International Journal for Uncertainty Quantification, 12(4):65–92 (2022)

L^p CONVERGENCE OF APPROXIMATE MAPS AND PROBABILITY DENSITIES FOR FORWARD AND INVERSE PROBLEMS IN UNCERTAINTY QUANTIFICATION

T. Butler,[1,] T. Wildey,[2] & W. Zhang[1]*

[1]*University of Colorado Denver, Department of Mathematical and Statistical Sciences, Denver, Colorado 80204, USA*

[2]*Sandia National Laboratories, Center for Computing Research, Albuquerque, New Mexico 87185, USA*

*Address all correspondence to: T. Butler, University of Colorado Denver, Department of Mathematical and Statistical Sciences, Denver, Colorado 80204, USA, E-mail: troy.butler@ucdenver.edu

Original Manuscript Submitted: 2/16/2021; Final Draft Received: 11/28/2021

This work analyzes the convergence of probability densities solving uncertainty quantification problems for computational models where the mapping between input and output spaces is itself approximated. Specifically, we assume the exact mapping is replaced by a sequence of approximate maps that converges in L^p for some $1 \leq p < \infty$. To each approximate map, we then consider probability densities associated with push-forward and pullback measures solving forward and inverse problems, respectively. While the push-forward density is uniquely defined for each map, this is generally not the case for pullback densities since the maps are not typically bijective. A recently developed data-consistent inversion approach is therefore utilized to construct a specific sequence of pullback densities associated with the approximate maps. Convergence results for the push-forward and pullback densities are then proven under some additional assumptions. This significantly advances the results from a previous study that analyzed the convergence of such probability densities under the restrictive assumption that the approximate maps converged in an essentially uniform sense. Moreover, this greatly expands the realm of data-consistent inversion to problems requiring surrogate techniques that only guarantee L^p convergence for some $1 \leq p < \infty$. Numerical examples are also included along with numerical diagnostics of solutions and numerical verification of assumptions.

KEY WORDS: *inverse problems, uncertainty quantification, density estimation, L^p convergence, surrogate modeling, approximate modeling, polynomial chaos*

1. INTRODUCTION

Many uncertainty quantification (UQ) problems involve the propagation of uncertainties, described using probability measures, between the input and output spaces of a computational model. When the spaces are Euclidean and equipped with the Lebesgue measure, probability measures are often expressed uniquely as a probability density function (PDF or density).[‡] We are interested in densities solving two types of UQ problems in this work. The first density is on the output space of a model defined by the push-forward of an initial probability density specified on the inputs of a model. This represents a solution to a forward UQ problem where the goal is to predict probable outputs of a model using some initial/prior knowledge of model inputs. The second density is on the input space of a model defined by the pullback of an observed (or specified) probability density on model outputs. This represents a solution to a particular type of inverse UQ problem where the goal is to determine a density on model inputs whose push-forward through the model reconstructs the observed (or specified) probability density. Such inverse UQ problems

[‡]Or as Radon-Nikodym derivatives in more general cases.

2152–5080/22/$35.00

arise naturally in engineering design under uncertainty. At a high level, consider the manufacturing process for some engineered system that relies on various materials and subcomponents critical to obtaining desirable system outputs. Due to naturally occurring defects in materials and the tolerances set for machined or electronic components, the system inputs are naturally variable (i.e., following a distribution), which impacts the ability to control system outputs. Prior to the manufacturing, solving this type of inverse UQ problem is akin to specifying a desirable distribution of system outputs and determining the distributions on system inputs that is consistent with such outputs. Such a solution can inform the manufacturing process as various tolerances of components and purity standards for raw materials are assessed. If the manufacturing process is already underway, then obtaining data from various systems to identify the distribution on outputs and solving the same inverse UQ problem can identify system components with undesirable distributions that must be addressed to achieve more desirable outputs. We direct the interested reader to [1] where such inverse UQ problems are formulated in terms of a verification and validation and quality control problem in the design and manufacturing of thin elastic membranes, and in [2] they are used to learn a distribution of microstructure parameters in the context of computational materials science.

While solutions to the forward UQ problem are unique for a given mapping between input and output spaces, this is not typically the case for the inverse UQ problem since the map is generally not a bijection. The formulation and solution of the inverse UQ problem considered in this work can take several similar but distinct forms as seen in [3–5]. In [3], a method referred to as "Bayesian melding" is utilized to avoid the Borel paradox by incorporating both an implicit and explicit prior. The work of [4] utilizes generative adversarial networks to solve a constrained-optimization problem that minimizes divergence between distributions. We utilize the approach found in [5] where it is shown that incorporating an initial density and its associated push-forward density is sufficient for constructing a specific pullback density. We refer to the framework in [5] as defining data-consistent inversion to differentiate it from other techniques in the UQ literature that solve different types of inverse problems. Moreover, we refer to a pullback constructed in this way as an *updated* density since it is defined as a multiplicative update of the initial density. This is the method for producing a specific sequence of pullback densities in this work.

The existence, uniqueness, and stability of push-forward and pullback densities are also studied in [5]. However, such densities are often numerically estimated using a finite number of computational solutions of the model relating input and output spaces. We are particularly interested in cases where the outputs are defined by functionals applied to the solution to the model, which defines a (measurable) map between input and output spaces. Thus, we focus on the underlying, and often implicitly defined, map between input and output spaces instead of the actual model itself.

The primary contribution of this work is the convergence analysis for push-forward and pullback densities defined by a sequence of approximate maps that converge in L^p with $1 \leq p < \infty$. For instance, we may construct approximate maps using generalized polynomial chaos expansions (PCEs) [6,7], sparse grid interpolation [8,9], and Gaussian process models [10,11]. Take PCEs as an example: according to the Cameron-Martin theorem, the Hermite-chaos expansion converges for any arbitrary random process with finite second-order moments; i.e., the convergence of approximate maps occurs in L^2. Consequently, the theoretical results of this work significantly generalize the analysis presented in [12] where it is assumed that the approximate maps converge *essentially uniformly* (i.e., in L^∞). *Moreover, this analysis greatly expands the realm of problems for which data-consistent inversion is reliably applied including problems requiring surrogate techniques commonly utilized in the UQ community.* Another contribution of this work is the weakening of a key assumption in [12] that required an asymptotic notion of continuity along with a uniform boundedness of the densities to hold almost everywhere (a.e.), i.e., everywhere except on a set of possibly zero measure. In this work, these conditions can fail on sets of positive measure.

While the theorems and proofs presented in this work are both quite technical and require a significant amount of measure theory, we provide straightforward interpretations of the main results throughout to make these results more accessible to a wider audience. To properly understand these interpretations, we emphasize that the updated density solving the data-consistent inverse problem for a particular map involves the combination of three distinct densities: the initial density on model inputs, the push-forward of the initial density on model outputs, and the observed density on model outputs. A key step in the construction of a particular updated density for a given map is the evaluation of both the push-forward of the initial density and observed density on the outputs obtained by evaluating the given map on model inputs. We refer to the outputs obtained by evaluating a given map on model inputs as the data points predicted by evaluating the map under consideration. We therefore state and interpret convergence results for the

approximate push-forward densities (associated with an approximate map) at data points predicted by evaluating the approximate maps. As an example, the main forward analysis result is interpreted simply as follows:

> As the approximate maps converge to the exact map, the approximate push-forward densities (evaluated at data points predicted from the associated approximate maps) converge to the exact push-forward density (evaluated at the data points predicted from the exact map).

This result permits the proof of convergence of approximate updated densities, which has the following interpretation:

> Under a predictability assumption, as the approximate maps converge to the exact map, the approximate updated densities associated with the approximate maps converge to the exact updated density associated with the exact map.

Most results are framed as convergence in L^p both for simplicity and to coincide with the convergence of the approximate maps. However, it is worth noting that either finite measurability of the parameter space or convergence of the approximate maps in L^s for any $1 \leq s \leq p$ immediately results in L^s convergence of both push-forward densities and updated densities for any $1 \leq s \leq p$ by standard results in measure theory (cf. Proposition 6.12 in [13]).

The remainder of the paper is organized as follows. In Section 2, we summarize most of the notation and terminology used in this work. We introduce the measure-theoretic notion of a condition holding in an *almost* sense in Section 3 where it is contrasted with the more familiar (and restrictive) *almost everywhere* condition. This sets the stage for the forward and inverse problem analyses that follow in Sections 4 and 5, respectively, as well as for how these analyses differ from the previous L^∞ analysis of [12]. We discuss the impact of computational estimates of densities in Section 6.1. While the two main assumptions in this work are difficult to prove analytically *a priori*, Section 6.2 discusses how we numerically verify these assumptions when they hold in a stronger sense. Section 7 includes some additional discussion on the generality of the assumptions used in this work in the context of an example involving a singular push-forward density. This includes discussion on how we may utilize the numerical tools and diagnostics described in Section 6.2 to identify sources of error and improve accuracy of estimated densities. The main numerical results that employ PCEs follow in Section 8. Concluding remarks are given in Section 9. Detailed proofs for the theoretical results presented in Sections 4 and 5 are provided in Appendix A and Appendix B, respectively. Finally, Appendix C details how to obtain the scripts and data sets utilized in creating all of the figures and table data in this manuscript.

2. THE SPACES AND MAPS

Here, we summarize some common terminology, notation, and implicit assumptions used for both the forward and inverse analysis in this work. Denote by $\mathbf{\Lambda} \subset \mathbb{R}^k$ the space of all *physically possible* inputs to the model, which we refer to as parameters. A quantity (or quantities) of interest (QoI) refers to the functional(s) applied to the solution space of the model corresponding to scalar (or vector-valued) outputs. Depending on the problem, a QoI map may correspond to either output values we wish to predict or values for which we have already obtained data, e.g., for the purpose of model validation or calibration. The parameter-to-QoI map (often referred to simply as the QoI map) is denoted by $Q(\lambda) : \mathbf{\Lambda} \to \mathcal{D} \subset \mathbb{R}^m$ to make explicit the dependence on model parameters. Let (Q_n) denote a sequence of approximate QoI maps, $Q_n : \mathbf{\Lambda} \to \mathcal{D}$. Here, $\mathcal{D}$ denotes the range of all QoI maps (i.e., $\mathcal{D}$ is defined as the union of the ranges), which indicates the space of output data that can be predicted by at least one of the maps under consideration.

Assume that $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ and $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, \mu_{\mathcal{D}})$ are both measure spaces. Here, $\mathcal{B}_{\mathbf{\Lambda}}$ and $\mathcal{B}_{\mathcal{D}}$ denote the Borel σ-algebras inherited from the metric topologies on their respective spaces, and $\mu_{\mathbf{\Lambda}}$ and $\mu_{\mathcal{D}}$ denote the dominating measures for which probability densities (i.e., Radon-Nikodym derivatives of probability measures) are defined on each space. It is implicitly assumed that every QoI map is a measurable map between these measure spaces. The assumption, stated explicitly throughout this work, is that $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$. The convergence of densities is then considered on either $L^r(\mathcal{D})$ or $L^p(\mathbf{\Lambda})$. By $L^r(\mathcal{D})$ or $L^p(\mathbf{\Lambda})$, unless stated otherwise, it is implicitly assumed that the integrals are with respect to the dominating measures $\mu_{\mathcal{D}}$ or $\mu_{\mathbf{\Lambda}}$, respectively. A practical assumption implicitly made

by computations involving finite sampling of these maps and standard density approximation techniques is that each map is also piecewise smooth.

3. THE "ALMOST" CONCEPT IN MEASURE THEORY

To both set the stage and highlight some theoretical challenges in generalizing the previous L^∞ analysis in [12], we briefly summarize some results and concepts from measure theory (see, e.g., [13] for a concise introduction or [14] for a more thorough treatise). First, it is well known that if a sequence of measurable functions converges in L^p, then there exists a subsequence which converges *almost everywhere* (a.e.). In other words, the subsequence converges except on a set of measure zero.

Egorov's Theorem (cf. Theorem 2.33 in [13]) states that any sequence converging a.e. actually has the property that it converges *almost uniformly* (i.e., in L^∞ in an almost sense). The measure-theoretic distinction between a property holding a.e. as opposed to holding in an almost sense is subtle when first encountered. Specifically, a property is said to hold in an almost sense on a measure space $(X, \mathcal{B}_X, \mu_X)$ with $\mu_X(X) < \infty$ if for any $\epsilon > 0$, there exists $A_\epsilon \in \mathcal{B}_X$ such that $\mu_X(A_\epsilon) < \epsilon$ and the property holds on $X \backslash A_\epsilon$. This is loosely interpreted as stating that a property holds except on sets of small, but positive, measure. If $\mu_X(X) = \infty$, then we instead say that a property holds in the almost sense if for any $A \in \mathcal{B}_X$ with $\mu_X(A) < \infty$ and $\epsilon > 0$, there exists $A_\epsilon \in \mathcal{B}_X$ such that $\mu_X(A_\epsilon) < \epsilon$ and the property holds on $A \backslash A_\epsilon$.

It follows that the results of [12] may apply to a subsequence of the L^p maps restricted to a "large" subset of the spaces. However, while the existence results of the subsequence and subset are powerful theoretical tools, they are of little use practically. Specifically, it is not at all clear, in general, how one is to determine either the subsequence of maps or the subset of the spaces on which to carry out the analysis. Subsequently, it is unclear which part of the sequence of approximate maps or subsets of the spaces we can even apply the results from [12] when the maps converge in L^p.

The convergence results in this work apply directly on the entire sequence (not a subsequence) of approximate solutions to both the forward and inverse problems when using approximate maps that converge in L^p. Moreover, this convergence is proven with a weaker assumption than used in [12] involving the asymptotic and boundedness properties of densities solving the forward problem. Specifically, we replace the a.e. criteria required in the analysis of [12] with an almost criteria as shown in Section 4.

4. FORWARD PROBLEM ANALYSIS

In this section we analyze the convergence of push-forward densities obtained from solving a forward UQ problem using a sequence of approximate maps which converge to the exact map in the L^p sense where p is fixed and $1 \leq p < \infty$.

4.1 Problem Definition

Definition 1 (Forward problem and push-forward measure)**.** Let $P_{\mathbf{\Lambda}}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ be a given probability measure that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$ and admits a density $\pi_{\mathbf{\Lambda}}$. The forward problem is to determine the push-forward probability measure,

$$P_{\mathcal{D}}^{Q}(A) = P_{\mathbf{\Lambda}}(Q^{-1}(A)), \quad \forall A \in \mathcal{B}_{\mathcal{D}},$$

on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ that is absolutely continuous with respect to $\mu_{\mathcal{D}}$ and admits a density $\pi_{\mathcal{D}}^{Q}$.

4.2 Convergence Analysis

Denote by $(\pi_{\mathcal{D}}^{Q_n})$ the sequence of approximate push-forward densities. The major takeaway of the forward analysis is that under certain conditions $\pi_{\mathcal{D}}^{Q_n}(Q_n(\cdot)) \to \pi_{\mathcal{D}}^{Q}(Q(\cdot))$ in $L^p(\mathbf{\Lambda})$. The interpretation of this result is as follows:

> The approximate push-forward densities evaluated at the approximate predicted QoI converge to the exact push-forward density evaluated at the exactly predicted QoI in the L^p sense.

Of particular note is that this convergence is over the parameter space.

The assumption of $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$ implies convergence in distribution (i.e., the push-forward distributions converge at all continuity points of the limit distribution). However, this is not sufficient to guarantee convergence of the densities. Specifically, we cannot guarantee that $\pi_{\mathcal{D}}^{Q_n}(\cdot) \to \pi_{\mathcal{D}}^{Q}(\cdot)$ converges even pointwise on $\mathcal{D}$ let alone the desired result that $\pi_{\mathcal{D}}^{Q_n}(Q_n(\cdot)) \to \pi_{\mathcal{D}}^{Q}(Q(\cdot))$ in $L^p(\mathbf{\Lambda})$. For this, we need to provide conditions that remove pathological cases from consideration. The work of [15] defines sufficient conditions in terms of weaker asymptotic notions of equicontinuity and uniform equicontinuity.

Definition 2. Using similar notation from [15], we say that a sequence of real-valued functions (u_n) defined on $\mathbb{R}^k$ is *asymptotically equicontinuous (a.e.c.)* at $x \in \mathbb{R}^k$ if

$$\forall \epsilon > 0,\ \exists \delta(x,\epsilon) > 0, n(x,\epsilon) \text{ s.t. } |y - x| < \delta(x,\epsilon), n > n(x,\epsilon) \Rightarrow |u_n(y) - u_n(x)| < \epsilon.$$

If $\delta(x,\epsilon) = \delta(\epsilon)$ and $n(x,\epsilon) = n(\epsilon)$, then we say that the sequence is *asymptotically uniformly equicontinuous (a.u.e.c.)*.

Below, we give our first formal assumption.

Assumption 1. The sequence of approximate push-forward densities $(\pi_{\mathcal{D}}^{Q_n})$ is almost uniformly bounded and almost a.e.c.

That the properties in Assumption 1 hold in an almost sense is similar to the concept of *tightness* of measures. For example, when X has a topology (so that $\mathcal{B}_X$ represents the Borel σ-algebra on X), then a family of probability measures $\mathcal{P}$ on $(X, \mathcal{B}_X)$ is considered tight if for any $\epsilon > 0$ there exists compact $K_\epsilon \in \mathcal{B}_X$ such that for any $\mathbb{P} \in \mathcal{P}$, $\mathbb{P}(K_\epsilon) > 1 - \epsilon$. In other words, there exists a compact set containing most of the probability no matter which probability measure is considered. Assumptions of tightness are often assumed in classical results involving the convergence of measures. For example, Prokhorov's Theorem [16] states that tightness of probability measures is a necessary and sufficient condition for the precompactness of these measures in the topology of weak convergence. It is therefore not surprising that the almost criteria play a featured role in Assumption 1, which is utilized in proving convergence of push-forward densities below. Moreover, we return to the concept of tightness more explicitly in Lemma 3 in Section 5 where Assumption 1 is key to proving a tightness result for the push-forward densities.

The next two lemmas describe the types of convergence of the approximate push-forward densities that occur on $\mathcal{D}$ under certain conditions. The proofs are provided in Appendix A.

Lemma 1. *Suppose* $1 \le p < \infty$ *and* $Q_n \to Q$ *in* $L^p(\mathbf{\Lambda})$. *If Assumption 1 holds, then*

$$\pi_{\mathcal{D}}^{Q_n} \to \pi_{\mathcal{D}}^{Q} \qquad \text{in } \mathcal{D} \text{ in an almost sense}. \tag{1}$$

Furthermore, for any compact subset $D_c \subset \mathcal{D}$ *and* $1 \le r \le \infty$,

$$\pi_{\mathcal{D}}^{Q_n} \to \pi_{\mathcal{D}}^{Q} \qquad \text{in } L^r(D_c) \text{ in an almost sense}. \tag{2}$$

To extend the convergence in Eq. (2) to all of D_c, we require some mechanism for controlling the size of the L^r-norms of the push-forward densities on the "small" sets of positive measure (i.e., the sets denoted by N_δ in the proof of Lemma 1 in Appendix A). In other words, we need to ensure that the probability mass of the push-forwards is not collecting in data sets of arbitrarily small measure. An assumption of uniform integrability[§] avoids such pathological families of densities.

Lemma 2. *Suppose* $1 \le p < \infty$ *and* $Q_n \to Q$ *in* $L^p(\mathbf{\Lambda})$. *If Assumption 1 holds and the family of push-forward densities defined by these maps is uniformly integrable in* $L^r(\mathcal{D})$ *for some* $1 \le r < \infty$, *then for any compact subset* $D_c \subset \mathcal{D}$,

$$\pi_{\mathcal{D}}^{Q_n} \to \pi_{\mathcal{D}}^{Q} \qquad \text{in } L^r(D_c). \tag{3}$$

[§] On a measure space $(X, \mathcal{B}_X, \mu_X)$, a family of measurable functions $\{f_\alpha\}_{\alpha \in \mathcal{A}}$ is uniformly integrable if given $\epsilon > 0$ there exists M such that $\int_{\{x\,:\,|f_\alpha(x)| > M\}} |f_\alpha(x)|\, d\mu_X < \epsilon$ for all $\alpha \in \mathcal{A}$. We say that the family is uniformly integrable in $L^r(X)$ for some $1 \le r < \infty$ if $\{f_\alpha^r\}$ is uniformly integrable.

We now state the main result of the forward analysis where we again emphasize the interpretation that the approximate push-forward densities evaluated at the approximate QoI converge to the exact push-forward density evaluated at the exact QoI in $L^p(\mathbf{\Lambda})$. The proof of this result is in Appendix A.

Theorem 1 (L^p convergence of push-forward densities)**.** *Suppose $1 \leq p < \infty$ and $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$. If Assumption 1 holds and the family of push-forward densities defined by these maps is uniformly integrable in $L^p(\mathcal{D})$, and $\mathcal{D}$ is compact,*

$$\pi_{\mathcal{D}}^{Q_n}(Q_n(\cdot)) \to \pi_{\mathcal{D}}^{Q}(Q(\cdot)) \text{ in } L^p(\mathbf{\Lambda}). \tag{4}$$

The following corollary is included for completeness and summarizes the conclusions when the conditions in Assumption 1 are strengthened to hold a.e. instead of in an almost sense. The result is a strengthening of the conclusions in Lemma 1, not requiring Lemma 2 (the uniform integrability in L^p is now a given), while the conclusion of Theorem 1 remains unchanged. The changes to the proofs given above are straightforward (one can entirely avoid the construction of the "small" sets in the proofs) and are therefore omitted.

Corollary 1. *Suppose $1 \leq p < \infty$ and $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$. If the criteria in Assumption 1 hold a.e. instead of in an almost sense, then (1) $\pi_{\mathcal{D}}^{Q}$ is a.e. continuous on $\mathcal{D}$, and (2)*

$$\pi_{\mathcal{D}}^{Q_n} \to \pi_{\mathcal{D}}^{Q} \qquad \text{for a.e. } q \in \mathcal{D}. \tag{5}$$

Furthermore, for any compact subset $D_c \subset \mathcal{D}$ and $1 \leq r \leq \infty$,

$$\pi_{\mathcal{D}}^{Q_n} \to \pi_{\mathcal{D}}^{Q} \qquad \text{in } L^r(D_c). \tag{6}$$

Moreover, if $\mathcal{D}$ is compact,

$$\pi_{\mathcal{D}}^{Q_n}(Q_n(\cdot)) \to \pi_{\mathcal{D}}^{Q}(Q(\cdot)) \qquad \text{in } L^p(\mathbf{\Lambda}). \tag{7}$$

5. INVERSE PROBLEM ANALYSIS

In this section we analyze the convergence of probability density functions obtained from solving the inverse problem using approximate maps that converge to the exact map in the L^p sense where p is fixed and $1 \leq p < \infty$.

5.1 Problem Definition and Solution

We begin by defining and summarizing the inverse problem and its solution considered in this work. For a more thorough discussion on this inverse problem and the theory of existence, uniqueness, and stability of solutions, we direct the interested reader to [5], which also includes a comparison to alternative formulations and solutions of inverse problems including a specific comparison to the popular Bayesian formulations of inverse problems [17–23].

Definition 3 (Inverse problem and consistent measure)**.** Let $P_{\mathbf{\Lambda}}$ on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}})$ be a given probability measure that is absolutely continuous with respect to $\mu_{\mathcal{D}}$ and admits a density $\pi_{\mathcal{D}}$. The inverse problem is to determine a probability measure $P_{\mathbf{\Lambda}}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ that is absolutely continuous with respect to $\mu_{\mathbf{\Lambda}}$ and admits a probability density $\pi_{\mathbf{\Lambda}}$, such that the subsequent push-forward measure induced by the map, $Q(\lambda)$, satisfies

$$P_{\mathbf{\Lambda}}(Q^{-1}(A)) = P_{\mathcal{D}}^{Q}(A) = P_{\mathbf{\Lambda}}(A), \tag{8}$$

for any $A \in \mathcal{B}_{\mathcal{D}}$. We refer to any probability measure $P_{\mathbf{\Lambda}}$ that satisfies Eq. (8) as a **consistent** solution to the inverse problem.

As noted in the Introduction, the solutions to the inverse problem are generally not unique since the map is not assumed to be a bijection. However, various approaches for solving the inverse problem produce a specific pullback such as Bayesian melding [3] and generative adversarial networks applied to a constrained-optimization version of the problem [4]. We follow the approach given in [5] and incorporate an *initial* density, which we denote by $\pi_{\mathbf{\Lambda}}^{\text{init}}$. This requires the following predictability assumption based on the push-forward densities of the initial density that ensures the solvability of the inverse problem using either the approximate or exact QoI maps.

Assumption 2. There exists $C > 0$ such that for any $n \in \mathbb{N}$ and a.e. $q \in \mathcal{D}$, $\pi_{\mathcal{D}}(q) \leq C\pi_{\mathcal{D}}^{Q}(q)$, and $\pi_{\mathcal{D}}(q) \leq C\pi_{\mathcal{D}}^{Q_n}(q)$. Here, $\pi_{\mathcal{D}}^{Q}$ and $\pi_{\mathcal{D}}^{Q_n}$ are the push-forward densities of a given initial density $\pi_{\mathbf{\Lambda}}^{\text{init}}$.

This assumption is a modified form of the predictability assumptions previously used in [5,12]. This is referred to as a predictability assumption because it ensures that QoI data that are likely to be observed are also likely to be predicted by the push-forward densities associated with the initial density. It is important to note that this is an assumption on *both* the initial density and the QoI maps.

To help the reader develop more intuition and appreciation for this assumption, we briefly describe its importance from a computational perspective and how it ensures that rejection sampling[¶] is a viable numerical approach to generate independent identically distributed (iid) sets of samples from a consistent solution to the inverse problem as established in [5]. Specifically, we can first generate a set of iid samples from the initial density and then propagate these samples to the output space using any fixed QoI map. This defines a set of *proposal samples* in the output space. Then, we can perform rejection sampling using the *target* observed density. This defines a set of accepted samples in the output space along with a corresponding set of parameter samples. The accepted samples in the output space are iid samples from the observed density. Subsequently, the corresponding set of parameter samples are iid samples from a consistent solution defined as the *updated density* to emphasize how the information from the QoI map has effectively updated the initial density. This updated density has a closed form expression that follows by an application of the Disintegration Theorem [24] for a QoI map satisfying a predictability assumption.

Theorem 2 (Existence and uniqueness of solutions [5])**.** *Let $P_{\mathbf{\Lambda}}^{init}$ denote an initial probability measure on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}}, \mu_{\mathbf{\Lambda}})$ with density $\pi_{\mathbf{\Lambda}}^{init}$, and let $P_{\mathbf{\Lambda}}$ denote an observed probability measure on $(\mathcal{D}, \mathcal{B}_{\mathcal{D}}, \mu_{\mathcal{D}})$ with density $\pi_{\mathcal{D}}$. For a given QoI map, Q, that satisfies Assumption 2, the updated probability measure $P_{\mathbf{\Lambda}}^{up}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$ defined by*

$$P_{\mathbf{\Lambda}}^{up}(A) = \int_{\mathcal{D}} \left(\int_{A \cap Q^{-1}(q)} \pi_{\mathbf{\Lambda}}^{init}(\lambda) \frac{\pi_{\mathcal{D}}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q}(Q(\lambda))} d\mu_{\mathbf{\Lambda},q}(\lambda) \right) d\mu_{\mathcal{D}}(q), \quad \forall A \in \mathcal{B}_{\mathbf{\Lambda}} \tag{9}$$

is a consistent solution to the inverse problem in the sense of (8) *and is uniquely determined for the given initial probability measure $P_{\mathbf{\Lambda}}^{init}$ on $(\mathbf{\Lambda}, \mathcal{B}_{\mathbf{\Lambda}})$. Here, $\mu_{\mathbf{\Lambda},q}$ denotes the disintegrated measure of $\mu_{\mathbf{\Lambda}}$.*[||]

The updated density to the exact QoI map is then identified as

$$\pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda) = \pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda) \frac{\pi_{\mathcal{D}}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q}(Q(\lambda))}, \quad \lambda \in \Lambda, \tag{10}$$

which we often write as

$$\pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda) = \pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda) r(\lambda), \quad \text{with } r(\lambda) := \frac{\pi_{\mathcal{D}}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q}(Q(\lambda))}. \tag{11}$$

Here, the ratio denoted by $r(\lambda)$ has a practical interpretation as a *rejection ratio* for a randomly generated sample from $\pi_{\mathbf{\Lambda}}^{\text{init}}$.

Similarly, for any of the approximate QoI maps that satisfy the predictability assumption, the approximate updated densities are expressed as

$$\pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) := \pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda) \frac{\pi_{\mathcal{D}}(Q_n(\lambda))}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} = \pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda) r_n(\lambda), \quad \text{with } r_n(\lambda) := \frac{\pi_{\mathcal{D}}(Q_n(\lambda))}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))}. \tag{12}$$

[¶] The algorithm for rejection sampling is quite simple, so we summarize it here. Suppose we wish to obtain a sample from distribution X with density f using samples from distribution Y with density g. To do this, it is first necessary that there exists $M > 0$ such that $f(x) \leq Mg(x)$ for a.e. x. Supposing this is true, we then generate a sample y from distribution Y and $u \sim U(0,1)$. If $u < f(y)/(Mg(y))$, then we accept y as a sample drawn from X. Otherwise, we reject y and resample from Y while also generating a new $u \sim U(0,1)$. For data-consistent inversion, $\pi_{\mathcal{D}}$ and $\pi_{\mathcal{D}}^{Q}$ play the roles of f and g, respectively, and the C in the predictability assumption plays the role of M.

[||] For those unfamiliar with disintegrations of measures, it is helpful to think of this as a nonlinear version of Fubini's Theorem (e.g., see Theorem 2.37 in [13]) that justifies rewriting an integral over a product space as an iterated integral over the component spaces. This practice of rewriting an integral in terms of an iterated integral is often first encountered in multivariate calculus where integrals over an area or volume are rewritten as "double" or "triple" iterated integrals, respectively.

5.2 Convergence Analysis

Before presenting the results, we emphasize the main takeaway and interpretation from the inverse analysis as follows:

> Under the additional constraint of Assumption 2 (i.e., the predictability assumption), the approximate updated densities converge to the exact updated density in the L^p sense.

The following lemma is interpreted as providing a specific form for the subset of the data space on which the family of probability measures defined by the exact push-forward and the tail end of the approximate push-forward probability measures (defined by their densities) is considered tight. The proof is provided in Appendix B.

Lemma 3. *Suppose $1 \leq p < \infty$ and $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$. If Assumption 1 holds, then for any $\delta > 0$, there exists $a > 0$, compact $D_a \in \mathcal{B}_{\mathcal{D}}$ and $N > 0$ such that for any $n > N$,*

$$\pi_{\mathcal{D}}^{Q}(q) > a,\ \pi_{\mathcal{D}}^{Q_n}(q) > a, \quad \forall q \in D_a. \tag{13}$$

$$\int_{D_a} \pi_{\mathcal{D}}^{Q}(q)\, d\mu_{\mathcal{D}} > 1 - \delta, \quad \int_{D_a} \pi_{\mathcal{D}}^{Q_n}(q)\, d\mu_{\mathcal{D}} > 1 - \delta. \tag{14}$$

In actuality, a weaker conclusion than this lemma provides is needed to prove Theorem 3 below. Specifically, it is sufficient to prove that there exists a sequence of subsets in the data space associated with each $\pi_{\mathcal{D}}^{Q_n}$ containing "most" of the probability for each density while simultaneously bounding each density from below. However, the existence of a common data set not only serves to simplify the notation in the proof of the following theorem (we can avoid applying a subscript n to sets), but it also provides a useful conceptualization of this set in terms of a tightness property of the densities. This common set also implies a type of "effective support" containing "most of the predicted probability" for both the exact push-forward density and the tail end of the sequence of approximate push-forward densities. With this lemma in hand, we now state the main result of this section giving the convergence of approximate updated densities in an L^p sense when the approximate maps converge in an L^p sense. The proof is provided in Appendix B.

Theorem 3 (Convergence of updated densities)**.** *Suppose $1 \leq p < \infty$ and $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$. If $\pi_{\mathbf{\Lambda}}^{init} \in L^p(\mathbf{\Lambda})$, $\pi_{\mathcal{D}} \in L^p(\mathcal{D})$, Assumptions 1 and 2 hold, and the family of push-forward densities defined by these maps are uniformly integrable in $L^p(\mathcal{D})$, then*

$$\pi_{\mathbf{\Lambda}}^{up,n} \to \pi_{\mathbf{\Lambda}}^{up} \ \text{ in } L^p(\mathbf{\Lambda}). \tag{15}$$

6. NUMERICAL CONSIDERATIONS, COMPUTATIONAL ESTIMATES, AND IMPACT ON ASSUMPTIONS

6.1 Finite Sampling and Density Estimation

We first discuss how estimating the push-forward of an initial density (for either the exact or approximate QoI maps) using straightforward finite sampling techniques impacts the solutions to the forward and inverse problems considered in this work. Specifically, suppose that we first generate a finite set of iid samples from the initial density. Then, propagating this sample set through the QoI map constructs an iid sample set from the (unknown) push-forward density. While this sample set comes from the correct push-forward density (relative to the map used), we ultimately perform analysis on the density.

Kernel density estimation (KDE) techniques generally work well for estimating the push-forward densities especially when $\mathcal{D}$ is low-dimensional (e.g., see [12] and the references therein). When $\mathcal{D}$ is high-dimensional and the number of iid samples we can generate from the push-forward is limited (e.g., due to a computationally expensive QoI map), we may instead opt for parametric estimations of the density. Recent studies have analyzed other approaches for approximating densities. For instance, [25] considers the use of truncated Gram-Charlier or Edgeworth series expansions for the random variables, [26] utilizes sparse polynomial expansions or ReLU (rectified linear unit) networks to achieve exponential convergence when the map is a triangular monotone transport, and [27] develops a novel spline-based algorithm whose convergence rate in L^p is polynomial with respect to the sampling resolution. Moreover, the

work of [27] provides convergence rates for density estimation with any surrogate model that approximates both the QoI map and its gradient in L^∞. Whatever density estimation scheme is used introduces an error in both the solution to the forward problem and subsequently in the solution to the inverse problem.

In [12], we analyze the impact of KDE error on the forward and inverse solutions when the QoI maps converge in L^∞. Perhaps the most direct approach to perform a similar analysis of KDE errors in this work is to appeal to the L^1-norm. This is possible since L^p-convergence implies L^1-convergence (for finite measure spaces such as probability spaces) for all $1 < p \leq \infty$. Specifically, the L^p-norm, when scaled by a constant related to the measure of the space, provides an upper bound for the L^1-norm. Then, depending on any additional smoothness assumptions in the densities, we can apply KDE error bounds available in the literature that are either written in the L^1-norm or can be appropriately rewritten in the L^1-norm. For example, [28] provides uniform convergence rates on KDEs dependent upon the assumed level of smoothness in the densities and [29] proves L^1-convergence without any continuity assumptions on the densities. However, for the sake of brevity, we omit this additional error analysis in this work. Moreover, in the numerical examples, we follow the same procedure to compute and compare densities to avoid addressing these errors. Specifically, we first generate a single initial set of samples for each problem. We then compute estimates of the various densities associated with the approximate and exact maps using the same initial set of samples and a standard Gaussian KDE. Finally, errors are computed between the estimated densities associated with the approximate maps and the estimated densities associated with the exact map. By avoiding computing the error between an estimated density associated with an approximate map and the exact density associated with the exact map, we omit the KDE contribution to the error in the numerical results.

6.2 Assumptions: Context, Generality, and Numerical Verification

We now discuss the two main assumptions in this work primarily in the context of using numerical estimates of the approximate push-forward densities. Below, we simply refer to these as the estimated densities to distinguish from the terminology of approximate push-forward densities previously used that are in fact "exact" relative to the associated approximate QoI maps. For notational simplicity, we abuse notation and use $\pi_{\mathcal{D}}^{Q_n}$ to denote the estimated densities. For both the sake of simplicity and in light of the fact that Lipschitz continuous functions with compact support are dense in L^r for any $1 \leq r < \infty$ (e.g., stated as Lemma 4 in Appendix A.3), we assume that Gaussian KDEs are used so that all numerical estimates of approximate densities are Lipschitz continuous. This has an added conceptual and computational advantage, as we discuss below, in contextualizing the equicontinuity criterion of Assumption 1. Specifically, a well-known result is that if a sequence of Lipschitz continuous functions has a common Lipschitz constant, then the sequence of functions is equicontinuous.

6.2.1 Assumption 1

Recall that Assumption 1 involves two criteria: uniform boundedness and an asymptotic notion of equicontinuity. The assumption is that these criteria hold in an almost sense, which is rather permissive and allows for the theory to apply even in the presence of a countably infinite number of singularities in push-forward densities. This is discussed further in Section 7 in the context of a simple example with a push-forward density exhibiting a singularity. Thus, we generally expect that this assumption holds except for pathological examples. Unfortunately, verifying any criterion holds only in an almost sense is not necessarily straightforward. However, it is relatively straightforward to use the estimated densities to investigate if these criteria hold in the stronger a.e. sense. We first require some notation that we formally define below for ease of reference.

Definition 4. Denote by $(B_{n,m})$ and $(L_{n,m})$ the sequence of bounds and Lipschitz constants, respectively, for the estimates of $(\pi_{\mathcal{D}}^{Q_n})$ where n refers to the use of the nth approximate map and m refers to the number of parameter samples used to produce the estimated density.

Estimates of $(B_{n,m})$ may be obtained, for instance, by simply sorting the evaluation of density estimates associated with the nth map at a set of m iid samples used to form the estimated densities. To obtain estimates of $(L_{n,m})$, one approach is to first compute linear combinations of gradients of the kernel used in the numerical estimates of the

densities, restrict evaluation to the m iid samples used to construct the estimate, take the norms of these gradients, and then finally apply a sorting algorithm. For many kernels, such as the Gaussian one used in this work, it is straightforward to obtain closed-form expressions for the gradients by applying elementary calculus results such as the chain rule.

As n and m are increased, convergence of $(B_{n,m})$ and $(L_{n,m})$ implies that uniform boundedness and equicontinuity, respectively, hold in an a.e. sense. In other words, the criteria of Assumption 1 hold in an a.e. sense. This is demonstrated in the numerical examples of Section 8 where we use the function `gaussian_kde` within the subpackage `stats` of the library `scipy` to estimate the densities and subsequently use the `gradient` function within the library `numpy` to estimate the derivatives of these estimated densities. In Section 7, we discuss one potential way to utilize the computations leading to (divergent) sequences of $(B_{n,m})$ and $(L_{n,m})$ to help improve accuracy in estimated densities when singularities are present, but we leave further in-depth investigation to future work.

6.2.2 A Conceptual Example and Assumption 2

From a measure-theoretic perspective, the predictability assumption (i.e., Assumption 2) ensures that the observed measure (defined by the observed density) is absolutely continuous with respect to the push-forward measures (defined by the push-forward densities) that are obtained using either the exact or approximate maps. Absolute continuity of measures is equivalent to stating the existence of Radon-Nikodym derivatives (i.e., densities). When this assumption holds, the updated measure and density exist and take the form given in Eq. (9). Since $\pi_{\mathbf{\Lambda}}^{\text{up}}$ is a density on $\mathbf{\Lambda}$, its integral over $\mathbf{\Lambda}$ is equal to 1, which is rewritten as

$$1 = \int_{\mathbf{\Lambda}} \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda)\, d\mu_{\mathbf{\Lambda}} = \int_{\mathbf{\Lambda}} \pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda) r(Q(\lambda))\, d\mu_{\mathbf{\Lambda}} = \int_{\mathbf{\Lambda}} r(Q(\lambda))\, dP_{\mathbf{\Lambda}}^{\text{init}} = \mathbb{E}_{\text{init}}(r(Q(\lambda))).$$

Here, $\mathbb{E}_{\text{init}}$ indicates the expected value of a random variable with respect to the initial density. To monitor if Assumption 2 is violated by any of the approximate maps, we compute Monte Carlo estimates of the value of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ using the m iid samples from the initial density used to construct the estimated push-forward densities. This provides a very cheap and useful diagnostic tool since values that are not close to unity indicate that Assumption 2 is not satisfied (see [5] for details). In particular, in Section 7.3, we build intuition about this diagnostic tool in the context of a simple example with a singularity in the push-forward density so that values of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ are influenced by both numerical errors and, in some cases, violations of the predictability assumption.

7. AN ALMOST EXAMPLE AND THE ASSUMPTIONS

Here, we discuss the permissiveness of Assumption 1 in the context of a simple example that highlights and builds intuition about several theoretical and computational points.

Let $\mathbf{\Lambda} = [-1, 1]$. Suppose the exact QoI map is given by $Q(\lambda) = \lambda^5$ and the approximate QoI maps, $(Q_n(\lambda))$, are given by piecewise-linear interpolating splines using $n+2$ knots with the kth knot given by $-1+[2/(n+1)](k-1)$. In other words, n refers to the number of regularly spaced interior points at which we evaluate the exact map to construct the approximate map. It is clear that $\mathcal{D} = [-1, 1]$ and $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$ for any $1 \le p \le \infty$. While we use simple spline constructions for illustrative purposes here, we direct the interested reader to [30–32] for more information on recent advances of spline-based surrogates.

Now, suppose that the initial density is given by a uniform distribution on $\mathbf{\Lambda}$. In this case, the exact push-forward density associated with the exact QoI map is given by

$$\pi_{\mathcal{D}}^{Q}(q) = \frac{1}{10} q^{-4/5},$$

which is neither continuous on $\mathcal{D}$ nor in $L^\infty(\mathcal{D})$. In fact, $\pi_{\mathcal{D}}^{Q} \notin L^r(\mathcal{D})$ for any $r \ge 5/4$. The approximate push-forward densities are defined by a sequence of simple functions since they are defined by mapping a uniform density through piecewise-linear 1-to-1 maps. When n is odd, a straightforward computation shows that the nth approximate

push-forward density is equal to the constant $(n+1)^4/2^5$ on the interval $\left(-2^5/(n+1)^5, 2^5/(n+1)^5\right) \subset \mathcal{D}$. It follows that there are subsequential pointwise limits of these approximate push-forward densities evaluated at or near $q = 0$ that either tend to infinity or can be made arbitrarily large.

While the features of the push-forward densities discussed above may seem problematic, there is no issue in applying the theory of this work to this problem. To see this, note that for any $\epsilon > 0$, there exists a uniform bound for the family of densities on $\mathcal{D}\backslash(-\epsilon/2, \epsilon/2)$. Moreover, since the simple function approximate densities are bounded above on $\mathcal{D}\backslash(-\epsilon/2, \epsilon/2)$ and converge at a.e. $q \in \mathcal{D}\backslash(-\epsilon/2, \epsilon/2)$, they converge almost uniformly and are subsequently almost a.e.c. on $\mathcal{D}$.

7.1 The Permissiveness of Assumptions

The above example illustrates the generality of the theory developed in this work for forward UQ problems. Specifically, having both criteria of Assumption 1 hold in an almost sense is rather permissive since the theory applies to problems with push-forward densities containing a countably infinite number of singularities. Thus, under what we refer to as "normal problem conditions" where initial densities are nonsingular and the QoI map possesses only a finite number of critical points (typically corresponding to the number of singularities in the push-forward density as seen in the above example), we expect Assumption 1 to hold. However, a theoretical guarantee of convergence is not necessarily observed by the estimated densities if a computational budget restricts the quality of the estimates. To ensure that computational estimates based on finite sampling are even remotely accurate, we consider a more practical restriction on the class of forward UQ problems for which we apply the theory. Specifically, we restrict ourselves to problems that involve push-forward densities containing relatively small numbers of singularities. This can generally be assured, for example, by restricting the class of initial densities to be nonsingular and considering QoI maps that have finite numbers of critical points over the parameter space.

7.2 Computational Approach for Verifying the Almost Criteria of Assumption 1

The locations of critical points of the maps and corresponding singularities in push-forward densities are typically not known *a priori*. Yet, it is our ability to sufficiently sample around these points in the parameter and data spaces that most impact the pointwise accuracy of the estimated densities since most density estimation techniques (and especially Gaussian KDE) oversmooth "peaks" in densities. Estimating locations of potential singularities may allow for significant improvement in estimated density accuracy over all of $\mathcal{D}$ using alternative sampling techniques. While such numerical issues are not the focus of this work, we outline one computational approach to tackle this issue based on applying clustering techniques borrowed from machine learning to investigate the criteria of Assumption 1 in the context of the above example.

Representative results of the quantitative analyses discussed in Section 6.2.1 are summarized in Tables 1 and 2. In these tables, we observe the slow divergence of both the bounds and Lipschitz constants as both m and n increase. This is to be expected because the exact density, $\pi_{\mathcal{D}}^{Q}$, has a singularity, and the smoothing effect of the KDE results in slow convergence near this singularity. While we omit further numerical analysis here, we comment on one potential approach to help improve accuracy of the estimated push-forward densities by identifying the regions in parameter space associated with data singularities. First, it may be possible to use the sorting algorithms on the arrays that

TABLE 1: Representative results for $(B_{n,m})$ (see Definition 4) for the estimated densities associated with the nth approximate map constructed from m iid samples for the almost example of Section 7

m \ n	**1**	**2**	**4**	**8**	**16**
1E3	0.54	1.52	2.50	2.94	3.19
1E4	0.53	2.09	3.42	4.10	4.43
1E5	0.51	3.03	4.44	5.63	6.12

TABLE 2: Representative results for $(L_{n,m})$ (see Definition 4) for the estimated densities associated with the nth approximate map constructed from m iid samples for the almost example of Section 7

m \ n	**1**	**2**	**4**	**8**	**16**
1E3	1.41	6.08	13.68	17.98	20.81
1E4	2.19	13.89	27.26	38.35	44.06
1E5	3.48	33.14	46.45	75.08	89.42

returned the bounds and Lipschitz constants to identify points in data space where potential singularities are nearby. Subsequently, applying clustering algorithms on the corresponding samples in the parameter space, it may be possible to identify the regions in parameter space that should be sampled more extensively to improve the accuracy of the push-forward density estimates.

7.3 Assumption 2 and the Diagnostic $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$

We now give some intuition about the range of values we typically encounter for the diagnostic tool described in Section 6.2.2 using the simple example presented at the start of Section 7. We use three different observed densities, which we refer to as

$$\text{Case I: } \pi_{\mathcal{D}} \sim N(0.50, 0.1^2), \quad \text{Case II: } \pi_{\mathcal{D}} \sim N(0.25, 0.1^2), \quad \text{and} \quad \text{Case III: } \pi_{\mathcal{D}} \sim N(1, 0.1^2).$$

These cases demonstrate the variation in values we expect for $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ under different scenarios related to Assumption 2 and the magnitude of errors present in the estimated density relative to the effective support of the observed density. In all cases, the effective support may be interpreted being within three standard deviations of the means of the observed densities. Below, we refer to the plots in Fig. 1 for a visual reference of these cases that demonstrate the extent to which either Assumption 2 is violated or where significant errors in the estimated densities occur. For simplicity in each of these plots, we only plot the exact push-forward density for the exact map, its estimation with $m = 1\text{E}5$ iid samples, and the different observed densities.

In Case I (see the left plot in Fig. 1), Assumption 2 "effectively" holds (relative to the effective support of the observed density) and the estimated density is an accurate estimate of the associated push-forward density over the effective support of the observed density. Case II (see the middle plot in Fig. 1) is similar to Case I except that significant errors in the estimated density are now present in the effective support of the observed density. In Case III (see the right plot in Fig. 1), Assumption 2 is violated since significant portions of the effective support of the observed density are not "predicted" by any of the push-forward densities or their estimates.

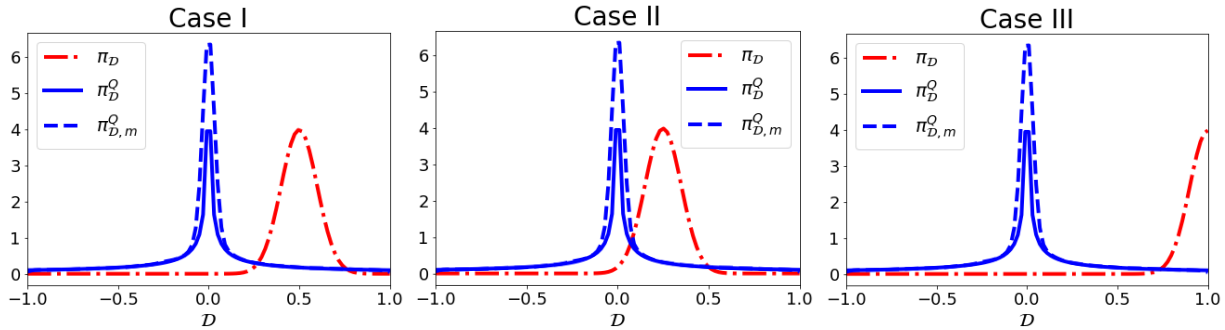

FIG. 1: Observed densities (red dash-dotted curves) used in Cases I (left), II (middle), and III (right). The solid blue curve is the exact push-forward for the exact map. The dashed blue curve is the estimated push-forward for this map using $m = 1\text{E}5$ iid samples. The errors in the estimated density for both this map and all approximate maps are primarily restricted to a narrow region around $q = 0$, which is in the effective support of the observed density in Case II. The effective support for the observed density in Case III extends beyond the range predicted by the exact push-forward density, which is a violation of Assumption 2.

In Table 3, we summarize the estimated values of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ for $n = 1, 2, 4, 8$, and 16. We see that in Case I, the values are all close to 1, while in Case II, there is some deviation away from 1 that is due to errors in the estimated push-forward densities present in the effective support of the observed density. However, in Case III, we see significant deviations away from 1 that are *primarily* due to the violation of Assumption 2. In fact, if we used the exact push-forward densities associated with each approximate map, then these values would be very close to 0.5 because we are missing half the support of the observed density. The reason for the values being slightly higher than 0.5 is due to the kernel estimate "extending" the effective support of the push-forward densities beyond the actual data space defined by $[-1, 1]$.

The takeaway is this: monitoring $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ is a good diagnostic for determining if either Assumption 2 is violated or if significant errors are present in the associated estimated push-forward density (and subsequently if errors are present in the associated estimate of the *approximate* updated density). While there are certainly exceptions, we generally find large deviations of this expected value away from 1 are due to violations of the assumption while smaller deviations away from 1 are often due to numerical errors in the estimated density. Finally, we note that just because estimated values of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ are near 1 does *not* mean that the estimated push-forward density or associated estimate of the updated density are accurate approximations for the *exact* push-forward density or *exact* updated density one would obtain using the *exact* QoI map. A "good value" of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ (i.e., a value near 1) simply speaks to both Assumption 2 not being violated and also to the relative accuracy of the estimated push-forward density with respect to its nonestimated counterpart.

8. NUMERICAL EXAMPLES

Surrogates of the QoI map can significantly reduce the computational cost of solving both forward and inverse UQ problems. The popularity of using stochastic spectral methods, and polynomial chaos expansions (PCEs) in particular, to approximate QoI maps arising in UQ problems dates back several decades; e.g., see [7,33–36]. The convergence theory of these methods dates back nearly a century starting with the homogeneous chaos introduced by Wiener [37] with further analysis by Cameron and Martin [38]. However, the interested reader may find the far more recent reference [7] to be a more accessible introduction to this topic, and we simply note here that PCEs are understood to converge in an L^2 sense. While the theory we developed applies to any sequence of approximate models that converge in an L^p sense, we focus on numerical examples using PCEs given their prevalence in the literature. While many other surrogate techniques exist that produce sequences of approximate models that converge in an L^p sense, we offer some specific remarks about Gaussian process emulation (sometimes referred to as Gaussian process regression) as it has become a more popular approach in recent years. This popularity is driven by multiple factors including the relationship between Gaussian processes and neural networks (e.g., see [39]) and the ability to "learn" these maps directly from noisy data of the underlying "true" response, which represents a type of "model free" approach that differs from most PCE-based approaches. Recent studies show that Gaussian process emulation has desirable convergence properties in L^2 [40] (or, more strongly in a H^β Sobolev space where β depends on the smoothness of the underlying response [41]), which immediately implies that the theory in this work also applies to Gaussian process emulation under certain constraints involving the continuity or smoothness of the exact QoI map. The interested reader should refer to [42] (and the references therein) for a comparison of PCE to Gaussian process emulation performance in solving UQ problems in computationally expensive problems. The main conclusion in [42] is that neither method

TABLE 3: Representative estimates of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ with $m = 1\text{E}4$ iid samples used for both density estimation of associated push-forward densities and Monte Carlo estimate of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ for the cases discussed in Section 7.3

Case \ n	1	2	4	8	16
I	1.00	1.00	1.00	0.98	0.99
II	0.99	0.88	0.85	0.92	0.92
III	0.69	0.66	0.63	0.61	0.61

unanimously outperforms the other. In general, the best choice of surrogate technique depends on many specific factors of the problem such as the dimension, regularity, and any potential correlation between inputs.

Two numerical examples are considered showing application of the theory to commonly studied ordinary and partial differential equation models. Moreover, we consider separate numerical approaches for generating the PCE approximations based on Galerkin projections in each example. The PCEs for the first example are extensively studied in [7] using an intrusive approach, which we choose to employ here for one particular PCE. This allows the interested reader to modify the PCE used here by following the steps outlined in [7]. The PCEs for the second example are obtained using a nonintrusive pseudo-spectral approach that exploits quadrature methods for accurate construction of the coefficients in the PCEs. For more information on these approaches, we direct the interested reader to [6,7,43] for traditional intrusive approaches, to [44,45] for nonintrusive approaches, and to [46] for a comparison of such different approaches. Moreover, Appendix C describes how to obtain and utilize the scripts and data sets to re-create all of the figures and tables presented below.

In both examples, we use standard Gaussian kernel density estimation (KDE) to approximate the push-forward density functions, which are subsequently used to estimate the errors in both the push-forward and updated densities. As noted in Section 6.1, the impact of the additional error arising from the use of KDEs is analyzed in [12] when the maps converge in L^∞, but we leave the modifications of that analysis for L^p-norms to future work.

Before we present the numerical examples, some final remarks are worth noting.

1. While it is possible to form PCEs with respect to distributions unrelated to the initial distribution, this is in general not considered optimal (e.g., see [36] where such inefficiencies are explored in a Bayesian setting). Thus, for both simplicity in computations and presentation, we choose to form the PCEs with respect to the initial distribution assumed for the parameters.

2. The initial distributions considered in the examples assume independent input parameters to simplify the PCE construction, but this is not a necessary assumption. While several methods exist for constructing PCEs of dependent variables, a novel Gram-Schmidt orthogonalization method studied in [47] is shown to produce PCEs of dependent variables that are orders of magnitude more accurate than other methods.

3. The convergence of PCEs is in an L^2 sense with respect to the probability measure of the random variable used to define the orthogonal polynomials in the expansion. Thus, we consider/interpret the initial distribution as the dominating measure in this case.

4. Finally, the convergence results guaranteed in Theorem 1 require compact $\mathcal{D}$. However, we may still observe these convergence results in practice when $\mathcal{D}$ is not compact if there is a compact "computational support" for the push-forward and observed densities. By "computational support" we mean the set that contains all typical random samples generated in practice by the various distributions. This is discussed more in the second example where a noncompact $\mathcal{D}$ is considered yet the convergence results are still observed. Moreover, we provide a variant of the first example in the scripts and data sets provided with this work (see Appendix C) that also demonstrates the convergence results holding when $\mathcal{D}$ is not compact yet has a compact "computational support."

8.1 ODE Example

Consider the ordinary differential equation

$$\frac{dy(t)}{dt} = -\lambda y, \qquad y(0) = 1, \tag{16}$$

where the decay rate coefficient λ is treated as a random variable. The QoI is taken as $Q(\lambda) = y(0.5, \lambda)$ where $y(t, \lambda) = e^{-\lambda t}$ is the exact solution of Eq. (16), which has a clear dependence on the parameter value. For this example, $\mathbf{\Lambda} = [-1, 1]$ and $\mathcal{D} = [e^{-0.5}, e^{0.5}]$.

8.1.1 Convergence of Forward Problem

Assume $\pi_{\mathbf{\Lambda}}^{\text{init}} \sim U([-1,1])$. Following [7], the PCE of the solution $y(t,\lambda)$ is given by

$$y(t,\lambda) = \sum_{i=0}^{\infty} y_i(t)\Phi_i(\lambda),$$

where y_i denotes the ith coefficient of the expansion of $y(t,\lambda)$, and $\{\Phi_i\}$ denotes a complete orthogonal polynomial basis from the Askey scheme associated with a choice of random variable for Φ_i. Since the initial distribution of λ is assumed uniform, we choose $\{\Phi_i\}$ as the Legendre polynomials and put the uniform distributed random variable in Φ_i for optimal convergence. The approximate maps are then defined by

$$Q_n(\lambda) = \sum_{i=0}^{n} y_i(0.5)\Phi_i(\lambda), \tag{17}$$

where n denotes the highest-order polynomial in the expansion used to approximate the map. Using the intrusive approach detailed in [7], a system of ordinary differential equations is formed and solved for each n to obtain the coefficients $y_i(t,\lambda)$ for $0 \le i \le n$, which are then evaluated at $t = 0.5$ to obtain $Q_n(\lambda)$. Reference results are obtained using the exact map defined by $Q(\lambda) = e^{-0.5\lambda}$ evaluated on the same set of parameter samples as the approximate maps to construct estimates of the exact push-forward and updated densities.

To numerically verify Assumption 1, we follow the approach discussed in Section 6.2.1 to compute sequences of bounds and Lipschitz constants. Tables 4 and 5 indicate convergence of both the computed bounds and Lipschitz constants as both m and n increase. This suggests that both criteria of Assumption 1 hold in the stronger a.e. sense.

We now observe the convergence guaranteed by Lemma 1 and Theorem 1. First, a single initial set of 1E4 iid parameter samples drawn from the initial distribution is generated. Then, the approximate push-forward densities are estimated using a standard Gaussian KDE applied to the approximate map evaluations of these parameter samples. Since $\mathcal{D}$ is compact, we compute Monte Carlo estimates of $\|\pi_{\mathcal{D}}^{Q}(q) - \pi_{\mathcal{D}}^{Q_n}(q)\|_{L^r}$ for various $r \ge 1$ using a fixed set of 1E4 iid uniform random samples of $\mathcal{D}$. The left plot in Fig. 2 shows the corresponding error plots for $r = 1, 2, \ldots, 5$ as functions of the approximate map number. These error plots indicate an exponential convergence of the approximate push-forward densities for $r \ge 1$, which verifies the convergence in Lemma 1. We next estimate $\|\pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))\|_{L^2(\Lambda)}$ using Monte Carlo estimates for each of the $n = 1, 2, \ldots, 5$ approximate maps

TABLE 4: Representative results for $(B_{n,m})$ (see Definition 4) for the estimated densities associated with the nth approximate map constructed from m iid samples for the example of Section 8.1

m \ n	**1**	**2**	**3**	**4**	**5**
1E3	1.06	1.28	1.24	1.24	1.24
1E4	1.03	1.49	1.42	1.42	1.42
1E5	1.00	1.61	1.49	1.50	1.50

TABLE 5: Representative results for $(L_{n,m})$ (see Definition 4) for the estimated densities associated with the nth approximate map constructed from m iid samples for the example of Section 8.1

m \ n	**1**	**2**	**3**	**4**	**5**
1E3	5.58	8.18	7.73	7.76	7.76
1E4	8.40	14.18	12.97	13.05	13.05
1E5	13.45	23.43	21.00	21.22	21.21

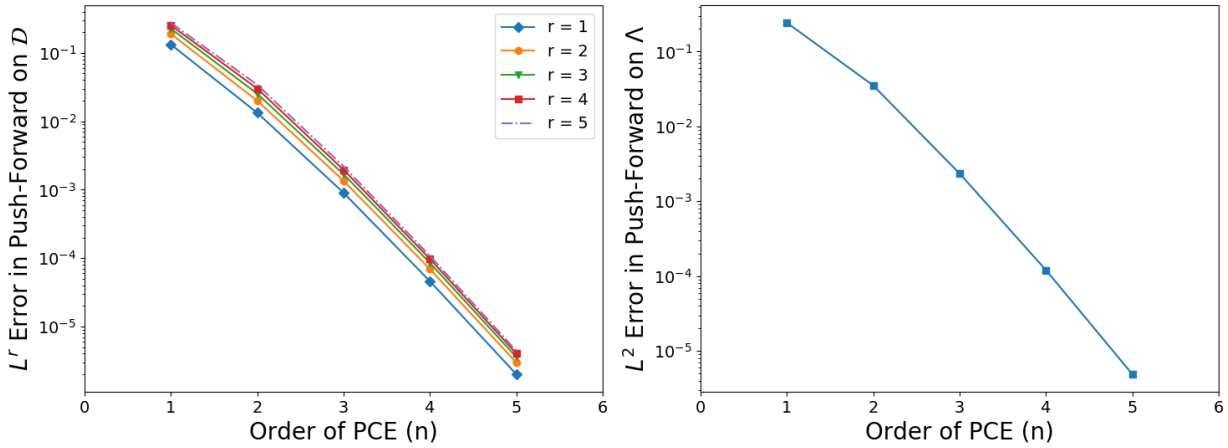

FIG. 2: Convergence of push-forward densities for ODE example. Left: L^r error plot of the push-forward densities on $\mathcal{D}$ for $r = 1, 2, \ldots, 5$. Right: L^2 error plot of the push-forward densities on $\mathbf{\Lambda}$.

and the same 1E4 iid samples drawn from the initial distribution as were used to construct the push-forwards. These errors are summarized in the right plot of Fig. 2 as a function of the approximate map number, which verifies the convergence in Theorem 1.

8.1.2 Convergence of Inverse Problem

We now observe the convergence guaranteed in Theorem 3. Using the same initial distribution as above, we assume the observed density is given by a $Beta(4,4)$ distribution on $[1, 1.25] \subset \mathcal{D}$. We use the same approximate maps and approximate push-forward densities as above. Before we analyze the convergence, we first numerically verify Assumption 2 using the expected value of the ratio, $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$, discussed in Section 6.2.2. Since all of the $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ estimates are close to 1 from Table 6, it appears there is no violation of Assumption 2 for these maps.

Figure 3 illustrates the error in the updated densities associated with the nth approximate map. Here, the error is given by $\|\pi_\Lambda^{u,n}(\lambda) - \pi_\Lambda^u(\lambda)\|_{L^2(\Lambda)}$, which is again estimated using Monte Carlo integration with the same 1E4 samples used to construct the push-forward densities. The plot demonstrates the L^2 convergence of the updated densities.

8.2 PDE Example

Consider the following PDE,

$$\begin{cases} -\nabla \cdot (A\nabla u) = (e^{\lambda_1}\lambda_1^2\pi^2 + e^{\lambda_2}\lambda_2^2\pi^2)u, & \text{in } \Omega \\ u = 0, & \text{on } \Gamma_0 \\ (A\nabla u)\cdot n = -e^{\lambda_2}\lambda_2\pi\sin(\lambda_1\pi x)\sin(\lambda_2\pi y), & \text{on } \Gamma_1, \\ (A\nabla u)\cdot n = e^{\lambda_2}\lambda_2\pi\sin(\lambda_1\pi x)\sin(\lambda_2\pi y), & \text{on } \Gamma_2 \\ (A\nabla u)\cdot n = e^{\lambda_1}\lambda_1\pi\cos(\lambda_1\pi x)\cos(\lambda_2\pi y), & \text{on } \Gamma_3 \end{cases} \tag{18}$$

where

TABLE 6: Representative estimates of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ with $m =$ 1E4 iid samples used for both density estimation of associated push-forward densities and Monte Carlo estimate of $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ for the example of Section 8.1

n	1	2	3	4	5
$\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$	1.003	0.979	0.978	0.978	0.978

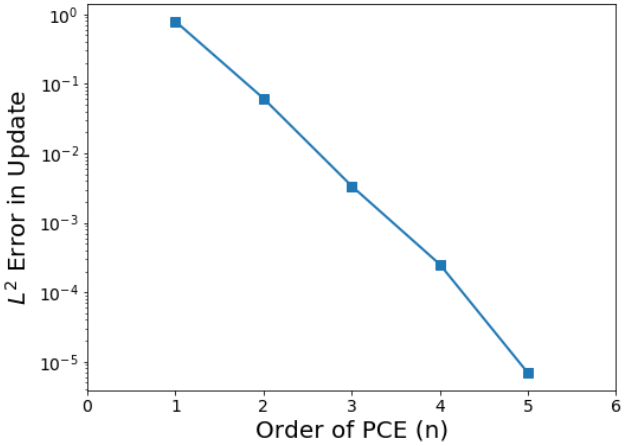

FIG. 3: Convergence of updated densities for ODE example in L^2

$$A = \begin{bmatrix} e^{\lambda_1} & 0 \\ 0 & e^{\lambda_2} \end{bmatrix}.$$

$\Gamma_0 = \{(x,y) : x = 0, 0 \le y \le 1\}$, $\Gamma_1 = \{(x,y) : y = 1, 0 \le x \le 1\}$, $\Gamma_2 = \{(x,y) : y = 0, 0 \le x \le 1\}$, $\Gamma_3 = \{(x,y) : x = 1, 0 \le y \le 1\}$, and $\Omega = (0,1) \times (0,1)$.

The QoI is the average value of the solution u in region $[a,b] \times [c,d]$,

$$Q(\lambda_1, \lambda_2) = \frac{1}{(b-a)(d-c)} \int_c^d \int_a^b u(x, y; \lambda_1, \lambda_2)\, dx\, dy, \tag{19}$$

where, in this example, we know $u(x, y; \lambda_1, \lambda_2) = \sin(\lambda_1 \pi x) \cos(\lambda_2 \pi y)$ is the exact solution of Eq. (18) associated with a particular sample $(\lambda_1, \lambda_2) \in \mathbf{\Lambda}$, and we choose $a = c = 0.4$, $b = d = 0.6$. For this example, $\mathbf{\Lambda} = (-\infty, \infty) \times (-\infty, \infty)$ and $\mathcal{D} = (-\infty, \infty)$.

8.2.1 Convergence of Forward Problem

As before, the goal of this section is to observe the convergence results of Lemma 1 and Theorem 1. Assume $\lambda_1 \sim N(\mu_1, \sigma_1^2)$ and $\lambda_2 \sim N(\mu_2, \sigma_2^2)$ with $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 0.1$. Then, we write the exact map $Q(\lambda_1, \lambda_2)$ as follows:

$$Q(\lambda_1, \lambda_2) = \sum_{i,j=0}^{\infty} q_{ij} \Phi_{ij} \left(\frac{\lambda_1 - \mu_1}{\sigma_1}, \frac{\lambda_2 - \mu_2}{\sigma_2} \right),$$

where $\{\Phi_{ij}\}$ (for $0 \le i, j < \infty$) denotes a complete 2-d orthogonal Hermite polynomial basis, $\{q_{i,j}\}$ (for $0 \le i, j < \infty$) denotes the corresponding coefficients of this PCE, and the equality is understood to hold in L^2. Here, instead of formulating and solving a system of PDEs to compute the coefficients $\{q_{ij}\}$, we apply a nonintrusive pseudo-spectral approach to approximate the $\{q_{ij}\}$. Specifically, starting with the PCE above, we compute (weighted) L^2-inner products of both sides with a fixed Φ_{ij}. Then, by exploiting the orthogonality of the polynomials, q_{ij} is given as an integral of $Q(\lambda_1, \lambda_2)$ weighted by the product of Φ_{ij} and the underlying normal distribution for which these polynomials are orthogonal (and normalized by the weighted L^2-norm of Φ_{ij}). In general, we do not expect to have $Q(\lambda_1, \lambda_2)$ available in closed form to compute this integral, so we instead turn to quadrature methods to discretize the integrals over $\mathbf{\Lambda}$ defining q_{ij} where numerical approximations of $Q(\lambda_1, \lambda_2)$ are used at each quadrature point. To this end, we use a triangulation of a 50×50 square mesh on Ω to obtain a numerical estimate of $u(x, y; \lambda_1, \lambda_2)$ and subsequently of $Q(\lambda_1, \lambda_2)$. Estimates of $Q(\lambda_1, \lambda_2)$ are obtained on a set of 400 parameter samples taken from the tensor product of 20-point Gauss-Hermite quadrature points in each dimension to ensure accuracy well beyond the degree of polynomials considered in this example. Using these computed q_{ij}, we write the approximate maps as

$$Q_n(\lambda_1, \lambda_2) = \sum_{i+j \leq n} q_{ij} \Phi_{ij}\left(\frac{\lambda_1 - \mu_1}{\sigma_1}, \frac{\lambda_2 - \mu_2}{\sigma_2}\right).$$

The reference results below are obtained by using the manufactured solution and evaluating a closed form expression for the resulting QoI on each of the parameter samples used by the approximate maps to construct estimates of the exact densities.

Before verifying assumptions and observing the convergence of densities, we remark on some interesting behavior with regards to the QoI maps. The closed-form expression for the QoI map reveals an odd function in the λ_1-direction and an even function in the λ_2-direction, which is due to the oscillatory nature of the manufactured solution and domain $[a, b] \times [c, d]$ used to compute the QoI. Subsequently, this implies the exact QoI map is orthogonal to many of the polynomials Φ_{ij}. In this example, we consider approximate maps of order $n = 1, 2, \ldots, 5$ where Φ_{ij} for any $i + j \leq n$ is given by the product $\Phi_i(\lambda_1)\Phi_j(\lambda_2)$. Thus, it is straightforward to see that the exact QoI map is orthogonal in the weighted L^2-inner product to many of the Hermite polynomials. In fact, the only (i, j)-pairs with $i + j \leq 5$ such that q_{ij} is nonzero are

$$S := \{(1, 0), (3, 0), (1, 2), (5, 0), (3, 2), (1, 4)\}.$$

This is numerically confirmed as well where the quadrature method produces estimated values for q_{ij} with $(i, j) \notin S$ that are orders of magnitude smaller than the coefficients associated with $(i, j) \in S$. Consequently, the approximate maps Q_1, Q_3, and Q_5 are all distinct approximations to Q whereas Q_2 and Q_4 are almost indistinguishable from Q_1 and Q_3, respectively, with only slight variations present due to the use of numerical quadrature and the numerical solution of the PDE used in the quadrature method. The impact of this on the convergence of densities is seen below.

As before, we first numerically verify Assumption 1 as described in Section 6.2.1 to compute sequences of bounds and Lipschitz constants. Tables 7 and 8 indicate convergence of both the computed bounds and Lipschitz constants as both m and n increase. As in the previous example, this suggests that both criteria of Assumpton 1 hold in the stronger a.e. sense.

We now observe the convergence guaranteed in Lemma 1 and Theorem 1. A single initial set of 1E4 iid parameter samples is generated and each of the approximate maps is evaluated on these parameter samples to generate different approximate QoI sample sets. Then, the push-forward densities for each approximate map are estimated using a standard Gaussian KDE. The range of output samples falls within the interval $[-1, 1]$, so we set $D_c = [-1, 1]$ and estimate $\|\pi_{\mathcal{D}}^{Q}(q) - \pi_{\mathcal{D}}^{Q_n}(q)\|_{L^r(D_c)}$ for various $r \geq 1$ using Monte Carlo estimates on a fixed set of 1E4 uniform random samples in D_c.

TABLE 7: Representative results for $(B_{n,m})$ (see Definition 4) for the estimated densities associated with the nth approximate map constructed from m iid samples for the example of Section 8.2

m \ n	**1**	**2**	**3**	**4**	**5**
1E3	2.63	2.63	2.61	2.61	2.61
1E4	2.64	2.64	2.61	2.61	2.61
1E5	2.60	2.60	2.57	2.57	2.57

TABLE 8: Representative results for $(L_{n,m})$ (see Definition 4) for the estimated densities associated with the nth approximate map constructed from m iid samples for the example of Section 8.2

m \ n	**1**	**2**	**3**	**4**	**5**
1E3	12.11	12.11	11.95	11.95	11.95
1E4	11.47	11.47	11.41	11.41	11.41
1E5	11.07	11.07	10.86	10.86	10.86

The left plot of Fig. 4 shows the corresponding error plots for $r = 1, 2, \ldots, 5$ as a function of the approximate map number. Note that the error decreases when n increases from an even to an odd integer but does not appear to change when n increases from an odd to an even integer. This is due to the symmetry of the QoI map and its orthogonality with respect to even-ordered polynomials as described above and is thus completely expected for this sequence of approximate maps. We note that $\mathcal{D} = (-\infty, \infty)$ so that the conclusions of Theorem 1 cannot be guaranteed. However, virtually all samples are generated within a compact set of $\mathcal{D}$ (in this case $[-1, 1]$), so we expect to see convergence. The set $[-1, 1]$ is the "computational support" referred to in the preamble of this section. We then estimate $\|\pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))\|_{L^2(\Lambda)}$ using Monte Carlo estimates with the same 1E4 samples used to estimate the approximate push-forwards for $n = 1, 2, \ldots, 5$, which is summarized in the right plot of Fig. 4 as a function of the approximate map number. The overall trend of decreasing errors in both plots verifies the theoretical results of Lemma 1 and Theorem 1.

8.2.2 Convergence of Inverse Problem

We now observe the convergence in Theorem 3. Using the same initial distribution as above, we assume the observed density is given by a $N(0.3, 0.1^2)$ distribution. We use the same approximate maps and approximate push-forward densities as above. Assumption 2 is again numerically verified using the expected value of the ratio, $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$, mentioned in Section 6.2.2. For each of the approximate maps, rounding $\mathbb{E}_{\text{init}}(r(Q_n(\lambda)))$ at the second decimal gives 1.00, so we omit a summary of these values in a table.

Figure 5 illustrates the error of updated densities as a function of approximate map number. Here, we again use Monte Carlo estimates of $\|\pi_{\Lambda}^{u,n}(\lambda) - \pi_{\Lambda}^{u}(\lambda)\|_{L^2(\Lambda)}$ using the same set of parameter samples used to construct the push-forwards, and we observe L^2 convergence of the updated densities.

9. CONCLUSIONS

We developed a theoretical framework for analyzing the convergence of probability density functions computed using approximate models for both forward and inverse problems. The theoretical results are quite general and apply to any L^p-convergent sequence of approximate models. This greatly extends previous work that required essentially uniform convergence of approximate models (i.e., L^∞ convergence). A simple numerical example producing a singular push-forward density is used to show the permissiveness of the two main assumptions used in this work. Moreover, these assumptions are verified in each of the main numerical examples that demonstrate the convergence results for polynomial chaos expansions, which are commonly used to build approximate models in the literature.

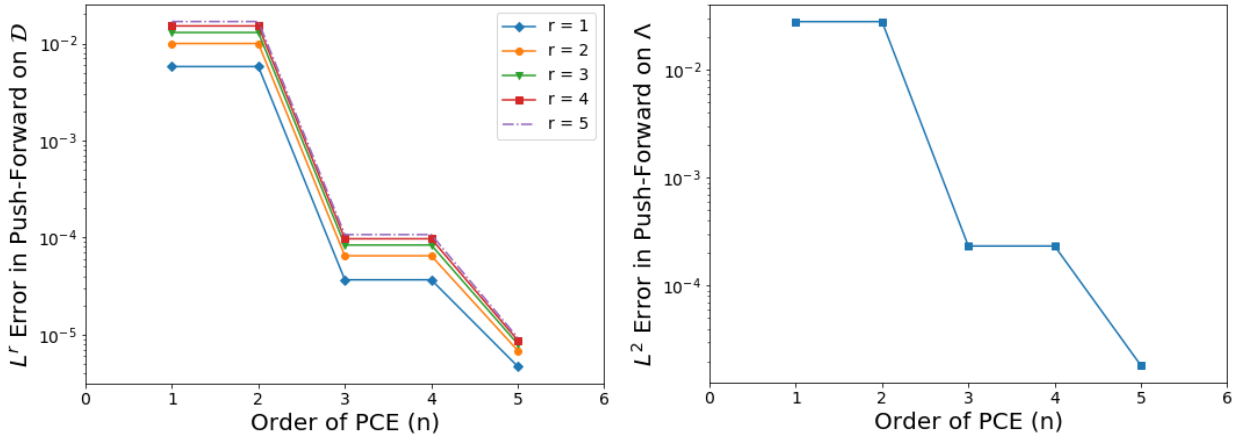

FIG. 4: Convergence of push-forward densities for PDE example. Left: L^r error plot of the push-forward densities on $\mathcal{D}$ for $r = 1, 2, \ldots, 5$. Right: L^2 error plot of the push-forward densities on $\mathbf{\Lambda}$.

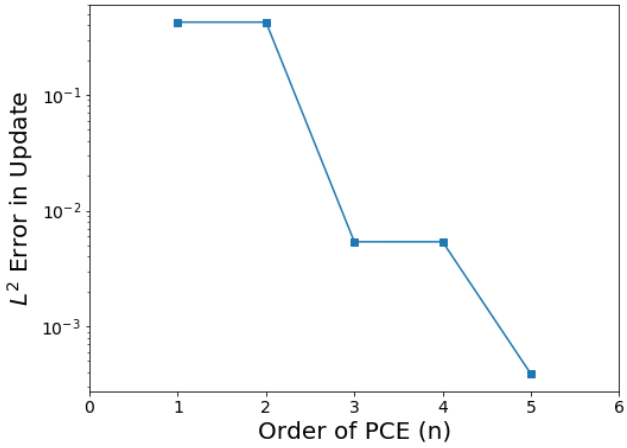

FIG. 5: Convergence of updated densities for PDE example in L^2

ACKNOWLEDGMENTS

T. Wildey's work is supported by the Office of Science Early Career Research Program. T. Butler's and W. Zhang's work is supported by the National Science Foundation under Grant No. DMS-1818941.

REFERENCES

1. Butler, T. and Hakula, H., What Do We Hear from a Drum? A Data-Consistent Approach to Quantifying Irreducible Uncertainty on Model Inputs by Extracting Information from Correlated Model Output Data, *Comput. Methods Appl. Mech. Eng.*, **370**:113228, 2020.
2. Tran, A. and Wildey, T., Solving Stochastic Inverse Problems for Property–Structure Linkages Using Data-Consistent Inversion and Machine Learning, *JOM*, **73**(1):72–89, 2021.
3. Poole, D. and Raftery, A.E., Inference for Deterministic Simulation Models: The Bayesian Melding Approach, *J. Am. Stat. Assoc.*, **95**(452):1244–1255, 2000.
4. Parikh, J., Kozloski, J., and Gurev, V., Integration of AI and Mechanistic Modeling in Generative Adversarial Networks for Stochastic Inverse Problems, *Stat. Mach. Learn.*, arXiv:2009.08267, 2020.
5. Butler, T., Jakeman, J., and Wildey, T., Combining Push-Forward Measures and Bayes' Rule to Construct Consistent Solutions to Stochastic Inverse Problems, *SIAM J. Sci. Comput.*, **40**(2):A984–A1011, 2018.
6. Ghanem, R. and Spanos, P., *Stochastic Finite Elements: A Spectral Approach*, New York: Springer Verlag, 2002.
7. Xiu, D. and Karniadakis, G., The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations, *SIAM J. Sci. Comput.*, **24**:619–644, 2002.
8. Barthelmann, V., Novak, E., and Ritter, K., High Dimensional Polynomial Interpolation on Sparse Grids, *Adv. Comput. Math.*, **12**:273–288, 2000.
9. Ma, X. and Zabaras, N., An Adaptive Hierarchical Sparse Grid Collocation Algorithm for the Solution of Stochastic Differential Equations, *J. Comput. Phys.*, **228**(8):3084–3113, 2009.
10. Scheuerer, M., Schaback, R., and Schlather, M., Interpolation of Spatial Data—A Stochastic or a Deterministic Problem?, *Eur. J. Appl. Math.*, **24**(4):601–629, 2013.
11. Rasmussen, C.E. and Williams, C.K.I., *Gaussian Processes for Machine Learning*, Cambridge, MA: The MIT Press, 2006.
12. Butler, T., Jakeman, J., and Wildey, T., Convergence of Probability Densities Using Approximate Models for Forward and Inverse Problems in Uncertainty Quantification, *SIAM J. Sci. Comput.*, **40**(5):A3523–A3548, 2018.
13. Folland, G.B., *Real Analysis: Modern Techniques and Their Applications*, New York: Wiley, 1999.
14. Bogachev, V., *Measure Theory*, Vol. 1, Berlin Heidelberg: Springer-Verlag 2007.
15. Sweeting, T.J., On a Converse to Scheffe's Theorem, *Ann. Stat.*, **14**(3):1252–1256, 1986.

16. Bogachev, V., *Measure Theory*, Vol. 2, Berlin Heidelberg: Springer-Verlag, 2007.
17. Stuart, A.M., Inverse Problems: A Bayesian Perspective, *Acta Numer.*, **19**:451–559, 2010.
18. Kaipio, J. and Somersalo, E., Statistical Inverse Problems: Discretization, Model Reduction and Inverse Crimes, *J. Comput. Appl. Math.*, **198**(2):493–504, 2007,
19. Berger, J.O., Moreno, E., Pericchi, L.R., Bayarri, M.J., Bernardo, J.M., Cano, J.A., De la Horra, J., Martín, J., Ríos-Insúa, D., Betrò, B., Dasgupta, A., Gustafson, P., Wasserman, L., Kadane, J.B., Srinivasan, C., Lavine, M., O'Hagan, A., Polasek, W., Robert, C.P., Goutis, C., Ruggeri, F., Salinetti, G., and Sivaganesan, S., An Overview of Robust Bayesian Analysis, *Test*, **3**(1):5–124, 1994.
20. Bernardo, J.M. and Adrian, F.M., *Bayesian Theory*, New York: Wiley, 1994.
21. Robert, C.P., *The Bayesian Choice—A Decision Theoretic Motivation*, 2nd ed., Berlin: Springer, 2001.
22. Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B., *Bayesian Data Analysis*, 3rd ed., Boca Raton, FL: Chapman and Hall/CRC, 2013.
23. Jaynes, E.T., *Probability Theory: The Logic of Science*, Cambridge, UK: Cambridge University Press, 1998.
24. Dellacherie, C. and Meyer, P., *Probabilities and Potential*, Amsterdam, The Netherlands: North-Holland Publishing Co., 1978.
25. Capodaglio, G., Gunzburger, M., and Wynn, H.P., Approximation of Probability Density Functions for PDEs with Random Parameters Using Truncated Series Expansions, *Vietnam J. Math.*, **49**:685–711, 2021.
26. Zech, J. and Marzouk, Y., Sparse Approximation of Triangular Transports on Bounded Domains, *Math. Stat. Theory*, arXiv:2107.13422, 2020.
27. Ditkowski, A., Fibich, G., and Sagiv, A., Density Estimation in Uncertainty Propagation Problems Using a Surrogate Model, *SIAM/ASA J. Uncertainty Quantif.*, **8**(1):261–300, 2020.
28. Hansen, B.E., Uniform Convergence Rates for Kernel Estimation with Dependent Data, *Econ. Theor.*, **24**(3):726–748, 2008.
29. Devroye, L.P. and Wagner, T.J., The L_1 Convergence of Kernel Density Estimates, *Ann. Stat.*, **7**(5):1136–1139, 1979.
30. Rahman, S., A Spline Chaos Expansion, *SIAM/ASA J. Uncertainty Quantif.*, **8**(1):27–57, 2020.
31. Beck, J., Tamellini, L., and Tempone, R., IGA-Based Multi-Index Stochastic Collocation for Random PDEs on Arbitrary Domains, *Comput. Methods Appl. Mech. Eng.*, **351**:330–350, 2019.
32. Piazzola, C., Tamellini, L., Pellegrini, R., Broglia, R., Serani, A., and Diez, M., Uncertainty Quantification of Ship Resistance via Multi-Index Stochastic Collocation and Radial Basis Function Surrogates: A Comparison, *AIAA Aviation 2020 Forum*, Virtual, 2020.
33. Ghanem, R. and Red-Horse, J., Propagation of Probabilistic Uncertainty in Complex Physical Systems Using a Stochastic Finite Element Approach, *Phys. D: Nonlinear Phenom.*, **133**:137–144, 1999.
34. Le Maître, O., Ghanem, R., Knio, O., and Najm, H., Uncertainty Propagation Using Wiener-Haar Expansions, *J. Comput. Phys.*, **197**(1):28–57, 2004.
35. Wan, X. and Karniadakis, G., Beyond Wiener-Askey Expansions: Handling Arbitrary PDFs, *J. Sci. Comput.*, **27**:455–464, 2006.
36. Marzouk, Y., Najm, H., and Rahn, L., Stochastic Spectral Methods for Efficient Bayesian Solution of Inverse Problems, *J. Comput. Phys.*, **224**:560–586, 2007.
37. Wiener, N., The Homogeneous Chaos, *Am. J. Math.*, **60**:897–936, 1938.
38. Cameron, R. and Martin, W., The Orthogonal Development of Non-Linear Functionals in Series of Fourier-Hermite Functionals, *Ann. Math.*, **48**:385–392, 1947.
39. Lee, J., Bahri, Y., Novak, R., Schoenholz, S.S., Pennington, J., and Sohl-Dickstein, J., Deep Neural Networks as Gaussian Processes, *Stat. Mach. Learn.*, arXiv:1711.00165, 2018.
40. Wang, W. and Jing, B.Y., Convergence of Gaussian Process Regression: Optimality, Robustness, and Relationship with Kernel Ridge Regression, *Math. Stat. Theory*, arXiv:2104.09778, 2021.
41. Teckentrup, A.L., Convergence of Gaussian Process Regression with Estimated Hyper-Parameters and Applications in Bayesian Inverse Problems, *Math. Numer. Anal.*, arXiv:1909.00232, 2020.
42. Owen, N.E., Challenor, P., Menon, P.P., and Bennani, S., Comparison of Surrogate-Based Uncertainty Quantification Methods for Computationally Expensive Simulators, *SIAM/ASA J. Uncertainty Quantif.*, **5**(1):403–435, 2017.

43. Deb, M., Babuška, I., and Oden, J., Solution of Stochastic Partial Differential Equations Using Galerkin Finite Element Techniques, *Comput. Methods Appl. Mech. Eng.*, **190**:6359–6372, 2001.

44. Reagan, M.T., Najm, H.N., Ghanem, R.G., and Knio, O.M., Uncertainty Quantification in Reacting-Flow Simulations through Non-Intrusive Spectral Projection, *Combust. Flame*, **132**(3):545–555, 2003.

45. Acharjee, S. and Zabaras, N., A Non-Intrusive Stochastic Galerkin Approach for Modeling Uncertainty Propagation in Deformation Processes, *Comput. Struct.*, **85**(5-6):244–254, 2007.

46. Constantine, P.G., Gleich, D.F., and Iaccarino, G., Spectral Methods for Parameterized Matrix Equations, *SIAM J. Matrix Anal. Appl.*, **31**:2681–2699, 2010.

47. Jakeman, J.D., Franzelin, F., Narayan, A., Eldred, M., and Plfüger, D., Polynomial Chaos Expansions for Dependent Random Variables, *Comput. Methods Appl. Mech. Eng.*, **351**:643–666, 2019.

48. Adams, R.A., *Sobolev Spaces*, Cambridge, MA: Academic Press, 1975.

49. Harris, C.R., Millman, K.J., van der Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M.H., Brett, M., Haldane, A., del R'ıo, J.F., Wiebe, M., Peterson, P., G'erard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T.E., Array Programming with NumPy, *Nature*, **585**(7825):357–362, 2020.

50. Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S.J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, L., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., van Mulbregt, P., and Contributors, S., SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, **17**:261–272, 2020.

51. Hunter, J.D., Matplotlib: A 2D Graphics Environment, *Comput. Sci. Eng.*, **9**(3):90–95, 2007.

52. Logg, A. and Wells, G.N., DOLFIN: Automated Finite Element Computing, *ACM Trans. Math. Software*, **37**:1–20, 2011.

APPENDIX A. PROOFS FOR FORWARD PROBLEM ANALYSIS

APPENDIX A.1 Proof of Lemma 1

Before we prove this lemma, we recall two useful results from measure theory. First, if X is a Euclidean space and $A \in \mathcal{B}_X$ with $\mu_X(A) < \infty$, then, for any $\epsilon > 0$, there exists an open $G \supset A$ and a compact $K \subset A$ such that $\mu_X(G \backslash K) < \epsilon$. In other words, any set of finite measure can be approximated arbitrarily well (in measure) by either an open set containing it or a compact set contained within it. Thus, without loss of generality, the set where the uniform bounded and a.e.c. criteria in Assumption 1 do *not* hold can be chosen as an open set. This is done in the proof below where this set is denoted by N_δ. Second, the general form of Lusin's Theorem[**] applies to the measure spaces considered in this work. In the context of this work, this implies that all the densities considered in this work are continuous functions with compact support in an almost sense. Subsequently, densities defined on either $\mathbf{\Lambda}$ or $\mathcal{D}$ are in L^r for $1 \leq r \leq \infty$ in an almost sense.

Proof. Since $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$, Q_n converges weakly to Q.[††] This along with Assumption 1 implies that $\pi_{\mathcal{D}}^{Q_n}$ converges to $\pi_{\mathcal{D}}^{Q}$ in an almost sense using Theorem 1 from [15].[‡‡] in an almost sense. This proves (1).

[**] The classical version of Lusin's theorem implies that measurable complex-valued functions defined on $[a, b]$ are almost continuous. The general form of Lusin's theorem extends this result to measurable functions defined on any Radon measure space that map into a second-countable topological space. The interested reader should refer to Section 7.14(ix) in [16] for a thorough treatise on generalizations of Lusin's theorem.

[††] Weak convergence of random variables is also called convergence in distribution as it implies that the corresponding distribution functions associated with the random variables converge at all continuity points of the limit distribution function. For a concise measure-theoretic summary of this connection between general measurable functions and their induced distribution functions, we direct the interested reader to Section 6.4 of [13].

[‡‡] Theorem 1 from [15] states that for a sequence of distributions (G_n) and corresponding densities (g_n) the following two statements are equivalent. (1) (g_n) is a.e.c. and bounded, and (G_n) converges to a distribution G. (2) $g_n \to g$ pointwise, uniformly on compact subsets of their domain, and g is the continuous density of the distribution G. In the context of this work, $Q_n \to Q$ weakly implies the push-forward distributions converge and Assumption 1 then implies the associated push-forward densities converge according to this result.

Let $\delta > 0$ and consider any compact subset $D_c \subset \mathcal{D}$. By Assumption 1 and the fact that $\pi_{\mathcal{D}}^{Q}$ is in $L^\infty(\mathcal{D})$ in an almost sense, there exists an open set N_δ such that $\mu_{\mathcal{D}}(N_\delta) < \delta$, $(\pi_{\mathcal{D}}^{Q_n})$ is uniformly bounded and a.e.c. on $\mathcal{D}\backslash N_\delta$, and $\pi_{\mathcal{D}}^{Q}$ is in $L^\infty(\mathcal{D}\backslash N_\delta)$. By the compactness of D_c and openness of N_δ, $(\pi_{\mathcal{D}}^{Q_n})$ is a.u.e.c. on $D_c\backslash N_\delta$. Then, by Theorem 2 from [15],[§§]

$$\pi_{\mathcal{D}}^{Q_n} \to \pi_{\mathcal{D}}^{Q} \text{ in } L^\infty(D_c\backslash N_\delta).$$

Finally, for any $1 \leq r < \infty$, the embedding $L^\infty(D_c\backslash N_\delta) \subset L^r(D_c\backslash N_\delta)$ implies $\pi_{\mathcal{D}}^{Q_n} \to \pi_{\mathcal{D}}^{Q}$ in $L^r(D_c\backslash N_\delta)$ which proves (2). □

APPENDIX A.2 Proof of Lemma 2

Before we prove this result, we recall the definitions of uniformly integrable and uniformly absolutely continuous functions on a measure space $(X, \mathcal{B}_X, \mu_X)$. A sequence of measurable functions (f_n) is *uniformly integrable* if given $\epsilon > 0$ there exists M such that $\int_{\{x\,:\,|f_n(x)|>M\}} |f_n(x)|\, d\mu_X < \epsilon$ for each n. A sequence of measurable functions (f_n) is *uniformly absolutely continuous* if given $\epsilon > 0$ there exists $\delta > 0$ such that $\left|\int_A f_n(x)\, d\mu_X\right| < \epsilon$ for each n whenever $\mu_X(A) < \delta$. When μ_X is a finite measure, (f_n) is uniformly integrable if and only if $\sup_n \int |f_n(x)|\, d\mu_X < \infty$ and (f_n) is uniformly absolutely continuous.[¶¶] This result is exploited in the proof below.

Proof. Let $\epsilon > 0$. Use the uniform integrability in $L^r(\mathcal{D})$ to choose $\delta > 0$ such that the integral of any push-forward density raised to the r power over a set $A \in \mathcal{B}_{\mathcal{D}}$ with $\mu_{\mathcal{D}}(A) < \delta$ is bounded by $\epsilon^r/2$. Use Assumption 1 to choose the N_δ set. Using the fact that

$$\left\|\pi_{\mathcal{D}}^{Q_n}(q) - \pi_{\mathcal{D}}^{Q}(q)\right\|_{L^r(D_c)} = \left[\left\|\pi_{\mathcal{D}}^{Q_n}(q) - \pi_{\mathcal{D}}^{Q}(q)\right\|_{L^r(D_c\backslash N_\delta)}^r + \left\|\pi_{\mathcal{D}}^{Q_n}(q) - \pi_{\mathcal{D}}^{Q}(q)\right\|_{L^r(N_\delta)}^r\right]^{1/r},$$

along with Eq. (2) to bound the first term by $\epsilon^r/2$ for sufficiently large n proves (3). □

APPENDIX A.3 Proof of Theorem 1

The proof of the following lemma is omitted since it immediately follows by applying the classical result that $C_c^\infty(\mathbb{R}^n)$ (i.e., the space of infinitely differentiable functions with compact support) is dense in $L^r(\mathbb{R}^n)$ [48].

Lemma 4. *Lipschitz continuous functions with compact support are dense in $L^r(X)$ for $X \in \{\mathbf{\Lambda}, \mathcal{D}\}$ and any $1 \leq r < \infty$.*

Proof. Let $\epsilon > 0$. By Lemma 4, there exists a Lipschitz continuous $\widetilde{\pi_{\mathcal{D}}^{Q}}$ approximating $\pi_{\mathcal{D}}^{Q}$ such that

$$\left\|\pi_{\mathcal{D}}^{Q}(q) - \widetilde{\pi_{\mathcal{D}}^{Q}}(q)\right\|_{L^p(\mathcal{D})} < \frac{\epsilon}{4}. \tag{A.1}$$

Applying the triangle inequality three times gives

$$\begin{aligned}\left\|\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) - \pi_{\mathcal{D}}^{Q}(Q(\lambda))\right\|_{L^p(\mathbf{\Lambda})} \leq{}& \left\|\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) - \pi_{\mathcal{D}}^{Q}(Q_n(\lambda))\right\|_{L^p(\mathbf{\Lambda})} \\ &+ \left\|\pi_{\mathcal{D}}^{Q}(Q_n(\lambda)) - \widetilde{\pi_{\mathcal{D}}^{Q}}(Q_n(\lambda))\right\|_{L^p(\mathbf{\Lambda})} \\ &+ \left\|\widetilde{\pi_{\mathcal{D}}^{Q}}(Q_n(\lambda)) - \widetilde{\pi_{\mathcal{D}}^{Q}}(Q(\lambda))\right\|_{L^p(\mathbf{\Lambda})} \\ &+ \left\|\widetilde{\pi_{\mathcal{D}}^{Q}}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q}(Q(\lambda))\right\|_{L^p(\mathbf{\Lambda})}.\end{aligned} \tag{A.2}$$

[§§] Theorem 2 from [15] states that for a sequence of distributions (G_n) and corresponding densities (g_n) the following two statements are equivalent. (1) (g_n) is a.u.e.c. and bounded, and (G_n) converges to a distribution G. (2) $g_n \to g$ pointwise, uniformly on their domain, where g is the uniformly continuous density of the distribution G.

[¶¶] See, for example, Proposition 4.5.3 in [14].

Recalling that $\widetilde{\pi_{\mathcal{D}}^{Q}}$ is Lipschitz continuous, there is a constant $C > 0$ such that

$$\left\| \widetilde{\pi_{\mathcal{D}}^{Q}}(Q_n(\lambda)) - \widetilde{\pi_{\mathcal{D}}^{Q}}(Q(\lambda)) \right\|_{L^p(\mathbf{\Lambda})} \leq C \| Q_n(\lambda) - Q(\lambda) \|_{L^p(\mathbf{\Lambda})}.$$

Then, $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$ implies that the third term on the right-hand side of Eq. (A.2) is bounded by $\epsilon/4$ by setting n sufficiently large.

The first, second, and fourth terms on the right-hand side of Eq. (A.2) are equivalently written, respectively, as

$$\begin{aligned}
\left\| \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) - \pi_{\mathcal{D}}^{Q}(Q_n(\lambda)) \right\|_{L^p(\mathbf{\Lambda})} &= \left\| \pi_{\mathcal{D}}^{Q_n}(q) - \pi_{\mathcal{D}}^{Q}(q) \right\|_{L^p(\mathcal{D})}, \\
\left\| \pi_{\mathcal{D}}^{Q}(Q_n(\lambda)) - \widetilde{\pi_{\mathcal{D}}^{Q}}(Q_n(\lambda)) \right\|_{L^p(\mathbf{\Lambda})} &= \left\| \pi_{\mathcal{D}}^{Q}(q) - \widetilde{\pi_{\mathcal{D}}^{Q}}(q) \right\|_{L^p(\mathcal{D})}, \\
\left\| \widetilde{\pi_{\mathcal{D}}^{Q}}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q}(Q(\lambda)) \right\|_{L^p(\mathbf{\Lambda})} &= \left\| \widetilde{\pi_{\mathcal{D}}^{Q}}(q) - \pi_{\mathcal{D}}^{Q}(q) \right\|_{L^p(\mathcal{D})}.
\end{aligned}$$

By Eq. (A.1), the second and fourth terms on the right-hand side of Eq. (A.2) are bounded by $\epsilon/4$. Finally, by Lemma 2, the first term on the right-hand side of Eq. (A.2) is bounded by $\epsilon/4$, which proves (4). □

APPENDIX B. PROOFS FOR INVERSE PROBLEM ANALYSIS

APPENDIX B.1 Proof of Lemma 3

Before we prove this lemma, we recall a few standard results from measure theory to simplify the first few steps of the proof. First, if $(X, \mathcal{B}_X, \mu_X)$ is a measure space and $f \in L^1(X)$, then for any $\epsilon > 0$ there exists $a > 0$ such that

$$\int_{\{x\,:\,|f(x)|>a\}} |f(x)|\, d\mu_X > \int_X |f(x)|\, d\mu_X - \epsilon.$$

Moreover, if the measure space is σ-finite, then there exists $A \in \mathcal{B}_X$ such that $\mu_X(A) < \infty$ and

$$\int_A |f(x)|\, d\mu_X > \int_X |f(x)|\, d\mu_X - \epsilon.$$

Combining these results, it is possible to choose $a > 0$ and A compact such that

$$\int_A |f(x)|\, d\mu_X > \int_X |f(x)|\, d\mu_X - \epsilon, \quad \text{and} \quad |f(x)| > a, \ \ \forall x \in A.$$

We also make use of the standard measure theory results involving approximating sets of finite measure with open or compact sets as discussed following Lemma 1.

Proof. Let $0 < \delta < 1$. Following the discussion above, there exists $a > 0$ and compact $D \in \mathcal{B}_{\mathcal{D}}$ (so $\mu_{\mathcal{D}}(D) < \infty$) such that

$$\int_D \pi_{\mathcal{D}}^{Q}(q)\, d\mu_{\mathcal{D}} > 1 - \frac{\delta}{4}, \quad \text{and} \quad \pi_{\mathcal{D}}^{Q}(q) > 2a, \ \ \forall q \in D.$$

Using Assumption 1, choose $\eta > 0$ sufficiently small and N_η an open set such that the sequence of approximate push-forwards is a.e.c. on $D_a := D \backslash N_\eta$, $\mu_{\mathcal{D}}(D_a) > \mu_{\mathcal{D}}(D) - \eta$, and

$$\int_{D_a} \pi_{\mathcal{D}}^{Q}(q)\, d\mu_{\mathcal{D}} > 1 - \frac{\delta}{2}, \quad \text{and} \quad \pi_{\mathcal{D}}^{Q}(q) > 2a, \ \ \forall q \in D_a.$$

Let $0 < \epsilon < \min\{a, \delta/(2\mu_{\mathcal{D}}(D_a))\}$. By design, D_a is itself compact, so the sequence of approximate push-forwards is in fact a.u.e.c. on D_a. Thus, by application of Theorem 2 in [15] on D_a there exists $N > 0$ such that for any $n > N$ and $q \in D_a$,

$$-\epsilon < \pi_{\mathcal{D}}^{Q_n}(q) - \pi_{\mathcal{D}}^{Q}(q) < \epsilon.$$

Then, $q \in D_a$ implies $\pi_{\mathcal{D}}^{Q_n}(q) > \pi_{\mathcal{D}}^{Q}(q) - \epsilon > a$, which proves (13).

Finally, for any $n > N$,

$$\begin{aligned}\int_{D_a} \pi_{\mathcal{D}}^{Q_n}(q)\, d\mu_{\mathcal{D}} &= \int_{D_a} \left[\pi_{\mathcal{D}}^{Q_n}(q) - \pi_{\mathcal{D}}^{Q}(q)\right] d\mu_{\mathcal{D}} + \int_{D_a} \pi_{\mathcal{D}}^{Q}(q)\, d\mu_{\mathcal{D}} \\ &> -\epsilon\mu_{\mathcal{D}}(D_a) + 1 - \frac{\delta}{2} > 1 - \delta.\end{aligned}$$

This proves (14). □

APPENDIX B.2 Proof of Theorem 2

Without loss of generality, we prove this theorem under some additional simplifying assumptions. Specifically, we assume that $\pi_{\mathcal{D}}$ and $\pi_{\mathbf{\Lambda}}^{\text{init}}$ are both Lipschitz continuous and that $\mathcal{D}$ and $\mathbf{\Lambda}$ are both compact. If this is not the case, we can carry out the analysis using "sufficiently good" approximations to $\pi_{\mathcal{D}}$ and $\pi_{\mathbf{\Lambda}}^{\text{init}}$ that are Lipschitz continuous with compact support by Lemma 4 (presented in Appendix A.3), and simply use triangle inequalities to prove the result for more general initial and observed densities and noncompact parameter and data spaces.

Proof. Let $\epsilon > 0$. Since $\pi_{\mathbf{\Lambda}}^{\text{init}} \in C(\mathbf{\Lambda})$ with compact support, there exists $M > 0$ such that for any $\lambda \in \mathbf{\Lambda}$

$$\left|\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)\right| < M.$$

Set $0 < \delta < \epsilon^p/(2^{p+1}C^pM^{p-1})$ (here, the C is from Assumption 2). By Lemma 3, there exists $a > 0$, $D_a \in \mathcal{B}_{\mathcal{D}}$, and $N_1 > 0$ such that for any $n > N_1$,

$$\pi_{\mathcal{D}}^{Q}(q) > a, \quad \pi_{\mathcal{D}}^{Q_n}(q) > a, \quad \forall q \in D_a$$

$$\int_{D_a} \pi_{\mathcal{D}}^{Q}(q)\, d\mu_{\mathcal{D}} > 1 - \delta, \quad \int_{D_a} \pi_{\mathcal{D}}^{Q_n}(q)\, d\mu_{\mathcal{D}} > 1 - \delta.$$

Denote $\mathbf{\Lambda}_{a,n} = Q_n^{-1}(D_a)$, $\mathbf{\Lambda}_a = Q^{-1}(D_a)$. Then, by linearity of the integral operator,

$$\begin{aligned}\left\|\pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) - \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda)\right\|_{L^p(\mathbf{\Lambda})}^p &= \int_{\mathbf{\Lambda}} \left|\pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) - \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda)\right|^p d\mu_{\mathbf{\Lambda}} \\ &= \int_{\mathbf{\Lambda}_{a,n}} \left|\pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) - \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda)\right|^p d\mu_{\mathbf{\Lambda}} + \int_{\mathbf{\Lambda}\setminus\mathbf{\Lambda}_{a,n}} \left|\pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) - \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda)\right|^p d\mu_{\mathbf{\Lambda}} \\ &= \int_{\mathbf{\Lambda}_{a,n}} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p |r_n(\lambda) - r(\lambda)|^p\, d\mu_{\mathbf{\Lambda}} + \int_{\mathbf{\Lambda}\setminus\mathbf{\Lambda}_{a,n}} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p |r_n(\lambda) - r(\lambda)|^p\, d\mu_{\mathbf{\Lambda}}.\end{aligned}$$

Observe that the difference in ratios given by $|r_n(\lambda) - r(\lambda)|$ can be rewritten as

$$|r_n(\lambda) - r(\lambda)| = \left|\frac{\pi_{\mathcal{D}}(Q_n(\lambda))\pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}(Q(\lambda))\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))\pi_{\mathcal{D}}^{Q}(Q(\lambda))}\right|.$$

Then, by adding and subtracting $\pi_{\mathcal{D}}(Q(\lambda))\pi_{\mathcal{D}}^{Q}(Q(\lambda))$ in the numerator, this difference is decomposed as

$$|r_n(\lambda) - r(\lambda)| \le \underbrace{\left|\frac{\pi_{\mathcal{D}}(Q_n(\lambda)) - \pi_{\mathcal{D}}(Q(\lambda))}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))}\right|}_{T_1(\lambda)} + \underbrace{\left|\frac{\pi_{\mathcal{D}}(Q(\lambda))\left[\pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))\right]}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))\pi_{\mathcal{D}}^{Q}(Q(\lambda))}\right|}_{T_2(\lambda)}. \tag{B.1}$$

Observe that

$$\int_{\mathbf{\Lambda}_{a,n}} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p |r_n(\lambda) - r(\lambda)|^p \, d\mu_{\mathbf{\Lambda}} \leq \int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_1(\lambda) + T_2(\lambda))^p \, d\mu_{\mathbf{\Lambda},q} d\mu_{\mathcal{D}}$$
$$\leq \int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (2\text{max}_{\lambda \in \mathbf{\Lambda}}\{T_1(\lambda), T_2(\lambda)\})^p \, d\mu_{\mathbf{\Lambda},q} d\mu_{\mathcal{D}}.$$

Thus, we show that $\int_{\mathbf{\Lambda}_{a,n}} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p |r_n(\lambda) - r(\lambda)|^p \, d\mu_{\mathbf{\Lambda}} < \epsilon^p/2$ by proving

$$\int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_1(\lambda))^p \, d\mu_{\mathbf{\Lambda},q} d\mu_{\mathcal{D}} < \frac{\epsilon^p}{2^{p+2}},$$

and

$$\int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_2(\lambda))^p \, d\mu_{\mathbf{\Lambda},q} d\mu_{\mathcal{D}} < \frac{\epsilon^p}{2^{p+2}}.$$

Denote the Lipschitz constant for $\pi_{\mathcal{D}}$ by $C_1 \geq 0$, then

$$T_1(\lambda) \leq \frac{C_1 |Q(\lambda) - Q_n(\lambda)|}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))}. \tag{B.2}$$

Hölder's inequality then implies

$$\int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_1(\lambda))^p \, d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q)$$
$$\leq (C_1)^p \int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} |Q(\lambda) - Q_n(\lambda)|^p \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q)$$
$$\leq (C_1)^p \int_{\mathcal{D}} \| |Q(\lambda) - Q_n(\lambda)|^p \|_{L^1(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))} \left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))} d\mu_{\mathcal{D}}(q)$$
$$\leq (C_1)^p \int_{\mathcal{D}} \|Q(\lambda) - Q_n(\lambda)\|^p_{L^p(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))} \left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))} d\mu_{\mathcal{D}}(q)$$
$$\leq (C_1)^p \left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n})} \int_{\mathcal{D}} \|Q(\lambda) - Q_n(\lambda)\|^p_{L^p(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))} \, d\mu_{\mathcal{D}}(q).$$

By the Disintegration Theorem,

$$\int_{\mathbf{\Lambda}_{a,n}} |Q(\lambda) - Q_n(\lambda)|^p \, d\mu_{\mathbf{\Lambda}} = \int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} |Q(\lambda) - Q_n(\lambda)|^p d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q)$$
$$= \int_{\mathcal{D}} \|Q(\lambda) - Q_n(\lambda)\|^p_{L^p(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))} \, d\mu_{\mathcal{D}}(q).$$

Then, we have

$$\int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_1(\lambda))^p \, d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q)$$
$$\leq (C_1)^p \left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n})} \int_{\mathbf{\Lambda}_{a,n}} |Q(\lambda) - Q_n(\lambda)|^p \, d\mu_{\mathbf{\Lambda}}$$
$$\leq (C_1)^p \left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n})} \|Q(\lambda) - Q_n(\lambda)\|^p_{L^p(\mathbf{\Lambda})}.$$

By construction,

$$\left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n})} \leq \frac{M^p}{a^p}.$$

Since $Q_n \to Q$ in $L^p(\mathbf{\Lambda})$, there exists $N_2 > 0$ such that for any $n > N_2$,

$$\|Q_n(\lambda) - Q(\lambda)\|_{L^p(\mathbf{\Lambda})}^p < \frac{a^p \epsilon^p}{2^{p+2}(C_1)^p M^p}.$$

Combining these inequalities gives

$$\int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_1(\lambda))^p \, d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q) < \frac{\epsilon^p}{2^{p+2}}.$$

We now bound the term involving $T_2(\lambda)$. First, rewrite Assumption 2 as $\pi_{\mathcal{D}}(q)/\pi_{\mathcal{D}}^{Q}(q) \leq C$ for a.e. $q \in \mathcal{D}$. Then, use Hölder's inequality and the Disintegration Theorem as before to get

$$\begin{aligned}
&\int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_2(\lambda))^p \, d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q) \\
&\leq C^p \int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} \left[\pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) \right]^p \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q) \\
&\leq C^p \int_{\mathcal{D}} \left\| \pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) \right\|_{L^p(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))}^p \left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q))} d\mu_{\mathcal{D}}(q) \\
&\leq C^p \left\| \left(\frac{\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda)}{\pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda))} \right)^p \right\|_{L^\infty(\mathbf{\Lambda}_{a,n})} \left\| \pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) \right\|_{L^p(\mathbf{\Lambda})}^p.
\end{aligned}$$

By Eq. (4) of Theorem 1, $\left\| \pi_{\mathcal{D}}^{Q}(Q(\lambda)) - \pi_{\mathcal{D}}^{Q_n}(Q_n(\lambda)) \right\|_{L^p(\mathbf{\Lambda})} \to 0$. It follows in a similar manner as before that there exists $N_3 > 0$ such that for any $n > N_3$,

$$\int_{\mathcal{D}} \int_{\mathbf{\Lambda}_{a,n} \cap Q_n^{-1}(q)} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p (T_2(\lambda))^p \, d\mu_{\mathbf{\Lambda},q}(\lambda) \, d\mu_{\mathcal{D}}(q) < \frac{\epsilon^p}{2^{p+2}}.$$

Set $N = \max\{N_1, N_2, N_3\}$. For any $n > N$,

$$\int_{\mathbf{\Lambda}_{a,n}} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p |r_n(\lambda) - r(\lambda)|^p \, d\mu_{\mathbf{\Lambda}} < \frac{\epsilon^p}{2}.$$

We now bound the other error term defined on the set $\mathbf{\Lambda} \backslash \mathbf{\Lambda}_{a,n}$.

$$\begin{aligned}
\int_{\mathbf{\Lambda} \backslash \mathbf{\Lambda}_{a,n}} (\pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda))^p |r_n(\lambda) - r(\lambda)|^p \, d\mu_{\mathbf{\Lambda}} &\leq \int_{\mathbf{\Lambda} \backslash \mathbf{\Lambda}_{a,n}} \pi_{\mathbf{\Lambda}}^{\text{init}}(\lambda) (2C)^p M^{p-1} \, d\mu_{\mathbf{\Lambda}} \\
&= (2C)^p M^{p-1} \int_{\mathcal{D} \backslash D_a} \pi_{\mathcal{D}}^{Q_n}(q) \, d\mu_{\mathcal{D}} \\
&< (2C)^p M^{p-1} \delta < \frac{\epsilon^p}{2}.
\end{aligned}$$

Thus, the sum of the error terms for any $n > N$ satisfies

$$\begin{aligned}
\left\| \pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) - \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda) \right\|_{L^p(\mathbf{\Lambda})}^p &= \int_{\mathbf{\Lambda}_{a,n}} \left| \pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) - \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda) \right|^p d\mu_{\mathbf{\Lambda}} + \int_{\mathbf{\Lambda} \backslash \mathbf{\Lambda}_{a,n}} \left| \pi_{\mathbf{\Lambda}}^{\text{up},n}(\lambda) - \pi_{\mathbf{\Lambda}}^{\text{up}}(\lambda) \right|^p d\mu_{\mathbf{\Lambda}} \\
&< \frac{\epsilon^p}{2} + \frac{\epsilon^p}{2} = \epsilon^p.
\end{aligned}$$

Raising each side to the $1/p$ power finishes the proof. □

APPENDIX C. OBTAINING DATA, FIGURES, AND SCRIPTS

The repository "Lp" (available at https://github.com/User-zwj/Lp.git) includes the scripts producing all the tables and figures in this article. The scripts are written in Python, and utilize libraries Numpy [49], SciPy [50], Matplotlib [51], and DOLFIN [52]. Python 3.6 or newer is recommended. Instructions are included in the README file.