{M}_2}$. Looking closer at this last condition, we require that:

$$\bar{\delta} < \overline{\lambda}_{\min}(\bar{\theta}_n)\left(\bar{\theta}_n - \bar{\delta}\right)^4 \left[\left(-2 - \frac{2\alpha+2}{n}\right)\left(\bar{\theta}_n - \bar{\delta}\right) + 6\left(\frac{\beta}{n} + \hat{\theta}_n\right)\right]^{-1}$$

Letting $0 < \tilde{\delta} := \bar{\theta}_n - \bar{\delta} < \bar{\theta}_n$, we therefore require

$$\tilde{\delta} + \overline{\lambda}_{\min}(\bar{\theta}_n)\tilde{\delta}^4 \left[\left(-2 - \frac{2\alpha+2}{n}\right)\tilde{\delta} + 6\left(\frac{\beta}{n} + \hat{\theta}_n\right)\right]^{-1} > \bar{\theta}_n$$

which is equivalent to:

$$\tilde{\delta} + \tilde{\delta}^4\overline{\lambda}_{\min}(\bar{\theta}_n)\left(2 + \frac{2\alpha+2}{n}\right)^{-1}\left[3\bar{\theta}_n - \tilde{\delta}\right]^{-1} > \bar{\theta}_n.$$

This condition will be satisfied if

$$\tilde{\delta} + \tilde{\delta}^4\overline{\lambda}_{\min}(\bar{\theta}_n)\left(2 + \frac{2\alpha+2}{n}\right)^{-1}\left[3\bar{\theta}_n\right]^{-1} > \bar{\theta}_n. \tag{E.1}$$

The left-hand side of (E.1) is increasing in $\tilde{\delta}$ and is clearly strictly greater than $\bar{\theta}_n$ for $\tilde{\delta} = \bar{\theta}_n$. This means that there exists a choice of $\bar{\delta}$ that yields $\bar{\delta} < \frac{\overline{\lambda}_{\min}(\bar{\theta}_n)}{\overline{M}_1}$ and the set of such choices can be obtained by solving (E.1) numerically. Finally, in order to make sure that the condition on δ is satisfied, we just need to check numerically how large n needs to be so that $\delta > \frac{1}{\sqrt{nJ_n(\hat{\theta}_n)}}$.