Random forward models and log-likelihoods in Bayesian inverse problems

H. C. Lie* T. J. Sullivan†‡ A. L. Teckentrup§¶

February 1, 2018

Abstract: We consider the use of randomised forward models and log-likelihoods within the Bayesian approach to inverse problems. Such random approximations to the exact forward model or log-likelihood arise naturally when a computationally expensive model is approximated using a cheaper stochastic surrogate, as in Gaussian process emulation (kriging), or in the field of probabilistic numerical methods. We show that the Hellinger distance between the exact and approximate Bayesian posteriors is bounded by moments of the difference between the true and approximate log-likelihoods. Example applications of these stability results are given for randomised misfit models in large data applications and the probabilistic solution of ordinary differential equations.

Keywords: Bayesian inverse problem, random likelihood, surrogate model, posterior consistency, uncertainty quantification, randomised misfit, probabilistic numerics.

2010 Mathematics Subject Classification: 62F15, 62G08, 65C99, 65D05, 65D30, 65J22, 68W20.

1 Introduction

Inverse problems are ubiquitous in the applied sciences and in recent years renewed attention has been paid to their mathematical and statistical foundations (Evans and Stark, 2002; Kaipio and Somersalo, 2005; Stuart, 2010). Questions of well-posedness — i.e. the existence, uniqueness, and stability of solutions — have been of particular interest for infinite-dimensional/non-parametric inverse problems because of the need to ensure stable and discretisation-independent inferences (Lassas and Siltanen, 2004) and develop algorithms that scale well with respect to high discretisation dimension (Cotter et al., 2013).

This paper considers the stability of the posterior distribution in a Bayesian inverse problem (BIP) when an accurate but computationally intractable forward model or likelihood is replaced by a random surrogate or emulator. Such stochastic surrogates arise often in practice. For example, an expensive forward model such as the solution of a PDE may replaced by a kriging/Gaussian process model (Stuart and Teckentrup, 2017). In the realm of "big data" a residual vector of prohibitively high dimension may be randomly subsampled or orthogonally projected onto a randomly-chosen low-dimensional subspace (Nemirovski et al., 2008; Le et al., 2017). In the field of probabilistic numerical methods (Hennig et al., 2015), a deterministic dynamical system may be solved stochastically, with the stochasticity representing epistemic uncertainty about the behaviour of the system below the temporal or spatial grid scale (Conrad et al., 2016; Lie et al., 2017).

In each of the above-mentioned settings, the stochasticity in the forward model propagates to associated inverse problems, so that the Bayesian posterior becomes a *random measure*, which we define precisely

*Institute of Mathematics, Free University of Berlin, Arnimallee 6, 14195 Berlin, Germany, hlie@math.fu-berlin.de
†Institute of Mathematics, Free University of Berlin, Arnimallee 6, 14195 Berlin, Germany, t.j.sullivan@fu-berlin.de
‡Zuse Institute Berlin, Takustrasse 7, 14195 Berlin, Germany, sullivan@zib.de
§School of Mathematics, University of Edinburgh, UK, a.teckentrup@ed.ac.uk
¶The Alan Turing Institute, 96 Euston Road, London, NW1 2DB, UK

in (3.1). Alternatively, one may choose to average over the randomness to obtain a *marginal posterior*, which we define precisely in (3.2). It is natural to ask in which sense the approximate posterior (either the random or the marginal version) is close to the real posterior of interest.

In earlier work, Stuart and Teckentrup (2017) examined the case in which the random surrogate was a Gaussian process. More precisely, the object subjected to Gaussian process emulation was either the forward model (i.e. the parameter-to-observation map) or the negative log-likelihood. The prior Gaussian process was assumed to be continuous, and was then conditioned upon finitely many observations (i.e. pointwise evaluations) of the parameter-to-observation map or negative log-likelihood as appropriate. That paper provided error bounds on the Hellinger distance between the BIP's exact posterior distribution and various approximations based on the Gaussian process emulator, namely approximations based on the mean of the predictive (i.e. conditioned) Gaussian process, as well as approximations based on the full Gaussian process emulator. Those results showed that the Hellinger distance between the exact BIP posterior and its approximations can be bounded by moments of the error in the emulator.

In this paper, we extend the analysis of Stuart and Teckentrup (2017) to consider more general (i.e. non-Gaussian) random approximations to forward models and log-likelihoods, and quantify the impact upon the posterior measure in a BIP. After establishing some notation in Section 2, we state the main approximation theorems in Section 3. Section 4 gives an application of the general theory to random misfit models, in which high-dimensional data are rendered tractable by projection into a randomly-chosen low-dimensional subspace. Section 5 gives an application to the stochastic numerical solution of deterministic dynamical systems, in which the stochasticity is a device used to represent the impact of numerical discretisation uncertainty. The proofs of all theorems are deferred to an appendix located after the bibliographic references.

2 Setup and notation

2.1 Spaces of probability measures

$(\Omega, \mathcal{F}, \mathbb{P})$ will denote an abstract probability space, assumed to be rich enough to serve as a common domain for all random variables of interest.

$\mathcal{M}_1(\mathcal{U})$ will denote the space of probability measures on the Borel σ-algebra on a topological space $\mathcal{U}$; in practice, $\mathcal{U}$ will be a separable Banach space.

When $\mu \in \mathcal{M}_1(\mathcal{U})$, integration of a measurable function (random variable) $f\colon \mathcal{U} \to \mathbb{R}$ will also be denoted by expectation, i.e. $\mathbb{E}_\mu[f] := \int_{\mathcal{U}} f(u)\,\mathrm{d}\mu(u)$.

The space $\mathcal{M}_1(\mathcal{U})$ will be endowed with the Hellinger metric $d_{\mathrm{H}}\colon \mathcal{M}_1(\mathcal{U})^2 \to \mathbb{R}_{\geq 0}$: for probability measures μ and ν on $\mathcal{U}$ that are both absolutely continuous with respect to a common reference measure r, such as $\pi := \mu + \nu$,

$$d_{\mathrm{H}}(\mu,\nu)^2 := \frac{1}{2}\int_{\mathcal{U}} \left| \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\pi}(u)} - \sqrt{\frac{\mathrm{d}\nu}{\mathrm{d}\pi}} \right|^2 \mathrm{d}\pi(u) = 1 - \int_{\mathcal{U}} \sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\pi}(u)\frac{\mathrm{d}\nu}{\mathrm{d}\pi}(u)}\,\mathrm{d}\pi(u) = 1 - \mathbb{E}_\nu\left[\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\nu}}\right]. \qquad (2.1)$$

The Hellinger distance is independent of the choice of reference measure π and defines a metric on $\mathcal{M}_1(\mathcal{U})$ (Bogachev, 2007, Lemma 4.7.35–36) with respect to which $\mathcal{M}_1(\mathcal{U})$ evidently has diameter at most 1. The Hellinger topology coincides with the total variation topology (Kraft, 1955); the Hellinger topology is strictly weaker than the Kullback–Leibler (relative entropy) topology (Pinsker, 1964); all these topologies are strictly stronger than the topology of weak convergence of measures.

As used in Sections 3–5, the Hellinger metric is useful for uncertainty quantification when assessing the similarity of Bayesian posterior probability distributions, since expected values of square-integrable functions are Lipschitz continuous with respect to the Hellinger metric:

$$\left|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]\right| \leq \sqrt{2}\sqrt{\mathbb{E}_\mu\left[|f|^2\right] + \mathbb{E}_\nu\left[|f|^2\right]}\, d_{\mathrm{H}}(\mu,\nu) \qquad (2.2)$$

when $f \in L^2(\mathcal{U},\mu) \cap L^2(\mathcal{U},\nu)$. In particular, for bounded f, $|\mathbb{E}_\mu[f] - \mathbb{E}_\nu[f]| \leq 2\|f\|_\infty d_{\mathrm{H}}(\mu,\nu)$.

2.2 Bayesian inverse problems

By an *inverse problem* we mean the recovery of $u \in \mathcal{U}$ from an imperfect observation $y \in \mathcal{Y}$ of $G(u)$, for a known forward operator $G\colon \mathcal{U} \to \mathcal{Y}$. In practice, the operator G may arise as the composition $G = O \circ S$ of the solution operator $S\colon \mathcal{U} \to \mathcal{V}$ of a system of ordinary or partial differential equations with an observation operator $O\colon \mathcal{V} \to \mathcal{Y}$, and it is typically the case that $\mathcal{Y} = \mathbb{R}^J$ for some $J \in \mathbb{N}$, whereas $\mathcal{U}$ and $\mathcal{V}$ can have infinite dimension. For simplicity, we assume an additive noise model

$$y = G(u) + \eta, \tag{2.3}$$

where the statistics but not the realisation of η are known. In the strict sense, this inverse problem is ill-posed in the sense that there may be no element u that satisfies (2.3), or there may be multiple such u that are highly sensitive to the observed data y.

The Bayesian perspective eases these problems by interpreting u, y, and η all as random variables or fields. Through knowledge of the distribution of η, (2.3) defines the conditional distribution of $y|u$. After positing a prior probability distribution $\mu_0 \in \mathcal{M}_1(\mathcal{U})$ for u, the Bayesian solution to the inverse problem is nothing other than the posterior distribution for the conditioned random variable $u|y$. This posterior measure, which we denote $\mu^y \in \mathcal{M}_1(\mathcal{U})$, is from the Bayesian point of view the proper synthesis of the prior information in μ_0 with the observed data y. The same posterior μ^y can also be arrived at via the minimisation of penalised Kullback–Leibler, χ^2, or Dirichlet energies (Dupuis and Ellis, 1997; Jordan and Kinderlehrer, 1996; Ohta and Takatsu, 2011), where the penalisation again expresses compromise between fidelity to the prior and fidelity to the data.

The rigorous formulation of Bayes' formula for this context requires careful treatment and some further notation (Stuart, 2010). The pair (u, y) is assumed to be a well-defined random variable with values in $\mathcal{U} \times \mathcal{Y}$. The marginal distribution of u is the Bayesian prior $\mu_0 \in \mathcal{M}_1(\mathcal{U})$. The observational noise η is distributed according to $\mathbb{Q}_0 \in \mathcal{M}_1(\mathcal{Y})$, independently of u. The random variable $y|u$ is distributed according to $\mathbb{Q}_u$, the translate of $\mathbb{Q}_0$ by $G(u)$, which is assumed to be absolutely continuous with respect to $\mathbb{Q}_0$, with

$$\frac{\mathrm{d}\mathbb{Q}_u}{\mathrm{d}\mathbb{Q}_0}(y) = \exp(-\Phi(u; y)).$$

The function $\Phi\colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ is called the *negative log-likelihood* or simply *potential*. In the elementary setting of centred Gaussian noise, $\eta \sim \mathcal{N}(0, \Gamma)$ on $\mathcal{Y} = \mathbb{R}^J$, and the potential is the non-negative quadratic misfit $\Phi(u; y) = \frac{1}{2}\big\|\Gamma^{-1/2}(y - G(u))\big\|^2$. However, particularly for cases in which $\dim \mathcal{Y} = \infty$, it may be necessary to allow Φ to take negative values and even to be unbounded below (Stuart, 2010, Remark 3.8).

With this notation, Bayes' theorem is then as follows (Dashti and Stuart, 2016, Theorem 3.4):

Theorem 2.1 (Generalised Bayesian formula). *Suppose that $\Phi\colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ is $\mu_0 \otimes \mathbb{Q}_0$-measurable and that*

$$Z(y) \coloneqq \mathbb{E}_{\mu_0}\big[\exp(-\Phi(u; y))\big]$$

satisfies $0 < Z(y) < \infty$ for $\mathbb{Q}_0$-almost all $y \in \mathcal{Y}$. Then, for such y, the conditional distribution μ^y of $u|y$ exists and is absolutely continuous with respect to μ_0 with density

$$\frac{\mathrm{d}\mu^y}{\mathrm{d}\mu_0}(u) = \frac{\exp(-\Phi(u; y))}{Z(y)}. \tag{2.4}$$

Note that, for (2.4) to make sense, it is essential to check that $Z(y)$ is strictly positive and finite. Hereafter, to save space, we regard the data y as fixed, and hence write $\Phi(u)$ in place of $\Phi(u; y)$, Z in place of $Z(y)$, and μ in place of μ^y. In particular, we shall redefine the negative log-likelihood as a function $\Phi\colon \mathcal{U} \to \mathbb{R}$, instead of a function $\Phi\colon \mathcal{U} \times \mathcal{Y} \to \mathbb{R}$ as in Theorem 2.1 above.

From the perspective of numerical analysis, it is natural to ask about the well-posedness of the Bayesian posterior μ: is it stable when the prior μ_0, the potential Φ, or the observed data y are slightly perturbed, e.g. due to discretisation, truncation, or other numerical errors? For example, what is the impact of using an approximate numerical forward operator G_N in place of G, and hence an approximate $\Phi_N : \mathcal{U} \to \mathbb{R}$ in place of Φ? Here, we quantify stability in the Hellinger metric d_{H} from (2.1).

The stability of Bayesian inverse problems with respect to the prior is a difficult topic (Owhadi et al., 2015; Owhadi and Scovel, 2017) and we will not address it here. However, stability with respect to the observed data y and the log-likelihood Φ can be established. Such stability results were proved for Gaussian priors by Stuart (2010) and for more general priors by many contributions since then (Dashti et al., 2012; Hosseini, 2017; Hosseini and Nigam, 2017; Sullivan, 2017). Typical approximation theorems for the replacement of the potential Φ by a deterministic approximate potential Φ_N, leading to an approximate posterior μ_N, aim to transfer the convergence rate of the forward problem to the inverse problem, i.e. to prove an implication of the form

$$\big|\Phi(u) - \Phi_N(u)\big| \leq M(\|u\|)\psi(N) \implies d_{\mathrm{H}}\big(\mu, \mu_N\big) \leq C\psi(N),$$

where $M\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ is suitably well-behaved, $\psi\colon \mathbb{N} \to \mathbb{R}_{\geq 0}$ quantifies the convergence rate of the forward problem, and C is a constant. Following Stuart and Teckentrup (2017), the purpose of this article is to extend this paradigm and these approximation results to the case in which the approximation Φ_N is a *random* object.

3 Well-posed Bayesian inverse problems with random likelihoods

In many practical applications, the negative log-likelihood Φ is computationally too expensive or impossible to evaluate exactly, and in computations, one therefore often uses an approximation Φ_N of Φ. This leads to an approximation μ_N of the exact posterior distribution μ, and the aim is to show convergence, in a suitable sense, of μ_N to μ as the approximation error $\Phi_N - \Phi$ in the misfit potential tends to zero.

The focus of this work is on random approximations Φ_N. One particular example of such random approximations are the Gaussian process emulators analysed in Stuart and Teckentrup (2017); other examples include the randomised misfit models in Section 4 and the probabilistic numerical methods in Section 5. The present section extends the analysis of Stuart and Teckentrup (2017) from the case of Gaussian process approximations of forward models or log-likelihoods to more general non-Gaussian approximations. In doing so, more precise conditions are obtained for the exact Bayesian posterior to be well approximated by its random counterpart.

Let now $\Phi_N : \Omega \times \mathcal{U} \to \mathbb{R}$ be a measurable function that provides a random approximation to $\Phi\colon \mathcal{U} \to \mathbb{R}$, where we recall that we have fixed the data y. Let ν_N be a probability measure on Ω such that the distribution of Φ_N is given by $\nu_N \otimes \mu_0$. We assume throughout that the randomness in the approximation Φ_N of Φ is independent of the randomness in the parameters being inferred.

Replacing Φ by Φ_N in (2.4), we obtain the approximation μ_N^{sample}, the random measure given by

$$\begin{aligned} \frac{\mathrm{d}\mu_N^{\mathrm{sample}}}{\mathrm{d}\mu_0}(u) &\coloneqq \frac{\exp(-\Phi_N(u))}{Z_N^{\mathrm{sample}}}, \\ Z_N^{\mathrm{sample}} &\coloneqq \mathbb{E}_{\mu_0}\big[\exp(-\Phi_N(\,\cdot\,))\big] = \int_{\mathcal{U}} \exp(-\Phi_N(\,\cdot\,, u'))\,\mathrm{d}\mu_0(u'). \end{aligned} \tag{3.1}$$

Thus, the measure μ_N^{sample} is a random approximation of the deterministic measure μ, and the normalisation constant $Z_N^{\mathrm{sample}}\colon \Omega \to \mathbb{R}$ is a random variable. A deterministic approximation of the posterior distribution μ can now be obtained by taking the expected value of the random likelihood $\exp(-\Phi_N(u))$. This results in the marginal approximation defined by

$$\frac{\mathrm{d}\mu_N^{\mathrm{marginal}}}{\mathrm{d}\mu_0}(u) \coloneqq \frac{\mathbb{E}_{\nu_N}\big[\exp(-\Phi_N(u))\big]}{\mathbb{E}_{\nu_N}\big[Z_N^{\mathrm{sample}}\big]}, \tag{3.2}$$

where $\mathbb{E}_{\nu_N}[Z_N^{\mathrm{sample}}] = \int_\Omega Z_N^{\mathrm{sample}}(\omega)\,\mathrm{d}\nu_N(\omega)$. We note that an alternative deterministic approximation can be obtained by taking the expected value of the density $(Z_N^{\mathrm{sample}})^{-1}e^{-\Phi_N(u)}$ in (3.1) as a whole; however, the marginal approximation presented here appears more intuitive and more amenable to applications. Firstly, the marginal approximation provides a clear interpretation as the posterior distribution obtained by the approximation of the true data likelihood $\exp(-\Phi(u))$ by $\mathbb{E}_{\nu_N}\big[\exp(-\Phi_N(u))\big]$. Secondly, the marginal approximation is more amenable to sampling methods such as Markov chain Monte Carlo, with clear connections to the pseudo-marginal approach (Beaumont, 2003; Andrieu and Roberts, 2009).

3.1 Random misfit models

This section considers the general setting in which the deterministic potential Φ is approximated by a random potential $\Phi_N \sim \nu_N$. Recall from (2.4) that Z is the normalisation constant of μ, and that for μ to be well-defined, we must have that $0 < Z < \infty$. The following two results, Theorems 3.1 and 3.2, extend Theorems 4.9 and 4.11 respectively of Stuart and Teckentrup (2017), in which the approximation is a Gaussian process model:

Theorem 3.1 (Deterministic convergence of the marginal posterior)**.** *Suppose there exist positive scalars C_1, C_2, C_3, that do not depend on N, such that for the Hölder conjugate exponent pairs (p_1, p_1'), (p_2, p_2'), and (p_3, p_3'), we have*

(a) $\min\Big\{\big\|\mathbb{E}_{\nu_N}[\exp(-\Phi_N)]^{-1}\big\|_{L^{p_1}_{\mu_0}(\mathcal{U})}, \big\|\exp(\Phi)\big\|_{L^{p_1}_{\mu_0}(\mathcal{U})}\Big\} \leq C_1(p_1)$;

(b) $\left\|\mathbb{E}_{\nu_N}\left[\big(\exp(-\Phi) + \exp(-\Phi_N)\big)^{p_2}\right]^{1/p_2}\right\|_{L^{2p_1'p_3}_{\mu_0}(\mathcal{U})} \leq C_2(p_1, p_2, p_3)$;

(c) $C_3^{-1} \leq \mathbb{E}_{\nu_N}[Z_N^{\mathrm{sample}}] \leq C_3$.

Then there exists a scalar $C = C(C_1, C_2, C_3, Z) > 0$ that does not depend on N, such that

$$d_{\mathrm{H}}\big(\mu, \mu_N^{\mathrm{marginal}}\big) \leq C\left\|\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{p_2'}\right]^{1/p_2'}\right\|_{L^{2p_1'p_3'}_{\mu_0}(\mathcal{U})}, \tag{3.3a}$$

$$C(C_1, C_2, C_3, Z) = \left(\frac{C_1(p_1)}{Z} + C_3\max\big\{Z^{-3}, C_3^3\big\}\right)C_2^2(p_1, p_2, p_3). \tag{3.3b}$$

Theorem 3.2 (Mean-square convergence of the sample posterior)**.** *Suppose there exist positive scalars D_1, D_2, that do not depend on N, such that for Hölder conjugate exponent pairs (q_1, q_1') and (q_2, q_2'), we have*

(a) $\left\|\mathbb{E}_{\nu_N}\left[\big(e^{-\Phi/2} + e^{-\Phi_N/2}\big)^{2q_1}\right]^{1/q_1}\right\|_{L^{q_2}_{\mu_0}(\mathcal{U})} \leq D_1(q_1, q_2)$;

(b) $\left\|\mathbb{E}_{\nu_N}\left[\Big(Z_N^{\mathrm{sample}}\max\Big\{Z^{-3}, (Z_N^{\mathrm{sample}})^{-3}\Big\}\big(e^{-\Phi} + e^{-\Phi_N}\big)^2\Big)^{q_1}\right]^{1/q_1}\right\|_{L^{q_2}_{\mu_0}(\mathcal{U})} \leq D_2(q_1, q_2)$.

Then

$$\mathbb{E}_{\nu_N}\left[d_{\mathrm{H}}\big(\mu, \mu_N^{\mathrm{sample}}\big)^2\right]^{1/2} \leq (D_1 + D_2)\left\|\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{2q_1'}\right]^{1/2q_1'}\right\|_{L^{2q_2'}_{\mu_0}(\mathcal{U})}, \tag{3.4}$$

We now show that the assumptions of Theorems 3.1 and 3.2 are satisfied when the exact potential Φ and the approximation quality $\Phi_N \approx \Phi$ are suitably well behaved. Since $0 < Z < \infty$, it follows that there exists a strictly positive, finite scalar C_3 such that $C_3^{-1} < Z < C_3$.

Assumption 3.3. There exists $C_0 \in \mathbb{R}$ that does not depend on N, such that, for all $N \in \mathbb{N}$,

$$\Phi \geq -C_0 \quad \text{and} \quad \nu_N(\{\Phi_N \mid \Phi_N \geq -C_0\}) = 1, \tag{3.5}$$

and for any $0 < C_3 < +\infty$ with the property that $C_3^{-1} < Z < C_3$, there exists $N^*(C_3) \in \mathbb{N}$ such that, for all $N \geq N^*$,

$$\mathbb{E}_{\mu_0}[\mathbb{E}_{\nu_N}[|\Phi_N - \Phi|]] \leq \frac{1}{2\exp(C_0)}\min\left\{Z - \frac{1}{C_3}, C_3 - Z\right\}. \tag{3.6}$$

Lemma 3.4. *Suppose that Assumption 3.3 holds with C_0 as in (3.5) and C_3 and $N^*(C_3)$ as in (3.6), that $\exp(\Phi) \in L^{p^*}_{\mu_0}(\mathcal{U})$ for some $1 \leq p^* \leq +\infty$ with conjugate exponent $(p^*)'$, and there exists some $C_4 \in \mathbb{R}$ that does not depend on N, such that, for all $N \in \mathbb{N}$,*

$$\nu_N(\{\Phi_N \mid \mathbb{E}_{\mu_0}[\Phi_N] \leq C_4\}) = 1. \tag{3.7}$$

Then the hypotheses of Theorem 3.1 hold, with

$$p_1 = p^*,\ p_2 = p_3 = +\infty,\ C_1 = \|\exp(\Phi)\|_{L^{p^*}_{\mu_0}},\ C_2 = 2\exp(C_0),$$

and C_3 as above. Moreover, the hypotheses of Theorem 3.2 hold, with

$$q_1 = q_2 = \infty,\ D_1 = 4\exp(C_0),\ D_2 = 4\exp(3C_0)\max\{C_3^{-3}, \exp(3C_4)\}.$$

Lemma 3.5. *Suppose that Assumption 3.3 holds with C_0 as in (3.5) and C_3 and $N^*(C_3)$ as in (3.6), and that there exists some $2 < \rho^* < +\infty$ such that $\mathbb{E}_{\nu_N}[\exp(\rho^*\Phi_N)] \in L^1_{\mu_0}$. Then the hypotheses of Theorem 3.1 hold, with*

$$p_1 = \rho^*,\ p_2 = p_3 = +\infty,\ C_1 = \|\mathbb{E}_{\nu_N}[\exp(\rho^*\Phi_N)]\|^{1/\rho^*}_{L^1_{\mu_0}},\ C_2 = 2\exp(C_0),$$

and C_3 as above. Moreover, the hypotheses of Theorem 3.2 hold, with

$$\begin{aligned} q_1 &= \frac{\rho^*}{2}, & q_2 &= +\infty, \\ D_1 &= 4\exp(C_0), & D_2 &= 4\exp(2C_0)\Big(C_3^{-3}\exp(C_0) + \|\mathbb{E}_{\nu_N}[\exp(\rho^*\Phi_N)]\|^{2/\rho^*}_{L^1_{\mu_0}}\Big). \end{aligned}$$

In Lemmas 3.4 and 3.5 above, we have specified the largest possible values of the exponents that are compatible with the hypotheses. This is because later, in Theorem 3.9, we will want to use the smallest possible values of the corresponding *conjugate* exponents in the resulting inequalities (3.3a) and (3.4).

3.2 Random forward models in quadratic potentials

In many settings, the potentials Φ and Φ_N have a common form and differ only in the parameter-to-observable map. In this section we shall assume that Φ and Φ_N are quadratic misfits of the form

$$\Phi(u) = \frac{1}{2}\big\|\Gamma^{-1/2}(G(u) - y)\big\|^2_{\mathcal{Y}} \quad \text{and} \quad \Phi_N(u) = \frac{1}{2}\big\|\Gamma^{-1/2}(G_N(u) - y)\big\|^2_{\mathcal{Y}}, \tag{3.8}$$

corresponding to centred Gaussian observational noise with symmetric positive-definite covariance Γ. Again, we assume that G is deterministic while G_N is random. In this section, for this setting, we show how the quality of the approximation $G_N \approx G$ transfers to the approximation $\Phi_N \approx \Phi$, and hence to the approximation $\mu_N \approx \mu$ (for either the sample or marginal approximate posterior).

Pointwise in u and ω, the errors in the misfit and the forward model are related according to the following lemma.

Proposition 3.6. *Let Φ and Φ_N be defined as in (3.8), where $\mathcal{Y} = \mathbb{R}^n$ for some $n \in \mathbb{N}$ and the eigenvalues of the operator Γ are bounded away from zero. Then, for some $C = C(\Gamma) > 0$,*

$$\big|\Phi(u) - \Phi_N(u)\big| \le 2C_\Gamma\Big(\Phi(u)^{1/2}\|G_N(u) - G(u)\| + \|G(u) - G_N(u)\|^2\Big), \tag{3.9}$$

and hence, for $q \in [1,\infty)$ and all $u \in \mathcal{U}$,

$$\begin{aligned} \mathbb{E}_{\nu_N}\big[\big|\Phi(u) - \Phi_N(u)\big|^q\big]^{1/q} \le 4C_\Gamma\Big(&\Phi(u)^{q/2}\mathbb{E}_{\nu_N}[\|G_N(u) - G(u)\|^q] \\ &+ \mathbb{E}_{\nu_N}\big[\|G(u) - G_N(u)\|^{2q}\big]\Big)^{1/q}. \end{aligned} \tag{3.10}$$

Corollary 3.7. *Let $1 \le q \le s$, and suppose that $\Phi \in L^s_{\mu_0}$. If there exists an $N^* \in \mathbb{N}$ such that, for all $N \ge N^*$,*

$$\Big\|\mathbb{E}_{\nu_N}\big[\|G_N - G\|^{2q}\big]^{1/q}\Big\|_{L^s_{\mu_0}} \le 1,$$

then, there exists some $C = C(s) > 0$ that does not depend on N such that for all $N \ge N^*$,

$$\Big\|\mathbb{E}_{\nu_N}\big[|\Phi - \Phi_N|^q\big]^{1/q}\Big\|_{L^s_{\mu_0}} \le C\Big\|\mathbb{E}_{\nu_N}\big[\|G_N - G\|^{2q}\big]^{1/q}\Big\|^{1/2}_{L^s_{\mu_0}}$$

where $C(s) = (8C_\Gamma)\big(\mathbb{E}_{\mu_0}[\Phi^s]^{1/2} + 1\big)^{1/s}$ and C_Γ is as in Proposition 3.6.

Lemma 3.8. *Let Φ and Φ_N be as in (3.8). If, for some $q, s \ge 1$,*

$$\lim_{N\to\infty}\Big\|\mathbb{E}_{\nu_N}\big[\|G_N - G\|^{2q}\big]^{1/q}\Big\|_{L^s_{\mu_0}} = 0, \tag{3.11}$$

then Assumption 3.3 holds.

We shall use the preceding results to obtain bounds on the Hellinger distance in terms of errors in the forward model, of the following form: for $C, D > 0$ and $r_1, r_2, s_1, s_2 \geq 1$ that do not depend on N,

$$d_{\mathrm{H}}\big(\mu, \mu_N^{\text{marginal}}\big) \leq C \Big\| \mathbb{E}_{\nu_N}\big[\|G_N - G\|^{2r_1}\big]^{1/r_1} \Big\|_{L^{r_2}_{\mu_0}(\mathcal{U})}^{1/2} \tag{3.12}$$

$$\mathbb{E}_{\nu_N}\Big[d_{\mathrm{H}}\big(\mu, \mu_N^{\text{sample}}\big)^2\Big]^{1/2} \leq D \Big\| \mathbb{E}_{\nu_N}\big[\|G_N - G\|^{2s_1}\big]^{1/s_1} \Big\|_{L^{s_2}_{\mu_0}(\mathcal{U})}^{1/2}. \tag{3.13}$$

For brevity and simplicity, the following result uses one pair $q, s \geq 1$ in (3.11) in order to obtain convergence statements for *both* μ_N^{marginal} and μ_N^{sample}. If one is interested in only one of these measures, then one may optimise q and s accordingly.

Theorem 3.9 (Convergence of posteriors for randomised forward models in quadratic potentials)**.** *Let* Φ *and* Φ_N *be as in* (3.8).

(a) Suppose there exists some $p^* > 1$ *with Hölder conjugate* $(p^*)'$ *such that* $\exp(\Phi) \in L^{p^*}_{\mu_0}(\mathcal{U})$, *and suppose that* (3.7) *holds for some* $C_4 \in \mathbb{R}$. *If* $G_N \to G$ *as in* (3.11) *with* $q = 2$ *and* $s = 2p^*/(p^* - 1)$, *then* (3.12) *holds with* $r_1 = 1$ *and* $r_2 = 2p^*/(p^* - 1)$, *and* (3.13) *holds with* $s_1 = 2$ *and* $s_2 = 2$.

(b) Suppose there exists some $2 < \rho^* < \infty$ *such that* $\mathbb{E}_{\nu_N}[\exp(\rho^* \Phi_N)] \in L^1_{\mu_0}$. *If* $G_N \to G$ *as in* (3.11) *with* $q = 2\rho^*/(\rho^* - 2)$ *and* $s = 2\rho^*/(\rho^* - 1)$, *then* (3.12) *holds with* $r_1 = 1$ *and* $r_2 = 2\rho^*/(\rho^* - 1)$ *and* (3.13) *holds with* $s_1 = 2\rho^*/(\rho^* - 2)$ *and* $s_2 = 2$.

In both cases, μ_N^{marginal} *and* μ_N^{sample} *converge to* μ *in the appropriate metrics given in* (3.12) *and* (3.13) *respectively.*

4 Application: randomised misfit models

This section considers a particular Monte Carlo approximation Φ_N of a quadratic potential Φ, proposed by Nemirovski et al. (2008) and Shapiro et al. (2009) and further applied and analysed in the MAP estimator context by Le et al. (2017). This approximation is particularly useful when the data $y \in \mathbb{R}^J$ is very high-dimensional, and is derived in the following way:

$$\begin{aligned}
\Phi(u) &:= \frac{1}{2}\Big\|\Gamma^{-1/2}(y - G(u))\Big\|^2 \\
&= \frac{1}{2}\big(\Gamma^{-1/2}(y - G(u))\big)^{\mathsf{T}} \mathbb{E}[\sigma\sigma^{\mathsf{T}}] \big(\Gamma^{-1/2}(y - G(u))\big) && \text{where } \mathbb{E}[\sigma] = 0 \in \mathbb{R}^J,\ \mathbb{E}[\sigma\sigma^{\mathsf{T}}] = I_{J\times J} \\
&= \frac{1}{2}\mathbb{E}\Big[\big|\sigma^{\mathsf{T}}\big(\Gamma^{-1/2}(y - G(u))\big)\big|^2\Big] \\
&\approx \frac{1}{2N}\sum_{i=1}^{N}\big|\sigma^{(i)\mathsf{T}}\big(\Gamma^{-1/2}(y - G(u))\big)\big|^2 && \text{for i.i.d. } \sigma^{(1)}, \dots, \sigma^{(N)} \overset{\mathrm{d}}{=} \sigma \\
&= \frac{1}{2}\Big\|\Sigma_N^{\mathsf{T}}\big(\Gamma^{-1/2}(y - G(u)\big)\Big\|^2 && \text{for } \Sigma_N := \frac{1}{\sqrt{N}}[\sigma^{(1)} \cdots \sigma^{(N)}] \in \mathbb{R}^{J\times N} \\
&=: \Phi_N(u).
\end{aligned}$$

The analysis and numerical studies in Le et al. (2017, Sections 3–4) suggest that a good choice for the random vector σ would be one with independent and identically distributed (i.i.d.) entries from a sub-Gaussian probability distribution on $\mathbb{R}$. Examples of sub-Gaussian distributions considered include

(a) the standard Gaussian distribution: $\sigma_j \sim \mathcal{N}(0, 1)$, for $j = 1, \dots, J$; and

(b) the ℓ-sparse distribution: for $\ell \in [0, 1)$, let $s := \frac{1}{1-\ell} \geq 1$ and set, for $j = 1, \dots, J$,

$$\sigma_j := \sqrt{s}\begin{cases} 1, & \text{with probability } \frac{1}{2s}, \\ 0, & \text{with probability } \ell = 1 - \frac{1}{s}, \\ -1, & \text{with probability } \frac{1}{2s}. \end{cases}$$

Le et al. (2017) observe that, for large J and moderate $N \approx 10$, the random potential Φ_N and the original potential Φ are very similar, in particular having approximately the same minimisers and minimum values. Statistically, these correspond to the maximum likelihood estimators under Φ and Φ_N

being very similar; after weighting by a prior, this corresponds to similarity of maximum a posteriori (MAP) estimators.

Here, we take a fully Bayesian perpective, and thus the corresponding conjecture is that the deterministic posterior $\mathrm{d}\mu(u) \propto \exp(-\Phi(u))\,\mathrm{d}\mu_0(u)$ is well approximated by the random posterior $\mathrm{d}\mu_N^{\mathrm{sample}}(u) \propto \exp(-\Phi_N(u))\,\mathrm{d}\mu_0(u)$. Indeed, via Theorem 3.2, we have the following convergence result for the case of a sparsifying distribution:

Proposition 4.1. *Suppose that the entries of σ are i.i.d. ℓ-sparse, for some $\ell \in [0, 1)$, and that $\Phi \in L^2_{\mu_0}(\mathcal{U})$. Then there exists a constant C, independent of N, such that*

$$\Big(\mathbb{E}_\sigma\big[d_{\mathrm{H}}\big(\mu, \mu_N^{\mathrm{sample}}\big)^2\big]\Big)^{1/2} \leq \frac{C}{\sqrt{N}}. \tag{4.1}$$

As the proof reveals, a valid choice of the constant C in (4.1) is

$$C = \sqrt{J^3 \mathbb{E}_\sigma[\sigma_j^4] - 1}\,\|\Phi\|_{L^2_{\mu_0}(\mathcal{U})} = \sqrt{J^3 s^3 - 1}\,\|\Phi\|_{L^2_{\mu_0}(\mathcal{U})}. \tag{4.2}$$

Thus, as one would expect, the accuracy of the approximation decreases as σ approaches the complete sparsification case $\ell = 1$ or as the data dimension J increases, but always with the same convergence rate $N^{-1/2}$ in terms of the approximation dimension N.

Remark 4.2. Extending the conclusion of Proposition 4.1 to the case of i.i.d. Gaussian random variables $\sigma_j \sim \mathcal{N}(0, 1)$ appears to be problematic. In the proof, we crucially made use of the bound $|\sigma_j| \leq \sqrt{s}$ to verify Assumption (b) of Theorem 3.2. For Gaussian random variables, we would similarly have to bound the exponential moments of

$$\max_{\substack{1 \leq i \leq N \\ 1 \leq j \leq J}} \sigma_j^{(i)},$$

independently of $N \in \mathbb{N}$, which is not possible.

5 Application: probabilistic integration of dynamical systems

The data-based inference of initial conditions or governing parameters for dynamical problems arises frequently in scientific applications, a prime example being data assimilation in numerical weather prediction (Law et al., 2015; Reich and Cotter, 2015). In this setting, the Bayesian likelihood involves a solution of the mathematical model for the dynamics, which is typically an ODE or time-dependent PDE; we focus here on the ODE situation. Even when the governing ODE is deterministic, it may be profitable to perform a probabilistic numerical solution: possible motivations for doing so include the representation of model error (model inadequacy) in the ODE itself, and the impact of discretisation uncertainty. When such a probabilistic solver is used for the ODE, the likelihood becomes random in the sense considered in this paper.

Random approximate solution of deterministic ODEs is an old idea (Diaconis, 1988; Skilling, 1992) that has received renewed attention in recent years (Schober et al., 2014; Conrad et al., 2016; Hennig et al., 2015; Lie et al., 2017). As random forward models, these probabilistic ODE solvers are amenable to the analysis of Section 3. Let $f \colon \mathbb{R}^d \to \mathbb{R}^d$ and consider the following parameter-dependent initial value problem for a fixed, parameter-independent duration $T > 0$:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} z(t; u) &= f(z(t; u); u), && \text{for } 0 \leq t \leq T, \\ z(0; u) &= z_0(u). \end{aligned} \tag{5.1}$$

In the context of the Bayesian inverse problem presented in Section 2, the unknown parameter u will appear in the definition of the initial condition $z_0 = z_0(u)$ or the right-hand side $f(z(t)) = f(z(t); u)$, resulting in the parameter-dependent solution $(z(t; u))_{t \in [0,T]}$. Define the solution operator

$$S \colon \mathcal{U} \to C([0, T]; \mathbb{R}^d), \quad u \mapsto S(u) := (z(t; u))_{t \in [0,T]}, \tag{5.2}$$

where $(z(t;u))_{t\in[0,T]}$ solves (5.1). We equip $C([0,T];\mathbb{R}^d)$ with the supremum norm.

For notational convenience, we will for the majority of this section not indicate the dependence of z_0 or f on u. We will, however, explicitly track the dependence on z_0 and f of the error analysis below.

Denote by $\Phi^t\colon \mathbb{R}^d \to \mathbb{R}^d$ the flow map associated to the initial value problem (5.1), i.e. $\Phi^t(z_0) \coloneqq z(t;u) = S(u)(t)$. Fix a constant time step $\tau > 0$ such that $N \coloneqq T/\tau \in \mathbb{N}$, and a time grid

$$t_k \coloneqq k\tau \text{ for } k \in [N] \coloneqq \{0, 1, \dots, N\}. \tag{5.3}$$

we shall denote by $z_k \coloneqq z(t_k) \equiv \Phi^\tau(z_{k-1})$ the value of the exact solution to (5.1) at time t_k. We shall sometimes abuse notation and write $[N] = \{0, 1, \dots, N-1\}$ or $[N] = \{1, 2, \dots, N\}$.

To a single-step numerical integration method (e.g. a Runge–Kutta method of some order) we shall associate a numerical flow map $\Psi^\tau\colon \mathbb{R}^d \to \mathbb{R}^d$. The numerical flow map approximates the sequence $(z_k)_{k\in[N]}$ by a sequence $(Z'_k)_{k\in[N]}$, where $Z'_k \coloneqq \Psi^\tau(Z'_{k-1})$. A fundamental task in numerical analysis is to determine sufficient conditions for convergence of the sequence $(Z'_k)_{k\in[N]}$ to $(z_k)_{k\in[N]}$. The investigations of Conrad et al. (2016) and Lie et al. (2017) concern a similar task in the context of uncertainty quantification. Given $\tau > 0$, consider a collection $(\xi_k)_{k\in[N]}$ of stochastic processes $\xi_k\colon \Omega \times [0,\tau] \to \mathbb{R}^d$ having almost-surely continuous paths. Define a stochastic process $(Z_t)_{t\in[0,T]}$ in terms of a new randomised integrator

$$Z(t_{k+1};u) \coloneqq \Psi^\tau(Z(t_k;u)) + \xi_k(\tau). \tag{5.4}$$

The stochastic processes $(\xi_k)_{k\in[N]}$ are intended to capture the effect of uncertainties, e.g. those that arise due to properties of the vector field that are not resolved at the resolution given by the time step τ. We extend the definition (5.4) to continuous time via

$$Z(t;u) \coloneqq \Psi^{t-t_k}(Z(t_k;u)) + \xi_k(t-t_k), \quad \text{for } t_k < t < t_{k+1}. \tag{5.5}$$

We shall use the $(\xi_k)_{k\in[N]}$ to construct our random approximations to Φ. Note therefore that, in order to be consistent with our assumption (see the third paragraph of Section 3) that the randomness in the approximation of Φ is independent of the randomness in the parameter u being inferred, we shall assume that the $(\xi_k)_{k\in[N]}$ do not depend on the parameter u. However, the map Ψ^τ does depend on the parameter $u \in \mathcal{U}$, because Ψ^τ involves the vector field $f(\,\cdot\,;u)$.

Define the random solution operator associated to the randomised numerical integrator (5.5):

$$S_N\colon \mathcal{U} \to C([0,T];\mathbb{R}^d), \quad u \mapsto S_N(u) \coloneqq (Z(t;u))_{t\in[0,T]}, \tag{5.6}$$

where $(Z(t;u))_{t\in[0,T]}$ satisfies (5.5), and is almost surely continuous.

Let $T_J \subset [0,T]$ be a strictly increasing sequence of time points, indexed by a finite, nonempty index set J with cardinality $|J| \in \mathbb{N}$. Define $\mathcal{Y} = \mathbb{R}^{d|J|}$, and equip it with the topology induced by the standard Euclidean inner product. Define the observation operator

$$O\colon C([0,T];\mathbb{R}^d) \to \mathcal{Y}, \quad \tilde{z} \mapsto O(\tilde{z}) \coloneqq (\tilde{z}(t_j))_{t_j\in T_J}, \tag{5.7}$$

which projects some $\tilde{z} \in C([0,T];\mathbb{R}^d)$ to a finite-dimensional vector in $\mathcal{Y}$ constructed by stacking the $\mathbb{R}^d$-valued vectors that result from evaluating $\tilde{z}$ at the time points in T_J. We take $\|\cdot\|_{\mathcal{Y}} = \|\cdot\|_{\ell_2^{d|J|}}$.

Given the operators S, O, and S_N defined in (5.2), (5.7), and (5.6), we define the forward operators $G, G_N\colon \mathcal{U} \to \mathcal{Y}$ by

$$G \coloneqq O \circ S, \quad G_N \coloneqq O \circ S_N, \tag{5.8}$$

and the associated likelihoods are the quadratic misfits given by (3.8) with some fixed, positive-definite matrix Γ.

We define the continuous-time error process by

$$e(t;u) \coloneqq z(t;u) - Z(t;u), \quad 0 \le t \le T. \tag{5.9}$$

Since T_J is a proper subset of $[0,T]$, it follows that

$$\|G_N(u) - G(u)\| \equiv \|G_N(u) - G(u)\|_{\mathcal{Y}} \le |J| \sup_{0\le t\le T} \|e(t;u)\|_{\ell_2^d}. \tag{5.10}$$

This completes our formulation of the probabilistic numerical integration of the ODE (5.1) as a random likelihood model of the type considered in Section 3.

5.1 Convergence in continuous time for Lipschitz flows

In this section, we quote some assumptions and results from Lie et al. (2017). The vector field f in (5.1) induces a flow map $\Phi^\tau : \mathbb{R}^d \to \mathbb{R}^d$ by

$$\Phi^\tau(a) = a + \int_0^\tau f(\Phi^t(a))\,\mathrm{d}t. \tag{5.11}$$

Assumption 5.1 (Assumption 3.1, Lie et al. (2017))**.** The vector field f admits $0 < \tau^* \leq 1$ and $C_\Phi \geq 1$, such that for $0 < \tau < \tau^*$, the flow map $\Phi^\tau : \mathbb{R}^d \to \mathbb{R}^d$ defined by (5.11) is globally Lipschitz, with

$$\|\Phi^\tau(z_0) - \Phi^\tau(v_0)\| \leq (1 + C_\Phi \tau)\|z_0 - v_0\|, \quad \text{for all } z_0, v_0 \in \mathbb{R}^d.$$

Recall that $\Psi^\tau : \mathbb{R}^d \to \mathbb{R}^d$ represents the numerical method that we use to integrate (5.1).

Assumption 5.2 (Assumption 3.2, Lie et al. (2017))**.** The numerical method Ψ^τ has uniform local truncation error of order $q + 1$: for some constant $C_\Psi \geq 1$ that does not depend on τ,

$$\sup_{u \in \mathbb{R}^d} \|\Psi^\tau(u) - \Phi^\tau(u)\| \leq C_\Psi \tau^{q+1}.$$

Now recall the collection $(\xi_k(\tau))_{k \in [N]}$ of random variables, where $\xi_k(\tau)$ is used in (5.4).

Assumption 5.3 (Assumption 5.1, Lie et al. (2017))**.** The stochastic processes $(\xi_k)_{k \in \mathbb{N}}$ admit $p \geq 1$, $R \in \mathbb{N} \cup \{+\infty\}$, and $C_{\xi,R} \geq 1$, independent of k and τ, such that for all $1 \leq r \leq R$ and all $k \in \mathbb{N}$,

$$\mathbb{E}_{\nu_N}\left[\sup_{0 < t \leq T/N} \|\xi_k(t)\|^r\right] \leq \left(C_{\xi,R}\left(\frac{T}{N}\right)^{p+1/2}\right)^r.$$

The assumption above quantifies the regularity of the $(\xi_k)_{k \in [N]}$ by specifying how many moments each $\xi_k(t)$ has and how quickly these decay with $\tau = T/N$. We do not require the $(\xi_k(t))_{k \in [N]}$ to have zero mean, to be independent, or to be identically distributed.

We have the following convergence theorem:

Theorem 5.4 (Theorem 5.2, Lie et al. (2017))**.** *Let $n \in \mathbb{N}$, suppose that $e_0 = 0$, and suppose that Assumptions 5.1, 5.2, and 5.3 hold with parameters τ^*, C_Φ, C_Ψ, q, $C_{\xi,R}$, p, and R. Then, for all $T/\tau^* < N$,*

$$\mathbb{E}_{\nu_N}\left[\sup_{0 \leq t \leq T} \|e(t;u)\|^n\right] \leq 3^{n-1}\big((1 + C_\Phi \tau^*)^n \overline{C} + C_\Psi^n (\tau^*)^n + T C_{\xi,R}^n\big)\left(\frac{T}{N}\right)^{n(q \wedge (p-1/2))}, \tag{5.12}$$

where

$$\begin{aligned}\overline{C} &\coloneqq 2T \max\{(4C_\Psi)^n, (2C_{\xi,R})^n\} \exp(T C_\Phi(n, \tau^*)) \\ C_\Phi(n, \tau^*) &\coloneqq \big[(1 + \tau^* 2^{n-1})^2 (1 + \tau^* C_\Phi)^n - 1\big](\tau^*)^{-1}.\end{aligned}$$

Note that the scalars $\overline{C}$ and $C_\Phi(n, \tau^*)$ depend on $u \in \mathcal{U}$, since C_Φ and C_Ψ depend on the vector field f, which in turn depends on the parameter u.

Corollary 5.5. *Fix $n \in \mathbb{N}$. Suppose that Assumptions 5.1 and 5.2 hold, and that 5.3 holds with $R = +\infty$ and $p \geq 1/2$. Then, for all $0 < \tau < \tau^*$,*

$$\mathbb{E}_{\nu_N}\left[\exp\left(\rho \sup_{0 \leq t \leq T} \|e(t)\|^n\right)\right] < \infty, \quad \textit{for all } \rho \in \mathbb{R}. \tag{5.13}$$

5.2 Effect of probabilistic integration on Bayesian posterior distribution

Define the approximate posteriors μ_N^{marginal} and μ_N^{sample} according to (3.2) and (3.1), using the quadratic misfits Φ and Φ_N from (3.8) and the forward models G and G_N given in (5.2) and (5.6) respectively.

Theorem 5.6. *Suppose that $\mathcal{U}$ is a compact subset of $\mathbb{R}^m$ for some $m \in \mathbb{N}$, and suppose that $S, S_N : \mathcal{U} \to C([0,T];\mathbb{R}^d)$ are continuous maps. Let $2 < \rho^* < \infty$ be arbitrary. Suppose that $e_0 = 0$, and that Assumptions 5.1, 5.2, and 5.3 hold with parameters τ^*, C_Φ, C_Ψ, q, $R = +\infty$, $C_{\xi,R}$, and p, and that these parameters depend continuously on u. Then, for $N \in \mathbb{N}$ such that $T/\tau^* < N$,* (3.12) *holds for $r_1 = 1$ and $r_2 = 2\rho^*/(\rho^* - 1)$, and* (3.13) *holds for $s_1 = 2\rho^*/(\rho^* - 2)$ and $s_2 = 2$.*

Note that the continuous dependence on u of the parameters of Assumptions 5.1, 5.2 and 5.3 also allows for parameters that do not depend on u, e.g. $R = +\infty$.

Acknowledgements

ALT is partially supported by The Alan Turing Institute under the EPSRC grant EP/N510129/1. HCL and TJS are partially supported by the Freie Universität Berlin within the Excellence Initiative of the German Research Foundation (DFG). This work was partially supported by the DFG through grant CRC 1114 "Scaling Cascades in Complex Systems", and by the National Science Foundation (NSF) under grant DMS-1127914 to the Statistical and Applied Mathematical Sciences Institute (SAMSI) and SAMSI's QMC Working Group II "Probabilistic Numerics". Any opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the above-named funding agencies and institutions.

References

C. Andrieu and G. O. Roberts. The pseudo-marginal approach for efficient Monte Carlo computations. *Ann. Statist.*, 37(2):697–725, 2009. doi:10.1214/07-AOS574.

M. A. Beaumont. Estimation of population growth or decline in genetically monitored populations. *Genetics*, 164(3):1139–1160, 2003. URL http://www.genetics.org/content/164/3/1139.

V. I. Bogachev. *Measure Theory. Vol. I, II.* Springer-Verlag, Berlin, 2007. doi:10.1007/978-3-540-34514-5.

P. R. Conrad, M. Girolami, S. Särkkä, A. M. Stuart, and K. C. Zygalakis. Statistical analysis of differential equations: introducing probability measures on numerical solutions. *Stat. Comput.*, 2016. doi:10.1007/s11222-016-9671-0.

S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. *Statist. Sci.*, 28(3):424–446, 2013. doi:10.1214/13-STS421.

M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. In R. Ghanem, D. Higdon, and H. Owhadi, editors, *Handbook of Uncertainty Quantification*, pages 311–428. 2016. doi:10.1007/978-3-319-11259-6_7-1.

M. Dashti, S. Harris, and A. M. Stuart. Besov priors for Bayesian inverse problems. *Inverse Probl. Imaging*, 6(2):183–200, 2012. doi:10.3934/ipi.2012.6.183.

P. Diaconis. Bayesian numerical analysis. In *Statistical Decision Theory and Related Topics, IV, Vol. 1 (West Lafayette, Ind., 1986)*, pages 163–175. Springer, New York, 1988.

P. Dupuis and R. S. Ellis. *A Weak Convergence Approach to the Theory of Large Deviations.* Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, 1997. doi:10.1002/9781118165904.

S. N. Evans and P. B. Stark. Inverse problems as statistics. *Inverse Probl.*, 18(4):R55–R97, 2002. doi:10.1088/0266-5611/18/4/201.

P. Hennig, M. A. Osborne, and M. Girolami. Probabilistic numerics and uncertainty in computations. *Proc. A.*, 471(2179):20150142, 17, 2015. doi:10.1098/rspa.2015.0142.

B. Hosseini. Well-posed Bayesian inverse problems with infinitely-divisible and heavy-tailed prior measures. *SIAM/ASA J. Uncertain. Quantif.*, 5(1):1024–1060, 2017. doi:10.1137/16M1096372.

B. Hosseini and N. Nigam. Well-posed Bayesian inverse problems: priors with exponential tails. *SIAM/ASA J. Uncertain. Quantif.*, 5(1):436–465, 2017. doi:10.1137/16M1076824.

R. Jordan and D. Kinderlehrer. An extended variational principle. In *Partial Differential Equations and Applications*, volume 177 of *Lecture Notes in Pure and Appl. Math.*, pages 187–200. Dekker, New York, 1996. doi:10.5006/1.3292113.

J. Kaipio and E. Somersalo. *Statistical and Computational Inverse Problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005. doi:10.1007/b138659.

C. H. Kraft. Some conditions for consistency and uniform consistency of statistical procedures. *Univ. California Publ. Statist.*, 2:125–141, 1955.

M. Lassas and S. Siltanen. Can one use total variation prior for edge-preserving Bayesian inversion? *Inverse Probl.*, 20(5):1537–1563, 2004. doi:10.1088/0266-5611/20/5/013.

K. Law, A. Stuart, and K. Zygalakis. *Data Assimilation: A Mathematical Introduction*, volume 62 of *Texts in Applied Mathematics*. Springer, 2015. doi:10.1007/978-3-319-20325-6.

E. B. Le, A. Myers, T. Bui-Thanh, and Q. P. Nguyen. A data-scalable randomized misfit approach for solving large-scale PDE-constrained inverse problems. *Inverse Probl.*, 33(6):065003, 2017. doi:10.1088/1361-6420/aa6cbd.

H. C. Lie, A. M. Stuart, and T. J. Sullivan. Strong convergence rates of probabilistic integrators for ordinary differential equations, 2017. arXiv:1703.03680.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM J. Optim.*, 19(4):1574–1609, 2008. doi:10.1137/070704277.

S. Ohta and A. Takatsu. Displacement convexity of generalized relative entropies. *Adv. Math.*, 228(3): 1742–1787, 2011. doi:10.1016/j.aim.2011.06.029.

H. Owhadi and C. Scovel. Qualitative robustness in Bayesian inference. *ESAIM Probab. Stat.*, 21: 251–274, 2017. doi:10.1051/ps/2017014.

H. Owhadi, C. Scovel, and T. J. Sullivan. Brittleness of Bayesian inference under finite information in a continuous world. *Electron. J. Stat.*, 9(1):1–79, 2015. doi:10.1214/15-EJS989.

M. S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, Inc., San Francisco, Calif.-London-Amsterdam, 1964.

S. Reich and C. Cotter. *Probabilistic Forecasting and Bayesian Data Assimilation*. Cambridge University Press, New York, 2015. doi:10.1017/CBO9781107706804.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer Texts in Statistics. Springer-Verlag, New York, 1999. doi:10.1007/978-1-4757-3071-5.

M. Schober, D. K. Duvenaud, and P. Hennig. Probabilistic ODE solvers with Runge–Kutta means. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 739–747. Curran Associates, Inc., 2014. http://papers.nips.cc/paper/5451-probabilistic-ode-solvers-with-runge-kutta-means.

A. Shapiro, D. Dentcheva, and A. Ruszczyński. *Lectures on Stochastic Programming: Modeling and Theory*, volume 9 of *MPS/SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Programming Society (MPS), Philadelphia, PA, 2009. doi:10.1137/1.9780898718751.

J. Skilling. Bayesian solution of ordinary differential equations. In C. R. Smith, G. J. Erickson, and P. O. Neudorfer, editors, *Maximum Entropy and Bayesian Methods*, volume 50 of *Fundamental Theories of Physics*, pages 23–37. Springer, 1992. doi:10.1007/978-94-017-2219-3.

A. M. Stuart. Inverse problems: a Bayesian perspective. *Acta Numer.*, 19:451–559, 2010. doi:10.1017/S0962492910000061.

A. M. Stuart and A. L. Teckentrup. Posterior consistency for Gaussian process approximations of Bayesian posterior distributions. *Math. Comput.*, 2017. doi:10.1090/mcom/3244.

T. J. Sullivan. Well-posed Bayesian inverse problems and heavy-tailed stable quasi-Banach space priors. *Inverse Probl. Imaging*, 11(5):857–874, 2017. doi:10.3934/ipi.2017040.

Appendix: Proofs of Results

The proofs in this section will make frequent use of the following inequalities for real a and b:

$$(a-b)^2 = \left(\frac{a^2-b^2}{a+b}\right)^2 \le \frac{(a^2-b^2)^2}{a^2+b^2}, \tag{A.1}$$

$$|\exp(a)-\exp(b)| \le (\exp(a)+\exp(b))|a-b|. \tag{A.2}$$

We also have, for arbitrary $N \in \mathbb{N}$ and $p \ge 1$ (not necessarily integer-valued), by the triangle inequality and Jensen's inequality,

$$\left|\sum_{j=1}^{N} s_j\right|^p \le N^p \left(\frac{1}{N}\sum_{j=1}^{N}|s_j|\right)^p \le N^{p-1}\sum_{j=1}^{N}|s_j|^p, \tag{A.3}$$

Proof of Theorem 3.1. Using (2.4) and (3.1), we have

$$\begin{aligned}
\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}} - \sqrt{\frac{\mathrm{d}\mu_N^{\mathrm{marginal}}}{\mathrm{d}\mu_0}} &= \frac{\sqrt{\exp(-\Phi(u))}}{Z^{1/2}} - \frac{\sqrt{\mathbb{E}_{\nu_N}\left[\exp(-\Phi_N(u))\right]}}{\mathbb{E}_{\nu_N}\left[Z_N^{\mathrm{sample}}\right]^{1/2}} \\
&= \frac{\sqrt{\exp(-\Phi(u))} - \sqrt{\mathbb{E}_{\nu_N}\left[\exp(-\Phi_N(u))\right]}}{Z^{1/2}} \\
&\quad + \sqrt{\mathbb{E}_{\nu_N}\left[\exp(-\Phi_N(u))\right]}\left(\frac{1}{Z^{1/2}} - \frac{1}{\mathbb{E}_{\nu_N}\left[Z_N^{\mathrm{sample}}\right]^{1/2}}\right).
\end{aligned}$$

Using the inequality $(a-b)^2 \le 2(a^2+b^2)$ with $a = Z^{-1/2}\big(e^{-\Phi(u)/2} - \mathbb{E}_{\nu_N}[e^{-\Phi_N(u)}]^{1/2}\big)$ and $b = \mathbb{E}_{\nu_N}[Z_N^{\mathrm{sample}}]^{1/2}(Z^{-1/2} - \mathbb{E}_{\nu_N}[Z_N^{\mathrm{sample}}]^{-1/2})$, and using the definition (2.1) of the Hellinger distance d_{H}, we have

$$\begin{aligned}
2\, d_{\mathrm{H}}\big(\mu, \mu_N^{\mathrm{marginal}}\big)^2 &= \int_{\mathcal{U}} \left(\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}}(u) - \sqrt{\frac{\mathrm{d}\mu_N^{\mathrm{marginal}}}{\mathrm{d}\mu_0}}(u)\right)^2 \mathrm{d}\mu_0(u) \\
&\le \frac{2}{Z}\left\|\left(\sqrt{\exp(-\Phi)} - \sqrt{\mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]}\right)^2\right\|_{L^1_{\mu_0}} \\
&\quad + 2\mathbb{E}_{\nu_N}\left[Z_N^{\mathrm{sample}}\right]\left(Z^{-1/2} - \mathbb{E}_{\nu_N}\left[Z_N^{\mathrm{sample}}\right]^{-1/2}\right)^2 =: I + II.
\end{aligned}$$

For the first term, we use inequality (A.1) with $a = e^{-\Phi(u)/2}$ and $b = \mathbb{E}_{\nu_N}[\exp(-\Phi_N(u))]^{1/2}$, together with Hölder's inequality with conjugate exponents p_1 and p_1', to derive

$$\begin{aligned}
\frac{Z}{2} I &\le \left\|\big(\exp(-\Phi) - \mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]\big)^2\big(\exp(-\Phi) + \mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]\big)^{-1}\right\|_{L^1_{\mu_0}} \\
&\le \left\|\big(\exp(-\Phi) - \mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]\big)^2\right\|_{L^{p_1'}_{\mu_0}} \left\|\big(\exp(-\Phi) + \mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]\big)^{-1}\right\|_{L^{p_1}_{\mu_0}}.
\end{aligned} \tag{A.4}$$

We estimate the second factor on the right-hand side of (A.4). Using the facts that $1/x$ is decreasing on $(0,\infty)$, that $(x+y)^{-1} \le \min\{x^{-1}, y^{-1}\}$ for all $x, y > 0$, and that both $\exp(-\Phi(u))$ and $\mathbb{E}_{\nu_N}[\exp(-\Phi_N(u))]$ are strictly positive, we obtain

$$\left\|\big(\exp(-\Phi) + \mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]\big)^{-1}\right\|_{L^{p_1}_{\mu_0}} \le \left\|\min\Big\{\exp(\Phi), \mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]^{-1}\Big\}\right\|_{L^{p_1}_{\mu_0}}.$$

For $f, g \in L^1_{\mu_0}(\mathcal{U};\mathbb{R})$, the decomposition of $\mathcal{U} = \{f < g\} \uplus \{f \ge g\}$, and the corresponding integral inequalities on $\{f < g\}$ and $\{f \ge g\}$, imply that $\|\min\{f, g\}\|_{L^1_{\mu_0}} \le \min\{\|f\|_{L^1_{\mu_0}}, \|g\|_{L^1_{\mu_0}}\}$. Hence,

$$\left\|\min\Big\{e^{-\Phi}, \mathbb{E}_{\nu_N}\left[e^{-\Phi_N}\right]^{-1}\Big\}\right\|_{L^{p_1}_{\mu_0}} \le \min\left\{\|e^{\Phi}\|_{L^{p_1}_{\mu_0}}, \left\|\mathbb{E}_{\nu_N}\left[e^{-\Phi_N}\right]^{-1}\right\|_{L^{p_1}_{\mu_0}}\right\} \le C_1, \tag{A.5}$$

where $C_1 = C_1(p_1)$ is the constant specified in assumption (a). This completes our estimate for the second factor on the right-hand side of (A.4). For the first factor, the linearity of expectation, inequality (A.2), and Hölder's inequality with conjugate exponents p_2, p_2' with respect to ν_N and p_3, p_3' with respect to μ_0 give

$$\begin{aligned}
&\left\|\big(\exp(-\Phi) - \mathbb{E}_{\nu_N}\left[\exp(-\Phi_N)\right]\big)^2\right\|_{L^{p_1'}_{\mu_0}} = \left\|\mathbb{E}_{\nu_N}\left[\exp(-\Phi) - \exp(-\Phi_N)\right]^2\right\|_{L^{p_1'}_{\mu_0}} \\
&\quad\le \left\|\mathbb{E}_{\nu_N}\left[|\exp(-\Phi) + \exp(-\Phi_N)||\Phi - \Phi_N|\right]^2\right\|_{L^{p_1'}_{\mu_0}} \\
&\quad\le \left\|\mathbb{E}_{\nu_N}\left[\big(\exp(-\Phi) + \exp(-\Phi_N)\big)^{p_2}\right]^{2/p_2}\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{p_2'}\right]^{2/p_2'}\right\|_{L^{p_1'}_{\mu_0}} \\
&\quad\le \left\|\mathbb{E}_{\nu_N}\left[\big(\exp(-\Phi) + \exp(-\Phi_N)\big)^{p_2}\right]^{1/p_2}\right\|^2_{L^{2p_1'p_3}_{\mu_0}} \left\|\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{p_2'}\right]^{1/p_2'}\right\|^2_{L^{2p_1'p_3'}_{\mu_0}}
\end{aligned} \tag{A.6}$$

Letting $C_2 = C_2(p_1', p_2, p_3)$ be the constant in assumption (b), and using (A.5), it then follows that

$$I \le \frac{2}{Z} \cdot C_1(p_1) \cdot C_2^2(p_1', p_2, p_3) \cdot \left\|\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{p_2'}\right]^{1/p_2'}\right\|^2_{L^{2p_1'p_3'}_{\mu_0}}.$$

Now apply inequality (A.1) with $a = \mathbb{E}_{\nu_N}[Z_N^{\text{sample}}]^{-1/2}$ and $b = Z^{-1/2}$, and the inequality $[(x + y)xy]^{-1} \le \max\{x^{-3}, y^{-3}\}$ for $x, y > 0$, to obtain

$$\begin{aligned}
\frac{1}{2\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]} II &= \left(Z^{-1/2} - \left(\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]\right)^{-1/2}\right)^2 \\
&= \left(\frac{\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right] - Z}{Z\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]}\right)^2 \frac{Z\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]}{Z + \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]} \\
&\le \left(\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right] - Z\right)^2 \max\left\{Z^{-3}, \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]^{-3}\right\} \\
&\le \left(\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right] - Z\right)^2 \max\left\{Z^{-3}, C_3^{-3}\right\},
\end{aligned}$$

where the last inequality follows from assumption (c).

Using Tonelli's theorem, Jensen's inequality, inequality (A.2), and Hölder's inequality with the same conjugate exponent pairs that we used to obtain (A.6),

$$\begin{aligned}
&\left(\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right] - Z\right)^2 \\
&\quad= \mathbb{E}_{\mu_0}\left[\mathbb{E}_{\nu_N}\left[\exp(-\Phi_N) - \exp(-\Phi)\right]\right]^{2p_1'/p_1'} \\
&\quad\le \left\|\mathbb{E}_{\nu_N}\left[\exp(-\Phi) - \exp(-\Phi_N)\right]^2\right\|_{L^{p_1'}_{\mu_0}} \\
&\quad\le \left\|\mathbb{E}_{\nu_N}\left[\big(\exp(-\Phi) + \exp(-\Phi_N)\big)^{p_2}\right]^{1/p_2}\right\|^2_{L^{2p_1'p_3}_{\mu_0}} \left\|\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{p_2'}\right]^{1/p_2'}\right\|^2_{L^{2p_1'p_3'}_{\mu_0}} \\
&\quad\le C_2^2(p_1, p_2, p_3)\left\|\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{p_2'}\right]^{1/p_2'}\right\|^2_{L^{2p_1'p_3'}_{\mu_0}},
\end{aligned}$$

where assumption (b) yields the last inequality. Combining the estimates for I and II yields (3.3). ☐

Proof of Theorem 3.2. This proof is similar to the proof of Theorem 3.1. Since

$$\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}} - \sqrt{\frac{\mathrm{d}\mu_N^{\text{sample}}}{\mathrm{d}\mu_0}} = \frac{e^{-\Phi(u)/2} - e^{-\Phi_N(u)/2}}{Z^{1/2}} - e^{-\Phi_N(u)/2}\left(\frac{1}{\sqrt{Z_N^{\text{sample}}}} - \frac{1}{Z^{1/2}}\right),$$

Tonelli's theorem, the inequality $(a-b)^2 \le 2(a^2+b^2)$ with suitable a and b, and Jensen's inequality yield

$$
\begin{aligned}
\mathbb{E}_{\nu_N}\left[d_{\mathrm{H}}\big(\mu,\mu_N^{\text{sample}}\big)^2\right] &= \frac{1}{2}\left\|\mathbb{E}_{\nu_N}\left[\left(\sqrt{\frac{\mathrm{d}\mu}{\mathrm{d}\mu_0}}-\sqrt{\frac{\mathrm{d}\mu_N^{\text{sample}}}{\mathrm{d}\mu_0}}\right)^2\right]\right\|_{L^1_{\mu_0}} \\
&\le \frac{1}{Z}\left\|\mathbb{E}_{\nu_N}\left[\Big(\sqrt{\exp(-\Phi)}-\sqrt{\exp(-\Phi_N)}\Big)^2\right]\right\|_{L^1_{\mu_0}} \\
&\quad + \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\big(Z^{-1/2}-\big(Z_N^{\text{sample}}\big)^{-1/2}\big)^2\right] \\
&=: I + II.
\end{aligned}
$$

For the first term I, inequality (A.2), and Hölder's inequality with conjugate exponent pairs (q_1,q_1') and (q_2,q_2') give

$$
\begin{aligned}
ZI &= \left\|\mathbb{E}_{\nu_N}\left[\Big(\sqrt{\exp(-\Phi)}-\sqrt{\exp(-\Phi_N)}\Big)^2\right]\right\|_{L^1_{\mu_0}} \\
&\le \frac{1}{4}\left\|\mathbb{E}_{\nu_N}\left[|\exp(-\Phi/2)+\exp(-\Phi_N/2)|^2|\Phi-\Phi_N|^2\right]\right\|_{L^1_{\mu_0}} \\
&\le \left\|\mathbb{E}_{\nu_N}\left[|\exp(-\Phi/2)+\exp(-\Phi_N/2)|^{2q_1}\right]^{1/q_1}\mathbb{E}_{\nu_N}\left[|\Phi-\Phi_N|^{2q_1'}\right]^{1/q_1'}\right\|_{L^1_{\mu_0}} \\
&\le \left\|\mathbb{E}_{\nu_N}\left[\big(\exp(-\Phi/2)+\exp(-\Phi_N/2)\big)^{2q_1}\right]^{1/q_1}\right\|_{L^{q_2}_{\mu_0}}\left\|\mathbb{E}_{\nu_N}\left[|\Phi-\Phi_N|^{2q_1'}\right]^{1/2q_1'}\right\|^2_{L^{2q_2'}_{\mu_0}}.
\end{aligned}
$$

By (a), we may bound the first factor on the right-hand side of the last inequality by $D_1(q_1,q_2)$. Now by (A.1) with $a=Z^{-1/2}$ and $b=(Z_N^{\text{sample}})^{-1/2}$, and by the inequality $[(x+y)xy]^{-1}\le\max\{x^{-3},y^{-3}\}$ for $x,y>0$, we obtain (see the proof of Theorem 3.1 after (A.6)) that

$$
II \le \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\max\big\{Z^{-3},\big(Z_N^{\text{sample}}\big)^{-3}\big\}\big(Z-Z_N^{\text{sample}}\big)^2\right].
$$

Jensen's inequality and another application of inequality (A.1) yield

$$
\big(Z-Z_N^{\text{sample}}\big)^2 \le \|\exp(-\Phi)-\exp(-\Phi_N)\|^2_{L^2_{\mu_0}} \le \left\|\big(\exp(-\Phi)+\exp(-\Phi_N)\big)^2(\Phi-\Phi_N)^2\right\|_{L^1_{\mu_0}}.
$$

Combining the preceding two estimates, using Tonelli's theorem and Hölder's inequality with the same conjugate exponent pairs (q_1,q_1') and (q_2,q_2') as used in the bound for I, and using (b), we get

$$
\begin{aligned}
II &\le \left\|\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\max\big\{Z^{-3},\big(Z_N^{\text{sample}}\big)^{-3}\big\}\big(e^{-\Phi}+e^{-\Phi_N}\big)^2(\Phi-\Phi_N)^2\right]\right\|_{L^1_{\mu_0}} \\
&\le \left\|\mathbb{E}_{\nu_N}\left[\Big(Z_N^{\text{sample}}\max\big\{Z^{-3},\big(Z_N^{\text{sample}}\big)^{-3}\big\}\big(e^{-\Phi}+e^{-\Phi_N}\big)^2\Big)^{q_1}\right]^{\frac{1}{q_1}}\mathbb{E}_{\nu_N}\left[|\Phi-\Phi_N|^{2q_1'}\right]^{\frac{1}{q_1'}}\right\|_{L^1_{\mu_0}} \\
&\le D_2(q_1,q_2)\left\|\mathbb{E}_{\nu_N}\left[|\Phi-\Phi_N|^{2q_1'}\right]^{1/2q_1'}\right\|^2_{L^{2q_2'}_{\mu_0}}.
\end{aligned}
$$

Combining the preceding estimates yields (3.4). $\square$

Proof of Lemma 3.4. Since $\exp(\Phi)\in L^{p^*}_{\mu_0}$, examination of assumption (a) of Theorem 3.1 indicates that we may set $p_1=p^*$ and $C_1:=\|\exp(\Phi)\|_{L^{p^*}_{\mu_0}}$. By (3.5), it follows that $\mathbb{E}_{\nu_N}[\exp(-\Phi)+\exp(-\Phi_N)]\le 2\exp(C_0)$; thus assumption (b) of Theorem 3.1 holds with $p_2=p_3=+\infty$ (so that $2p_1'p_3=+\infty$) and $C_2=2\exp(C_0)$. We now prove that Assumption (c) of Theorem 3.1 holds. It follows by setting $x=-\Phi$ and $y=-\Phi_N$ in inequality (A.2) that $|\exp(-\Phi)-\exp(-\Phi_N)|\le 2\exp(C_0)|\Phi-\Phi_N|$. Thus

$$
\begin{aligned}
\big|Z_N^{\text{sample}}-Z\big| &= \big|\mathbb{E}_{\mu_0}\big[\exp(-\Phi_N)-\exp(-\Phi)\big]\big| \\
&\le \mathbb{E}_{\mu_0}\big[|\exp(-\Phi_N)-\exp(-\Phi)|\big] \\
&\le 2\exp(C_0)\mathbb{E}_{\mu_0}\big[|\Phi-\Phi_N|\big]. \qquad \text{(A.7)}
\end{aligned}
$$

Using Jensen's inequality, (A.7), Tonelli's theorem, and (3.6),

$$\begin{aligned}\left|\mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]-Z\right| &\leq \mathbb{E}_{\nu_N}\left[\left|Z_N^{\text{sample}}-Z\right|\right]\\ &\leq 2\exp(C_0)\left\|\mathbb{E}_{\nu_N}\left[\left|\Phi-\Phi_N\right|\right]\right\|_{L^1_{\mu_0}}\\ &\leq \min\left\{Z-\frac{1}{C_3},C_3-Z\right\}.\end{aligned}$$

The last inequality implies that assumption (c) of Theorem 3.1 holds with the same C_3 as in (3.6), since for any $0<C_3<+\infty$ that satisfies $C_3^{-1}<Z<C_3$ and (3.6), we have

$$C_3^{-1}-Z\leq \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]-Z\leq Z-C_3^{-1}\implies C_3^{-1}\leq \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]$$

and

$$Z-C_3\leq \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]-Z\leq C_3-Z\implies \mathbb{E}_{\nu_N}\left[Z_N^{\text{sample}}\right]\leq C_3,$$

and combining both the implied statements yields assumption (c) of Theorem 3.1; thus (3.3) holds, as desired.

Now note that (3.5) implies that assumption (a) of Theorem 3.2 holds with $q_1=q_2=+\infty$ and $D_1=4\exp(C_0)$. Furthermore, (3.5) also implies that $Z_N^{\text{sample}}=\mathbb{E}_{\mu_0}[\exp(-\Phi_N)]\leq\exp(C_0)$ for all Φ_N. Thus, given that Z is ν_N-a.s. constant, and given that there exists some $0<C_3<\infty$ such that $C_3^{-1}<Z<C_3$,

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\left(Z_N^{\text{sample}}\right)^{q_1}\max\left\{Z^{-3},\left(Z_N^{\text{sample}}\right)^{-3}\right\}^{q_1}\left(\exp(-\Phi(u))+\exp(-\Phi_N(u))\right)^{2q_1}\right]^{1/q_1}\\ &\quad\leq 4\exp(3C_0)\mathbb{E}_{\nu_N}\left[\max\left\{C_3^{-3},\left(Z_N^{\text{sample}}\right)^{-3}\right\}^{q_1}\right]^{1/q_1}.\end{aligned}\tag{A.8}$$

A necessary and sufficient condition for setting $q_1=+\infty$ above (and therefore also in assumption (b) of Theorem 3.2) is that Z_N^{sample} is ν_N-almost surely bounded away from zero by a constant that does not depend on N. By Jensen's inequality applied to the function $x\mapsto\exp(x)$, and by the monotonicity of this function,

$$Z_N^{\text{sample}}=\mathbb{E}_{\mu_0}[\exp(-\Phi_N)]\geq\exp(\mathbb{E}_{\mu_0}[-\Phi_N])\geq\exp(-C_4),$$

for C_4 as in (3.7). In particular, if (3.7) holds, then so does assumption (b) of Theorem 3.2, with $q_1=q_2=+\infty$ and $D_2=4\exp(3C_0)\max\{C_3^{-3},\exp(3C_4)\}$, by inequality (A.8). □

Proof of Lemma 3.5. The proof proceeds in the same way as the proof of Lemma 3.4, with the exception that we need to prove that the assumption that $\mathbb{E}_{\nu_N}[\exp(\rho^*\Phi_N)]\in L^1_{\mu_0}$ for some $\rho^*>2$ implies that assumption (a) of Theorem 3.1 and assumption (b) of Theorem 3.2 hold with the stated parameters. Therefore, the proof will only concern these two assertions. Since $x\mapsto x^{-t}$ is strictly convex on $\mathbb{R}_{>0}$ for any $t>0$, Jensen's inequality yields that $\|\mathbb{E}_{\nu_N}[\exp(-\Phi_N)]^{-1}\|_{L^t_{\mu_0}}\leq\|\mathbb{E}_{\nu_N}[\exp(t\Phi_N)]\|_{L^1_{\mu_0}}^{1/t}$. Therefore, setting $t=\rho^*$, we find that assumption (a) of Theorem 3.1 holds, with $p_1=\rho^*$ and $C_1=\|\mathbb{E}_{\nu_N}[\exp(\rho^*\Phi_N)]\|_{L^1_{\mu_0}}^{1/\rho^*}$. Using the inequality $\max\{x,y\}\leq x+y$ for $x,y\geq 0$, we have

$$\begin{aligned}&\mathbb{E}_{\nu_N}\left[\max\left\{Z_N^{\text{sample}}Z^{-3},\left(Z_N^{\text{sample}}\right)^{-2}\right\}^{q_1}\left(\exp(-\Phi(u))+\exp(-\Phi_N(u))\right)^{2q_1}\right]^{1/q_1}\\ &\quad\leq 4\exp(2C_0)\left(C_3^{-3}\exp(C_0)+\mathbb{E}_{\nu_N}\left[\left(Z_N^{\text{sample}}\right)^{-2q_1}\right]^{1/q_1}\right),\end{aligned}$$

while Jensen's inequality, Tonelli's theorem, and the definition of the $L^1_{\mu_0}$-norm yield that

$$\begin{aligned}\mathbb{E}_{\nu_N}[(Z_N^{\text{sample}})^{-2q_1}]&\leq\mathbb{E}_{\nu_N}[\mathbb{E}_{\mu_0}[\exp(2q_1\Phi_N)]]\\ &=\mathbb{E}_{\mu_0}[\mathbb{E}_{\nu_N}[\exp(2q_1\Phi_N)]]\\ &=\|\mathbb{E}_{\nu_N}[\exp(2q_1\Phi_N)]\|_{L^1_{\mu_0}}.\end{aligned}$$

Since the last term is finite for $q_1 \le \rho^*/2$ by the hypothesis that $\mathbb{E}_{\nu_N}[\exp(\rho^* \Phi_N)] \in L^1_{\mu_0}$, it follows that assumption (b) of Theorem 3.2 holds with the parameters $q_1 = \rho^*/2$, $q_2 = +\infty$, and the scalar $D_2 = 4\exp(2C_0)(C_3^{-3}\exp(C_0) + \|\mathbb{E}_{\nu_N}[\exp(\rho^*\Phi_N)]\|_{L^1_{\mu_0}}^{2/\rho^*})$. $\square$

Proof of Proposition 3.6. Recall (3.8), and fix an arbitrary $u \in \mathcal{U}$. We have

$$\big|\Phi(u) - \Phi_N(u)\big| = \frac{1}{2}\big|\big\langle G(u) - y), \Gamma^{-1}(G(u) - y)\big\rangle - \big\langle G_N(u) - y, \Gamma^{-1}(G_N(u) - y)\big\rangle\big|.$$

Adding and subtracting $\big\langle G_N(u) - y), \Gamma^{-1}(G(u) - y)\big\rangle$ inside the absolute value, rearranging terms, applying the Cauchy–Schwarz inequality, and letting C_Γ be the largest eigenvalue of Γ^{-1} yields

$$\begin{aligned}\big|\Phi(u) - \Phi_N(u)\big| &= \frac{1}{2}\big|\big\langle \Gamma^{-1}(G(u) - y), G(u) - G_N(u)\big\rangle + \big\langle \Gamma^{-1}(G_N(u) - y), G(u) - G_N(u)\big\rangle\big| \\ &= \frac{1}{2}\big|\big\langle G(u) - y + G_N(u) - y, \Gamma^{-1}(G(u) - G_N(u))\big\rangle\big| \\ &\le C_\Gamma \|G(u) + G_N(u) - 2y\| \|G(u) - G_N(u)\|. \end{aligned} \tag{A.9}$$

By the triangle inequality,

$$\|G(u) + G_N(u) - 2y\| \le 2\max\{\|G(u) - y\|, \|G_N(u) - y\|\} = 2\max\{\Phi(u)^{1/2}, \Phi_N(u)^{1/2}\},$$

and the triangle inequality and (3.8) yield

$$\begin{aligned}\Phi_N(u)^{1/2} &= 2^{-1/2}\|G_N(u) - y\| \\ &= 2^{-1/2}\|G(u) - y + G_N(u) - G(u)\| \\ &\le 2^{-1/2}(2^{1/2}\Phi(u)^{1/2} + \|G_N(u) - G(u)\|) \\ &= \Phi(u)^{1/2} + 2^{-1/2}\|G_N(u) - G(u)\|.\end{aligned}$$

Together, these inequalities yield

$$\|G(u) - y + G_N(u) - y\| \le 2(\Phi(u)^{1/2} + 2^{-1/2}\|G_N(u) - G(u)\|),$$

and substituting the above into (A.9) yields

$$\big|\Phi(u) - \Phi_N(u)\big| \le 2C_\Gamma\Big(\Phi(u)^{1/2}\|G_N(u) - G(u)\| + \|G(u) - G_N(u)\|^2\Big),$$

thus proving (3.9). Using (A.3) yields

$$\big|\Phi(u) - \Phi_N(u)\big|^q \le 2^{q-1}(2C_\Gamma)^q\Big(\Phi(u)^{q/2}\|G_N(u) - G(u)\|^q + \|G(u) - G_N(u)\|^{2q}\Big).$$

Now take expectations with respect to ν_N: since G and Φ are constant with respect to ν_N,

$$\begin{aligned}\mathbb{E}_{\nu_N}\big[\big|\Phi(u) - \Phi_N(u)\big|^q\big] \le (4C_\Gamma)^q\Big(&\Phi(u)^{q/2}\mathbb{E}_{\nu_N}[\|G_N(u) - G(u)\|^q] \\ &+ \mathbb{E}_{\nu_N}\big[\|G(u) - G_N(u)\|^{2q}\big]\Big).\end{aligned}$$

and taking the q^{th} root of both sides proves (3.10). $\square$

Proof of Corollary 3.7. Taking the $L^s_{\mu_0}$ norm of both sides of the second inequality in Proposition 3.6, and applying (A.3) with $s/q \ge 1$, we obtain

$$\begin{aligned}&\big\|\mathbb{E}_{\nu_N}[|\Phi - \Phi_N|^q]^{1/q}\big\|_{L^s_{\mu_0}} \\ &\le (4C_\Gamma)\mathbb{E}_{\mu_0}\left[\Big(\Phi^{q/2}\mathbb{E}_{\nu_N}[\|G_N - G\|^q] + \mathbb{E}_{\nu_N}\big[\|G_N - G\|^{2q}\big]\Big)^{s/q}\right]^{1/s} \\ &\le (4C_\Gamma)2^{1/q-1/s}\Big(\mathbb{E}_{\mu_0}\Big[\Phi(u)^{s/2}\mathbb{E}_{\nu_N}[\|G_N - G\|^q]^{s/q}\Big] + \mathbb{E}_{\mu_0}\Big[\mathbb{E}_{\nu_N}\big[\|G_N - G\|^{2q}\big]^{s/q}\Big]\Big)^{1/s}.\end{aligned}$$

By the Cauchy–Schwarz inequality and Jensen's inequality,

$$\begin{aligned}\mathbb{E}_{\mu_0}\left[\Phi^{s/2}\mathbb{E}_{\nu_N}[\|G_N-G\|^q]^{s/q}\right] &\leq \left(\mathbb{E}_{\mu_0}[\Phi^s]\mathbb{E}_{\mu_0}\left[\mathbb{E}_{\nu_N}[\|G_N-G\|^q]^{2s/q}\right]\right)^{1/2} \\ &\leq \left(\mathbb{E}_{\mu_0}[\Phi^s]\mathbb{E}_{\mu_0}\left[\mathbb{E}_{\nu_N}\left[\|G_N-G\|^{2q}\right]^{s/q}\right]\right)^{1/2}.\end{aligned}$$

Given that $0 \leq a \leq 1$ implies that $a \leq a^{1/2}$, the hypothesis of the corollary and the preceding imply that

$$\left\|\mathbb{E}_{\nu_N}[|\Phi-\Phi_N|^q]^{1/q}\right\|_{L^s_{\mu_0}} \leq (4C_\Gamma)2^{1/q-1/s}\left(\mathbb{E}_{\mu_0}[\Phi^s]^{1/2}+1\right)^{1/s}\left\|\mathbb{E}_{\nu_N}\left[\|G_N-G\|^{2q}\right]^{1/q}\right\|^{1/2}_{L^s_{\mu_0}}.$$

Since $2^{1/q-1/s} \leq 2^{1/q} \leq 2$, the proof is complete. ☐

Proof of Lemma 3.8. Given (3.8), we may choose the parameter C_0 in (3.5) to be $C_0 = 0$. By Jensen's inequality, (3.11) implies (3.6). ☐

Proof of Theorem 3.9. We first verify that Assumption 3.3 holds. Since Φ and Φ_N satisfy (3.8), it follows that we may set $C_0 = 0$ in (3.5). Since we assume throughout that $0 < Z = \mathbb{E}_{\mu_0}[\exp(-\Phi)] < \infty$, it follows that Φ has moments of all orders, and hence belongs to $L^s_{\mu_0}$ for all $s \in \mathbb{N}$. Therefore, given that (3.11) holds for $q, s \geq 1$, it follows from Jensen's inequality and Corollary 3.7 that we can make $\|\mathbb{E}_{\nu_N}[|\Phi_N-\Phi|]\|_{L^1_{\mu_0}}$ as small as desired. In particular, for any $0 < C_3 < +\infty$ that satisfies $C_3^{-1} < Z < C_3$, there exists a $N^*(C_3) \in \mathbb{N}$ such that for all $N \geq N^*(C_3)$, (3.6) holds.

The rest of the proof consists of applying Lemma 3.4 or 3.5, Corollary 3.7 and Lemma 3.8.

Case (i). The hypotheses in this case ensure that we may apply Lemma 3.4. Set $p_1 = p^*$ and $p_2 = p_3 = +\infty$, so that $p_1' = (p^*)' = p^*/(p^*-1)$ and $p_2' = p_3' = 1$. Substituting these exponents into (3.3a) and applying Corollary 3.7 with $s = 2p_1'p_3' = 2p^*/(p^*-1)$ and $q = p_2' = 1$ (note that $s \geq q \geq 1$), we obtain

$$d_{\mathrm{H}}\big(\mu,\mu_N^{\mathrm{marginal}}\big) \leq C\big\|\mathbb{E}_{\nu_N}\left[|\Phi-\Phi_N|\right]\big\|_{L^{2p^*/(p^*-1)}_{\mu_0}(\mathcal{U})} \leq C\big\|\mathbb{E}_{\nu_N}\left[\|G-G_N\|^2\right]\big\|^{1/2}_{L^{2p^*/(p^*-1)}_{\mu_0}(\mathcal{U})},$$

where $C > 0$ changes value between inequalities. Thus we have proven that (3.12) holds with $r_1 = 1$ and $r_2 = 2p^*/(p^*-1)$.

To prove that (3.13) holds with the desired exponents, we again use Lemma 3.4 to set $q_1 = q_2 = +\infty$, so that $q_1' = q_2' = 1$. Substituting these exponents into (3.4), and applying Corollary 3.7 with $s = 2q_2' = 2$ and $q = 2q_1' = 2$, we obtain

$$\mathbb{E}_{\nu_N}\left[d_{\mathrm{H}}\big(\mu,\mu_N^{\mathrm{sample}}\big)^2\right]^{1/2} \leq D\left\|\mathbb{E}_{\nu_N}\left[|\Phi-\Phi_N|^2\right]^{1/2}\right\|_{L^2_{\mu_0}} \leq D\left\|\mathbb{E}_{\nu_N}\left[\|G-G_N\|^4\right]^{1/2}\right\|^{1/2}_{L^2_{\mu_0}(\mathcal{U})},$$

where $D > 0$ changes value between inequalities. Thus we have proven that (3.13) holds with $s_1 = s_2 = 2$.

It remains to ensure that both the rightmost terms above converge to zero. Since (3.11) holds with $q = 2$ and $s = 2p^*/(p^*-1)$, the desired convergence follows from the nesting property of L^p-spaces defined on finite measure spaces. Therefore, both $\mu_N^{\mathrm{marginal}}$ and μ_N^{sample} converge to μ in the appropriate metrics.

Case (ii). Since the arguments in this case are the same as in the previous case, we only record the different material.

The hypotheses ensure that we may apply Lemma 3.5. Set $p_1 = \rho^*$ and $p_2 = p_3 = +\infty$, so that $(p_1)' = \rho^*/(\rho^*-1)$ and $p_2' = p_3' = 1$. Substituting these exponents into (3.3a) and applying Corollary 3.7 with $s = 2p_1'p_3' = 2\rho^*/(\rho^*-1)$ and $q = p_2' = 1$, we obtain

$$d_{\mathrm{H}}\big(\mu,\mu_N^{\mathrm{marginal}}\big) \leq C\big\|\mathbb{E}_{\nu_N}\left[|\Phi-\Phi_N|\right]\big\|_{L^{2\rho^*/(\rho^*-1)}_{\mu_0}(\mathcal{U})} \leq C\big\|\mathbb{E}_{\nu_N}\left[\|G-G_N\|^2\right]\big\|^{1/2}_{L^{2p^*/(p^*-1)}_{\mu_0}(\mathcal{U})},$$

where $C > 0$ changes value between inequalities. Thus we have proven that (3.12) holds with $r_1 = 1$ and $r_2 = 2\rho^*/(\rho^*-1)$.

To prove that (3.13) holds with the desired exponents, we again use Lemma 3.5 to set $q_1 = \frac{\rho^*}{2}$ and $q_2 = +\infty$, so that $q_1' = \rho^*/(\rho^*-2)$ and $q_2' = 1$. Substituting these exponents into (3.4), and applying Corollary 3.7 with $s = 2q_2' = 2$ and $q = 2q_1' = 2\rho^*/(\rho^*-2)$, we obtain

$$\begin{aligned}\mathbb{E}_{\nu_N}\left[d_{\mathrm{H}}\big(\mu, \mu_N^{\text{sample}}\big)^2\right]^{1/2} &\le D\left\|\mathbb{E}_{\nu_N}\left[|\Phi - \Phi_N|^{2\rho^*/(\rho^*-2)}\right]^{(\rho^*-2)/(2\rho^*)}\right\|_{L^2_{\mu_0}} \\ &\le D\left\|\mathbb{E}_{\nu_N}\left[\|G - G_N\|^{4\rho^*/(\rho^*-2)}\right]^{(\rho^*-2)/(2\rho^*)}\right\|^{1/2}_{L^2_{\mu_0}(\mathcal{U})},\end{aligned}$$

where $D > 0$ changes value between inequalities. Thus (3.13) holds with $s_1 = 2\rho^*/(\rho^*-2)$ and $s_2 = 2$. Since (3.11) holds with $q = 2\rho^*/(\rho^*-2)$ and $s = 2\rho^*/(\rho^*-1)$, it follows from the nesting property of L^p-spaces defined on finite measure spaces that both

$$\left\|\mathbb{E}_{\nu_N}\left[\|G - G_N\|^2\right]\right\|^{1/2}_{L^{2p^*/(p^*-1)}_{\mu_0}(\mathcal{U})} \text{ and } \left\|\mathbb{E}_{\nu_N}\left[\|G - G_N\|^{4\rho^*/(\rho^*-2)}\right]^{(\rho^*-2)/(2\rho^*)}\right\|^{1/2}_{L^2_{\mu_0}(\mathcal{U})}$$

converge to zero. □

Proof of Proposition 4.1. We start by verifying the assumptions of Theorem 3.2. Firstly, since $\Phi(u) \ge 0$ for all $u \in \mathcal{U}$, and $\Phi_N(u) \ge 0$ for all $u \in \mathcal{U}$ and all $\{\sigma^{(i)}\}_{i=1}^N$, we have that assumption (a) is satisfied for $q_1 = q_2 = \infty$. For assumption (b), we then have for any $q_2 \in [1, \infty]$

$$\begin{aligned}&\left\|\Big(\mathbb{E}_\sigma\Big[\big(Z_N^{\text{sample}}\big)^{q_1}\max\{Z^{-3}, \big(Z_N^{\text{sample}}\big)^{-3}\}^{q_1}\big(\exp\big(-\Phi(u)\big) + \exp\big(-\Phi_N(u)\big)\big)^{2q_1}\Big]^{1/q_1}\right\|_{L^{q_2}_{\mu_0}(\mathcal{U})} \\ &\le 4\,\mathbb{E}_\sigma\Big[\big(Z_N^{\text{sample}}\big)^{q_1}\max\{Z^{-3}, \big(Z_N^{\text{sample}}\big)^{-3}\}^{q_1}\Big]^{1/q_1} \\ &\le 4\Big(Z^{-3q_1}\mathbb{E}_\sigma\big[\big(Z_N^{\text{sample}}\big)^{q_1}\big] + \mathbb{E}_\sigma\big[\big(Z_N^{\text{sample}}\big)^{-2q_1}\big]\Big)^{1/q_1}.\end{aligned}$$

Since $\Phi_N(u) \ge 0$ for all $u \in \mathcal{U}$ and all $\{\sigma^{(i)}\}_{i=1}^N$, we have for any $q_1 \in [1, \infty]$

$$\mathbb{E}_\sigma\Big[\big(Z_N^{\text{sample}}\big)^{q_1}\Big]^{1/q_1} = \mathbb{E}_\sigma\left[\left(\int_{\mathcal{U}}\exp(-\Phi_N(u))\,\mathrm{d}\mu_0(u)\right)^{q_1}\right]^{1/q_1} \le 1. \tag{A.10}$$

Using the ℓ-sparse distribution of σ, we further have $|\sigma_j^{(i)}| \le \sqrt{s}$ and

$$\Phi_N(u) = \frac{1}{2N}\sum_{i=1}^N\big|\sigma^{(i)\mathrm{T}}\big(\Gamma^{-1/2}(y - G(u))\big)\big|^2 \le \frac{s}{2}\big\|\big(\Gamma^{-1/2}(y - G(u))\big)\big\|^2 = s\Phi(u),$$

which implies that $Z_N^{\text{sample}} \ge Z_s = \int_{\mathcal{U}}\exp(-s\Phi(u))\,\mathrm{d}\mu_0(u)$. It follows that for any $q_1 \in [1, \infty]$

$$\mathbb{E}_\sigma\big[\big(Z_N^{\text{sample}}\big)^{-2q_1}\big]^{1/q_1} \le \mathbb{E}_\sigma\big[Z_s^{-2q_1}\big]^{1/q_1} = Z_s^{-2},$$

and assumption (b) is hence also satisfied for $q_1 = q_2 = \infty$. By Theorem 3.2, we hence have

$$\Big(\mathbb{E}_\sigma\big[d_{\mathrm{H}}\big(\mu, \mu_N^{\text{sample}}\big)^2\big]\Big)^{1/2} \le C\Big\|\big(\mathbb{E}_\sigma\big[|\Phi(u) - \Phi_N(u)|^2\big]\big)^{1/2}\Big\|_{L^2_{\mu_0}(\mathcal{U})}.$$

Using standard properties of Monte Carlo estimators (see e.g. Robert and Casella (1999)), we have

$$\Big(\mathbb{E}_\sigma\big[|\Phi(u) - \Phi_N(u)|^2\big]\Big)^{1/2} = \sqrt{\frac{\mathbb{V}_\sigma\big[\frac{1}{2}\big|\sigma^{\mathrm{T}}\Gamma^{-1/2}(y - G(u))\big|^2\big]}{N}}.$$

Now, using $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$, $\left(\sum_{j=1}^J x_j\right)^4 \le J^3 \sum_{j=1}^J x_j^4$, the linearity of expectation, the ℓ-sparse distribution of σ, and $\|x\|_4 \le \|x\|_2$, we have

$$
\begin{aligned}
0 &\le \mathbb{V}_\sigma\left[\frac{1}{2}\left|\sigma^{\mathrm{T}}\Gamma^{-1/2}(y - G(u))\right|^2\right] \\
&= \mathbb{E}_\sigma\left[\frac{1}{4}\left|\sigma^{\mathrm{T}}\Gamma^{-1/2}(y - G(u))\right|^4\right] - \mathbb{E}_\sigma\left[\frac{1}{4}\left|\sigma^{\mathrm{T}}\Gamma^{-1/2}(y - G(u))\right|^2\right]^2 \\
&= \mathbb{E}_\sigma\left[\frac{1}{4}\left(\sum_{j=1}^J \sigma_j \big(\Gamma^{-1/2}(y - G(u))\big)_j\right)^4\right] - \frac{1}{4}\big\|\Gamma^{-1/2}(y - G(u))\big\|^4 \\
&\le \frac{1}{4} J^3 \sum_{j=1}^J \mathbb{E}_\sigma[\sigma_j^4]\big(\Gamma^{-1/2}(y - G(u))\big)_j^4 - \frac{1}{4}\big\|\Gamma^{-1/2}(y - G(u))\big\|^4 \\
&= \frac{1}{4} J^3 \mathbb{E}_\sigma[\sigma_j^4]\big\|\Gamma^{-1/2}(y - G(u))\big\|_4^4 - \frac{1}{4}\big\|\Gamma^{-1/2}(y - G(u))\big\|^4 \\
&\le \big(J^3 \mathbb{E}_\sigma[\sigma_j^4] - 1\big)\Phi(u)^2.
\end{aligned}
$$

The claim (4.1) now follows, with the choice of constant as in (4.2). □

Proof of Theorem 5.6. Recall that T_J is the set of time points in $[0, T]$ with cardinality $|J| \in \mathbb{N}$. In (5.10), we observed that

$$
\|G_N(u) - G(u)\| \equiv \|G_N(u) - G(u)\|_{\mathcal{Y}} \le |J| \sup_{0 \le t \le T} \|e(t; u)\|_{\ell_2^d}.
$$

Fix $\rho^* > 2$. Omitting the argument u of Φ_N, Φ, G_N and G, we have

$$
\begin{aligned}
\exp\big(\rho^*\Phi_N\big) &= \exp\big(\rho^*\big(\Phi_N - \Phi + \Phi\big)\big) \\
&\le \exp\big(\rho^*|\Phi_N - \Phi| + \rho^*\Phi\big) \\
&\le \exp\big(\rho^*|\Phi_N - \Phi|\big)\exp(\rho^*\Phi) \\
&\le \exp\big(2\rho^* C_\Gamma\big(\Phi^{1/2}\|G_N - G\| + \|G - G_N\|^2\big)\big)\exp(\rho^*\Phi) \\
&\le \frac{\exp(\rho^*\Phi)}{2}\big[\exp\big(4\rho^* C_\Gamma \Phi^{1/2}\|G_N - G\|\big) + \exp\big(4\rho^* C_\Gamma\|G - G_N\|^2\big)\big]
\end{aligned}
$$

where the last two inequalities follow from (3.9) and Young's inequality $ab \le (a^2 + b^2)/2$ for $a, b \ge 0$. Using (5.10), we therefore obtain

$$
\begin{aligned}
\exp(\rho^*\Phi_N) \le \frac{\exp(\rho^*\Phi)}{2}\Bigg[&\exp\bigg(4\rho^* C_\Gamma \Phi^{1/2}|J| \sup_{0 \le t \le T}\|e(t)\|_{\ell_2^d}\bigg) \\
&+ \exp\bigg(4\rho^* C_\Gamma |J|^2 \sup_{0 \le t \le T}\|e(t)\|_{\ell_2^d}^2\bigg)\Bigg],
\end{aligned}
$$

where we note that we have suppressed the u-dependence of $e(t; u)$ and simply written $e(t)$. Since $\mathcal{U}$ is compact and S is continuous, it follows that G and hence Φ are continuous on $\mathcal{U}$; by the extreme value theorem, Φ is bounded on $\mathcal{U}$, i.e. $\|\Phi\|_{L^\infty_{\mu_0}(\mathcal{U})}$ is finite. Using this fact and taking expectations with respect to ν_N we obtain

$$
\begin{aligned}
\mathbb{E}_{\nu_N}[\exp(\rho^*\Phi_N(u))] \le \frac{\exp(\rho^*\|\Phi\|_{L^\infty_{\mu_0}(\mathcal{U})})}{2}\Bigg(&\mathbb{E}_{\nu_N}\bigg[\exp\bigg(4\rho^* C_\Gamma \|\Phi\|_{L^\infty_{\mu_0}(\mathcal{U})}^{1/2}|J| \sup_{0 \le t \le T}\|e(t; u)\|_{\ell_2^d}\bigg)\bigg] \\
&+ \mathbb{E}_{\nu_N}\bigg[\exp\bigg(4\rho^* C_\Gamma |J|^2 \sup_{0 \le t \le T}\|e(t; u)\|_{\ell_2^d}^2\bigg)\bigg]\Bigg).
\end{aligned}
$$

By Corollary 5.5, the two terms on the right-hand side are finite for every $u \in \mathcal{U}$. Given the continuous dependence of the parameters of Assumptions 5.1, 5.2, and 5.3 on u, and given that $\mathcal{U}$ is a compact subset of a finite-dimensional Euclidean space, it follows that the right-hand side can be bounded by a scalar that does not depend on any u. Hence, the function $u \mapsto \mathbb{E}_{\nu_N}[\exp(\rho^* \Phi_N(u))]$ belongs to $L^\infty_{\mu_0}(\mathcal{U})$, and certainly in $L^1_{\mu_0}(\mathcal{U})$, so that the first hypothesis of Theorem 3.9(b) holds. For the second hypothesis, observe that, since Assumption 5.3 holds for $R = +\infty$, it follows that (5.12) holds for any $n \in \mathbb{N}$, and thus (3.11) holds for any $q, s \geq 1$. Therefore the hypotheses of Theorem 3.9(b) are satisfied, and the desired conclusion follows from Theorem 3.9. $\square$