ABSTRACT. One approach to noisy inverse problems is to use Bayesian methods. In this work, the statistical inverse problem of estimating the probability distribution of an infinite-dimensional unknown given its noisy indirect infinite-dimensional observation is studied in the Bayesian framework. The motivation for the work arises from the fact that the Bayesian computations are usually carried out in finite-dimensional cases, while the original inverse problem is often infinite-dimensional. A good understanding of an infinite-dimensional problem is, in general, helpful in finding efficient computational approaches to the problem.

The fundamental question of well-posedness of the infinite-dimensional statistical inverse problem is considered. In particular, it is shown that the continuous dependence of the posterior probabilities on the realizations of the observation provides a certain degree of uniqueness for the posterior distribution.

Special emphasis is on finding tools for working with non-Gaussian noise models. Especially, the applicability of the generalized Bayes formula is studied. Several examples of explicit posterior distributions are provided.

1. Introduction. Statistically oriented infinite-dimensional inverse problems are often described as problems where one wants to estimate an unknown function given its randomly perturbed indirect observation [15, 35, 43, 82, 95, 124]. We prefer the following description which suits well in the Bayesian framework.

The statistical inverse problem is to estimate the probability distribution of the unknown given its randomly perturbed indirect observation.

In this paper, the unknown $X$ and its observation $Y$ are modeled as random mappings from a complete probability space $(\Omega, \Sigma, P)$ into some locally convex Souslin topological vector spaces $F$ and $G$ equipped with their Borel $\sigma$-algebras $\mathcal{F}$ and $\mathcal{G}$, respectively. Recall, that a Souslin space is a Hausdorff topological space that is an image of a complete separable metric space under a continuous mapping. The observations are taken to be of the form \( Y = L(X) + \varepsilon \), where $\varepsilon$ represents random noise, $\varepsilon$ and $X$ are statistically independent, and $L : F \to G$ is a continuous mapping. The image measure $\mu_X := P \circ X^{-1}$ on $F$ is called the prior distribution, and it represents our beliefs about the unknown without any given observations.
Typically, we are given a sample \( Y(\omega_0) = L(X(\omega_0)) + \varepsilon(\omega_0) \), which is produced by an unknown \( X(\omega_0) \) and a perturbation \( \varepsilon(\omega_0) \) for some \( \omega_0 \in \Omega \). Ultimately, we pursue after the probability measure \( U \mapsto 1_U(X(\omega_0)) \) defined on the Borel sets \( U \subseteq F \). This measure would determine the unknown \( X(\omega_0) \) uniquely since \( F \) is a Hausdorff space, which implies that the singletons are closed sets and belong therefore to the Borel \( \sigma \)-algebra \( \mathcal{F} \). We get a simple approximation of the function \( \omega \mapsto 1_U(X(\omega)) \) on the basis of the given \( Y(\omega_0) \) by taking its orthogonal projection from \( L^2(\Omega, \Sigma, \mathcal{P}) \) onto \( L^2(\Omega, \sigma(Y), \mathcal{P}) \), where \( \sigma(Y) = Y^{-1}(\mathcal{G}) \) denotes the \( \sigma \)-algebra generated by \( Y \). Recall, that for any \( f \in L^2(\Omega, \Sigma, \mathcal{P}) \), this projection coincides \( \mathcal{P} \)-almost surely with the conditional expectation \( E[f|\sigma(Y)] \) of \( f \) given the \( \sigma \)-algebra generated by \( Y \) (see [34]). Moreover, there exists a measurable real-valued function \( \lambda_f \) on \( G \) such that \( \lambda_f(Y(\omega)) = E[f|\sigma(Y)](\omega) \) \( \mathcal{P} \)-almost surely. We take \( E[1_U(X)|\sigma(Y)](\omega_0) \) (or more precisely, \( \lambda_{1_U}(Y(\omega_0)) \)) as an estimate of the probability that the unknown \( X(\omega_0) \) belongs to the set \( U \in \mathcal{F} \).

When the mappings \( U \mapsto E[1_U(X)|\sigma(Y)](\omega_0) \) form a probability measure on \( (F, \mathcal{F}) \), which is denoted here with \( \mu(U, X(\omega_0)) \), this measure is called the posterior distribution of \( X \) given a sample \( Y(\omega_0) \) of \( Y \). From the posterior distribution one may extract information about the unknown \( X(\omega_0) \). For example, the posterior mean may serve as an estimate of the unknown.

The above estimation of the probability distribution of the unknown is generally known as the statistical inverse theory (also known as the statistical inversion or the Bayesian inversion). We postpone a literature review on the statistical inverse theory to Section 1.3. The present work concentrates on the following three topics in the statistical inverse theory inspired by a paper of Lassas et al [81].

(i) Applicability of the generalized Bayes formula for statistical inverse problems in locally convex Souslin topological vector spaces.
(ii) Well-posedness of the Bayesian statistical inverse problem.
(iii) Convergence of posterior distributions and posterior means for approximated unknowns. Especially, finding conditions that guarantee the convergence of the posterior distributions when the corresponding approximated prior distributions converge.

Cases (i) and (ii) are studied in the present Part I and Case (iii) in Part II.

1.1. Case (i): The generalized Bayes formula. When \( X \) and \( Y \) have continuous probability densities with respect to the Lebesgue measure, the conditional expectations lead to the Bayes formula

\[
D(x|y)D(y) = D(x, y) = D(y|x)D(x)
\]

which defines the unique continuous posterior probability density \( D(x|y) \) for any occurred observation \( y \) such that \( 0 < D_Y(y) < \infty \) (see [65]). In (1), the functions \( D(x), D(y), \) and \( D(x, y) \) denote the probability densities of \( P \circ X^{-1}, P \circ Y^{-1}, \) and \( P \circ (X, Y)^{-1} \) at \( x, y, \) and \( (x, y) \), respectively. If the observation is of the form \( Y = L(X) + \varepsilon, \) where \( X \) and the noise \( \varepsilon \) are statistically independent, then the conditional density of \( Y \) given \( X = x \) has the special form \( D(y|x) = D_{\varepsilon}(y - L(x)) \), where \( D_{\varepsilon} \) is the continuous probability density of the noise \( \varepsilon \).

The availability of the conditional density \( D(y|x) \) from the relationship between unknowns and observations is the key element for the statistical inverse theory. It makes the expression of the posterior density \( D(x|y) \) explicit, opening the way for exploring the posterior distribution numerically. Unfortunately, infinite-dimensional
probability measures lack probability density functions since there is no infinite-dimensional Lebesgue measure. Instead of (1), we have

\[ E[1_U(X)\sigma(Y)] = P(X \in U \cap Y \in V) = E[1_U(X)E[1_V(Y)|\sigma(X)]] \]

for all Borel sets \( U \in \mathcal{F} \) and \( V \in \mathcal{G} \). The distributions of \( X, Y \), and \((X,Y)\) are, in principle, known. However, determining \( E[1_U(X)\sigma(Y)] \) explicitly from the first equality in (2) is in general a hard task, where an explicit expression of the distribution of \((X,Y)\) is helpful, as in the case of linear Gaussian problems [85, 87, 90]. On the other hand, the second equality in (2) looks easy enough. For instance, the dominated convergence of simple functions to the exponential function shows [90]. On the other hand, the second equality in (2) looks easy enough. For instance, the dominated convergence of simple functions to the exponential function shows

\[ E[e^{i(X,\phi)+i(Y,\psi)}] = E[e^{i(X,\phi)}E[e^{i(Y,\psi)}|\sigma(X)]] \]

for all \( \phi \) and \( \psi \) in the dual spaces \( F' \) and \( G' \), respectively. This suggests that, after verifying some measurability conditions, we may take

\[ E[1_V(Y)|\sigma(X)](\omega) = \mu_{\epsilon+L(X(\omega))}(V) \]

\( P \)-almost surely since \( X \) and \( \epsilon \) are statistically independent. Does knowing the conditional probabilities \( E[1_V(Y)|\sigma(X)](\omega) \) help in determining the posterior distribution? The answer is positive in some cases. If the \( \sigma \)-algebra \( \mathcal{G} \) in question is countably generated and the conditional distributions of \( Y \) given \( X \) are regular and \( P \)-almost surely absolutely continuous with respect to some fixed \( \sigma \)-finite measure \( \lambda \) on \( \mathcal{G} \) (i.e. they are dominated by \( \lambda \)), then the generalized Bayes formula

\[ \mu(U, y) = \int_U \frac{d\mu_{Y|X}(\cdot|x)}{d\lambda}(y) d\mu_X(x) \]

is known to hold for \( U \in \mathcal{F} \) and \( \mu_{Y|X} \)-almost every given observation \( Y = y \) such that the denominator is finite and non-zero [68, 111]. In (3), it is required that the Radon-Nikodym densities \( \frac{d\mu_{Y|X}(\cdot|x)}{d\lambda}(y) \) of the conditional measure \( \mu_{Y|X}(\cdot|x) \) of \( Y \) given \( X = x \) with respect to \( \lambda \) are jointly measurable. This is sometimes achieved by defining the Radon-Nikodym densities with the help of a fixed joint density as is done in [111]. In (3), the form of \( Y \) is allowed to be more general than in our restricted case of \( Y = L(X) + \varepsilon \), where the posterior distribution has, for suitable \( L, X, \) and \( \varepsilon \), the form

\[ \mu(U, y) = \int_U \frac{d\mu_{Y+L(\varepsilon)}(\cdot|x)}{d\lambda}(y) d\mu_X(x) \]

for all \( U \in \mathcal{F} \) and \( \mu_{Y+L(\varepsilon)} \)-a.e. \( y \in \mathcal{G} \) such that the denominator is finite and non-zero. When the Radon-Nikodym densities in (4) are known, the posterior distribution on \( F \) has an explicit representation for all admissible \( y \in \mathcal{G} \).

In statistical inverse problems, the generalized Bayes formula for function-valued unknowns has been used before in the case of finite-dimensional noise models that have probability density functions with respect to the Lebesgue measure [22, 39, 82, 120] and in the case of infinite-dimensional Gaussian noise models, using in (4) the Cameron-Martin formula [54, 81, 120]. The starting point in [22, 23, 120] is that the posterior distribution is assumed to have Radon-Nikodym density with respect to the prior distribution. Therefore, a similar formula like (3) is used in [22, 23, 120], but not derived. In [54, 81, 82], the unknown and the noise are statistically independent. The same seems to be the case in same examples in [22, 23, 120] but the fact is not emphasized.
Fitzpatrick [39] studied (separable) Banach-space valued unknowns, and wrote the expression (4) in the case of finite-dimensional observations. As a concrete example, he used a Gaussian prior distribution on \( C([0, 1]) \) in the ill-posed inverse problem of determining the function \( q \) in the differential equation \(- (q u')' = f\) on \( (0, 1) \) from finitely many noisy values of the solution \( u \) satisfying the Dirichlet boundary condition. Lassas and Siltanen [82] used the generalized Bayes formula for certain prior random variables on \( C([0, 1]) \) and assumed the finite-dimensional noise to be Gaussian. Lassas et al [81] and Helin [54] had emphasis on edge-preserving prior distributions and used linear forward theory with Gaussian noise, but they allowed in (4) also other separable Banach and Hilbert space-valued unknowns, respectively. The forward mapping \( L \) was assumed to be linear in [54, 81, 82]. Cotter et al [22] studied the case of finite-dimensional observations and Banach space-valued unknowns, and required \( L \) to be measurable. Stuart [120] assumed \( L \) to be locally Lipschitz continuous and assumed finite-dimensional or Gaussian noise. Stuart allowed prior distributions that are absolutely continuous with respect to some Gaussian measure. Theorem 4.1 in [120] is an abstract generalization towards allowing certain infinite-dimensional non-Gaussian noise distributions but the identification of the used notation to any statistical inverse problem is omitted. The same approach is used in [23].

In the present paper, we provide (abstract) assumptions on the forward theory and the noise that are sufficient for the generalized Bayes formula in the case of statistical inverse problems in locally convex Souslin topological vector spaces. However, such a generalization is not particularly novel by itself, and the generalized Bayes formula is treated in this work as an important tool for achieving other results. For example, the study of Case (iii) exploits the generalized Bayes formula.

In this work, we allow infinite-dimensional noise models (similarly as in [54, 80, 81, 99]). One may ask, what are the benefits of such models because any feasible measuring instrument produces only finite-dimensional observations. For example, an analog-to-digital converter performs the weighted averaging and quantization of the signal; an X-ray imaging device has a finite number of projection angles and a limited resolution of the projection images. In this light, there is no immediate need for infinite-dimensional noise models. However, changes in the measuring instrument can lead to different posterior distributions and one may wish to choose the best finite-dimensional measurement configuration for the problem. As noted in [81], the mathematical formulation of the infinite-dimensional noise model, when possible, may be helpful, as it provides an overall framework for the studies. For some noise sources there even exists physically motivated infinite-dimensional noise models, like the model of the thermal noise in electric circuits, which arises from the thermal motion of the charge carriers. Particular emphasis in this work is on finding tools for dealing with non-Gaussian noise in infinite-dimensional statistical inverse problems. There are three reasons why the Gaussian noise model is not satisfactory.

1. Noise does not always follow well enough a Gaussian distribution. In Section 4.4 we discuss the appearance of \( \alpha \)-stable noise in statistical inverse problems. An evaluation of finite-dimensional noise models in medical imaging can be found in [49].
2. Model approximations – which were studied first by Kaipio and Somersalo [66] for finite-dimensional observations – can also produce infinite-dimensional non-Gaussian errors (cf. Remark 13).
3. Some noise statistics may not be exactly known. In statistical inverse theory the inaccuracies in the noise model are further modeled with hierarchical distributions (see Section 4.5 for a special case). A wrong noise distribution may cause poor performance of the estimators of the unknown.

Posterior distributions in locally convex Souslin (not metrizable) topological vector spaces have been considered before in the special case of Gaussian distribution-valued random variables in [80, 85] and in the general abstract case of Souslin space-valued random variables only in [99]. The first part of Section 2 is therefore devoted to the basics of the abstract statistical inverse theory in locally convex Souslin topological vector spaces. Unlike in the work of Piiroinen [99], where general Souslin space-valued random variables in statistical inverse problems were first studied, we use the generalized Bayes formula for the proofs. We also derive the generalized Bayes formula from the equation \( Y = L(X) + \varepsilon \), which supplements also the formulation presented in [22, 23, 120] where the starting point is a given form of the posterior distribution. This makes it easier to recognize situations where the generalized Bayes formula holds.

In this work, the use of locally convex Souslin topological vector spaces is mostly motivated by the fact that in Souslin spaces the Borel \( \sigma \)-algebras are regular enough for the existence of regular conditional measures [10]. Moreover, the class of such spaces contains many useful spaces, like complete separable metric vector spaces and spaces of (Schwartz) distributions [112]. We remark that the distribution spaces are sometimes preferred since the convergence of the characteristic functions of measures implies the weak convergence of measures for them (this fact is shown e.g. in [10] and used in [80]). We require \( G \) to be a topological vector space since \( Y \) is defined as the sum of two \( G \)-valued random variables. However, it is well-known that the sum of two random variables is not always a random variable in arbitrary topological vector spaces. In Lemma 3.1, we check that \( Y \) is indeed a random variable because the sample space \( G \) is a Souslin space (the fact is known but the author was unable to find a reference for the proof in the literature). We require \( G \) and \( F \) to be locally convex topological vector spaces since locally convex spaces have rich enough dual spaces that for example allow the use of characteristic functions in the identification of measures. In Remark 1, we note that a locally convex Souslin sample space of \( Y \) is allowed to be mis-specified by a continuous linear injection without altering the posterior distributions. This holds for any statistical inverse problem, not just for those admitting the representation (3).

In Section 4, we are able to derive (with the help of the generalized Bayes formula) explicit posterior distributions in some new cases where the noise has non-Gaussian infinite-dimensional distribution (Sections 4.3–4.7). It turns out that in some cases (Section 4.4) the posterior distribution has a simple expression for the infinite-dimensional observations but not for the truncated finite-dimensional observations. As a further motivation for the study of infinite-dimensional noise models, we suggest that the solutions of infinite-dimensional Bayesian problems may give rise to new numerically feasible, but non-Bayesian, approximations of the finite-dimensional posterior distributions.

1.2. Case (ii): Well-posedness of the Bayesian statistical inverse problem.
The projection operator from \( L^2(\Omega, \Sigma, P) \) onto \( L^2(\Omega, \sigma(Y), P) \) determines posterior probabilities \( \mu(U, Y(\omega)) \) only up to \( P \)-almost every \( \omega \in \Omega \). The uniqueness of the
posterior distribution for a given \( y \in \mathcal{R}(Y) \) is therefore unsettled (note that such form of nonuniqueness has nothing to do with the uniqueness of the deterministic inverse problem of recovering \( x_0 \) from \( L(x_0) \)). The nonuniqueness is fairly well understood in Gaussian linear problems \([80, 85, 87, 90, 116]\), where the posterior mean is known to be determined up to a set of probability zero, but has received limited attention in the general case. For a Bayesian scientist, such nonuniqueness is disconcerting. Two Bayesians using the same prior distribution and the same observations can, in principle, have different posterior distributions for some observations (in a set of probability zero). One aim of the present work is to make the two Bayesians agree on the form of their posteriors for a given \( y \in G \), at least in some special cases. In Theorem 2.4, we first carefully identify the nonuniqueness of the posterior distributions in locally convex Souslin topological vector spaces by adopting a new concept, the essential uniqueness, from the theory of conditional measures to the statistical inverse theory. Then, we apply a choice first made in \([81]\) and appearing also in \([22, 23, 54, 120]\), which is to work, if possible, with a fixed version of the posterior distribution depending continuously on observations in certain sense. Evans and Stark suggested even earlier that certain non-uniqueness problems with conditional expectations could be avoided by using dominated probabilities (see Remark 3.7 in \([35]\)).

The original part of this work begins in Section 2.2, where we achieve partial uniqueness of those posterior distributions that depend continuously on observations in the sense that posterior probabilities of Borel sets depend continuously on observations (cf. Theorem 2.7). The partial uniqueness gives an unambiguous meaning to the posterior distribution at a fixed observation. Moreover, it shows that then the Bayesian statistical inverse problem is well-posed – there exists a unique posterior distribution that depends continuously on the observations. The method of using continuous probability densities is widely used in the finite-dimensional case (see \([65]\)), but seems not to have been taken before within the abstract infinite-dimensional problems.

The posterior distributions are further studied in Theorem 2.8, where it is shown that the continuous dependence of the posterior probabilities of Borel sets on the observations implies the absolute continuity of the posterior distribution with respect to the prior distribution. We remark that Theorem 2.8 clarifies some of the differences between the undominated and the \( \mu_X \)-a.s. dominated cases. Indeed, if \( F \) and \( G \) are Polish vector spaces, a result of Macci \([88]\) says that the absolute continuity of \( \mu_Y \)-almost all posterior distributions with respect to the prior distribution is equivalent to the absolute continuity of the measures \( \mu_{\epsilon + L(x)} \) with respect to the measure \( \mu_Y \) for \( \mu_X \)-a.e. \( x \in F \). Hence, the continuous dependence of the posterior probabilities of Borel sets on observations is possible only when the measures \( \mu_{\epsilon + L(x)} \) are dominated by some \( \sigma \)-finite measure for \( \mu_X \)-a.e. \( x \in F \). What does this mean for the undominated cases? The posterior probability of at least one Borel set will be discontinuous as a function of observations. Hence, the corresponding Bayesian problem is ill-posed – small perturbations of the given sample can lead to large perturbations of some posterior probabilities. The ill-posedness in the linear Gaussian statistical inverse problems has been considered before by Florens and Simoni \([41, 116]\), who noted that the posterior mean in the Gaussian linear case can be ill-posed. Florens and Simoni also showed that the regularizing effect of the prior distribution has a limited power in such a case. They suggested using an additional Tikhonov regularization in the Gaussian linear case in order
to obtain approximations of the posterior means that depend continuously on the observations.

We note that the worst-case scenario for discontinuous posterior distributions on complete separable metric spaces is somewhat characterized in [13], where it is proved that either the set of all \( y \in G \) such that the posterior distributions \( \mu(\cdot, y) \) are mutually singular is (at most) countable or there exists a non-empty compact perfect set \( C \subset G \) and a Borel set \( B \subset F \times G \) such that \( 1 = \mu(B_y, y) = \mu(F \setminus B_y, y') \) for all \( y, y' \in C \) such that \( y \neq y' \). Here \( B_y = \{ x \in F : (x, y) \in B \} \).

In Theorem 3.4, we present some sufficient conditions that guarantee continuous dependence of the posterior posterior probabilities of Borel sets on observations by using the generalized Bayes formula. Cotter et al. [22, 23] have shown a closely related result which states that under certain conditions (including domination and Gaussian prior distribution), their version of the posterior distribution is Lipschitz continuous in finite-dimensional observations with respect to the Hellinger distance. Our proof relies on the Borel measurability of separately continuous functions – a result first obtained by Lebesgue and later generalized by Rudin [107]. The author was unable to find the proof of the measurability of separately continuous Souslin space-valued functions so the proof is included.

Note that even in a dominated case the posterior probabilities of Borel sets need not be continuous on any measurable linear subspace of full \( \mu_Y \)-measure (see Section 4.4 and Remark 11 for an example). However, the posterior probabilities of Borel sets in a dominated case are always continuous on certain compact sets of nearly full measure (cf. Theorem 2.9). Unfortunately, in infinite-dimensional normed spaces the interior of any compact set is empty. The partial uniqueness of Theorem 2.7 is therefore not generic for infinite-dimensional normed spaces, unless there is a locally finite union of compact sets \( K \) such that \( \bigcup_{i=1}^{\infty} K_i \) has full \( \mu_Y \)-measure and the restriction of \( \mu(U, \cdot) \) onto each \( K_i \) is continuous, which guarantees that \( \mu(U, \cdot) \) is continuous on the whole \( \bigcup_{i=1}^{\infty} K_i \). However, in Remark 3 we note that for any version of the posterior distribution there always exists some stronger topology on \( G \) that generates the same Borel sets, but makes the version continuous.

Notations: When \( G \) is a topological vector space, we denote with \( G' \) its topological dual space. If \( m \) is a measure on \( G \), we sometimes denote \( m(f) := \int_G f(x) \, dm(x) \). If \( Z : \Omega \to G \) is a random variable, its image measure \( P \circ Z^{-1} \) on \( G \) is denoted with \( \mu_Z \). A Borel measure and its Lebesgue’s completion are denoted with the same symbol.

1.3. A literature review. Statistical inverse theory became a popular method for solving geophysical problems in 1980’s [121, 122], and has since spread into many other fields (see [65, 120]). In this short review, we focus on general theoretical developments that lead to the modern description of statistical inverse theory. A more problem-oriented review of infinite-dimensional Bayesian statistical inverse problems can be found in [120], and reviews of statistically oriented inverse problems can be found in [16, 124]. A good reference to the computational aspects of finite-dimensional statistical inverse problems is [65], to Bayesian statistics [40, 105, 111], and to measure theory [10].

The statistical background of the statistical inverse theory belongs to the field of nonparametric Bayesian inference. Nonparametric statistics is concerned with making inferences about infinite-dimensional unknowns whereas parametric statistics studies finite-dimensional unknowns [6]. The function-valued prior models in
statistical inverse problems are therefore well within the scope of nonparametric statistics. We briefly review Bayesian nonparametric statistics and clarify its relations to statistical inverse problems.

Important nonparametric problems are the density estimation problem and the regression problem [91]. These two problems have guided the modern development of Bayesian nonparametric statistics.

In the density estimation problem, the observations are i.i.d. samples obeying some unknown probability distribution that has a density function \( f \) (usually on \( \mathbb{R} \)), and the objective is to estimate the density function \( f \). This problem is not directly related to our statistical inverse problem but is connected to the general development of the research field. It should be mentioned that Wolpert et al [142, 143] have described a semidiscrete Fredholm integral equation of the first kind as a Bayesian density estimation problem. On the other hand, the positron emission tomography (PET) imaging is an inverse problem that is usually described as a special density estimation problem where only indirect samples are available [62]. Hence, certain inverse problems lead to density estimation problems.

In the regression problem, the observations are of the type

\[
y_i = K(x_i) + \varepsilon_i.
\]

where \( x_i \in \mathbb{R}^n, i = 1, ..., n \), and the noise terms \( \varepsilon_i \) are typically independent and identically distributed. The objective is to estimate the unknown function \( K \). This problem has connections to statistical inverse problems. For example, if the realizations of \( X \) and \( Y = L(X) + \varepsilon \) are functions on \( \mathbb{R} \) and \( L \) is the identity mapping, then \( X \) is identified as \( K \).

One difference between the density estimation problem and the regression problem is the nature of given samples. In the regression problem, a single sample can also be infinite-dimensional, at least in theory. When the noise is Gaussian, such infinite-dimensional observation models are often called white noise models (see the short review in [147]).

The main questions in Bayesian nonparametrics have been the construction of the prior models, the utilization of the posterior distribution, and the consistency of the posterior distributions.

1.3.1. Prior models. The problem of finding good infinite-dimensional prior models has a long history. An early application of a function-valued prior model was carried out in 1896 by Poincaré [101], who applied a random series of the type

\[
X(t) = \sum_{i=1}^{\infty} X_i t^i
\]

in a regression problem on [0, 1]. He assumed independent normal distributions on coefficients \( X_i \) and calculated the posterior mean estimate on the basis of the given values \( y_i = X(t_i), i = 1, ..., n \). In Section V of Chapter 11 in [100] Poincaré discussed, in his visionary manner, the noisy regression problem. He proposed that the smoothness of the regression curve follows from the prior information of the unknown curve described in the form of probability distributions. In 1950, Grenander [50] applied a Gaussian process prior in a linear regression problem with additive Gaussian process noise. In 1957-58 Whittle [136, 137] discussed prior information on the smoothness of the unknown in certain density estimation problems, and later Kimeldorf and Wahba [71] clarified the relations between smoothing and Gaussian prior models. Nowadays, regularity of functions is one of the most important guidelines in constructing infinite-dimensional prior models in statistical inverse problems. This follows from the fact that the priors in statistical inverse
problems have two objectives. They express the prior beliefs about the unknown and are countermeasures against the ill-posedness of the deterministic inverse problem.

In general, the knowledge on infinite-dimensional random variables (and on their distributions) started to increase after Wiener published his construction of the Brownian motion in the beginning of 1920’s [139]. A decade later, Kolmogorov [73] introduced a constructive method for defining general infinite-dimensional random variables in the abstract setting. His method suited well for countably many random variables, but Doob noticed that the constructed $\sigma$-algebra was somewhat limited: for continuous parameter processes certain interesting sets, such as the set of all continuous functions, were not measurable with respect to the constructed $\sigma$-algebra. Doob’s remedy was the careful definition of the continuous-parameter stochastic processes in 1937 [31]. The theory of stochastic processes Doob’s definition of the separable stochastic processes provides the tools but not immediate answers for certain questions in statistical inverse problems. Namely, can the stochastic process be interpreted as a function-valued random variable that has values in some nice function space? The question is quite relevant since the direct theory is a mapping between two function spaces. One can e.g. apply Kolmogorov’s continuity theorem (proven by Kolmogorov in 1934, see [117]). Another approach is to directly define probability measures on function spaces. Jessen [59] carried out integration on infinite-dimensional dimensional torus equipped with the coordinate-wise convergence. M. Fréchet initiated the study of random variables in metric spaces (see [44]). His emphasis was on different modes of convergence and typical values of random variables, like the mean and the median. Significant contributions to the theory of probability measures on topological spaces were given by Alexandrov and Prohorov (see [132]). Later developments can be found in the books of Bogachev [9, 10], Gelfand and Vilenkin [46], Gihman and Skorokhod [48], Kahane [63], Kuo [78], Ledoux and Talagrand [83], Schwartz [112], Vakhania [130], Xia [144], and Yamasaki [146]. Typical points discussed in these books are the existence of measures, invariance properties of measures, topological supports, the equivalence and the equality of measures, the convergence of random series and the convergence of measures, all relevant properties for prior distributions. The existence of measures is often based on the Bochner-Minlos theorem that gives conditions for the one-to-one correspondence between measures $\mu$ and their characteristic functions $L(\phi) = \int e^{i\langle x, \phi \rangle} d\mu(x)$ on certain spaces. From the point of view of statistical inverse problems, it is unfortunate that direct connections between the characteristic function and the included prior information are not known. Therefore, it is no wonder that popular prior models have been described by other means, for example with infinite product measures and random series expansions. The works of Karhunen [69] in 1940’s on a series expansions of Gaussian random variables, nowadays known as the Karhunen-Loève expansion, are in this sense important. The Karhunen-Loève expansion was first used in 1950’s and 1960’s for expanding infinite-dimensional data [25, 50, 70], which made the Bayesian method of conditional mean estimation and the non-Bayesian method of likelihood ratio testing tractable. It was later adopted to describing infinite-dimensional unknowns (see for example [24]), but its main application has been in providing finite-dimensional approximations of Gaussian random variables. At present, other orthogonal expansions of Gaussian random variables are available [9]. The pioneering work of Mandelbaum [90] from 1984 on linear Gaussian statistical inverse problems relies on such series expansions of the
Gaussian random variables. Other works on Gaussian priors in statistical inverse problems are [41, 80, 85, 87, 116].

In 1963, Freedman introduced the class of tail-free priors for the density estimation problem [45]. In 1970’s the density estimation and the regression problem evolved further in different directions. Wahba et al [71, 133] took the approach with smoothing splines and Gaussian random series in the regression problem, and Ferguson [38] constructed Dirichlet process priors, which are certain random measure-valued unknowns, for the density estimation problem. In the case of Dirichlet processes, the space of the unknowns is the space of all probability measures on the fixed measure space equipped with the Borel $\sigma$-algebra with respect to the weak topology of measures. The Dirichlet process priors have similar properties in the density estimation problem as Gaussian priors have in the linear statistical inverse problems. Namely, the posterior distribution is the distribution of another Dirichlet process with updated parameters. In the both cases, the calculations of the posterior distribution are based on similar elements, which are the properties of the finite-dimensional distributions and the properties of the martingales [38, 90]. The Dirichlet process priors were generalized later to mixtures of Dirichlet processes (see [36]). Summaries of the prior distributions applied in modern density estimation problems can be found in [18, 134].

In 1990, Steinberg [118] suggested a prior model defined as a random series in which Hermite polynomials were multiplied by either improper or Gaussian coefficients. During 1998-2000, Abramovich et al [1, 2, 3] suggested random wavelet expansions in Besov spaces, with hierarchical coefficients whose hyperparameters guaranteed the sparseness of the expansions, as priors for the regression problem. In 1990’s also mixtures of Gaussian measures were suggested as priors for the regression problem (see [147]). Recently, Lassas et al [81] and Helin [54] constructed non-Gaussian edge-preserving priors suitable for statistical inverse problems. Besov space priors introduced by Lassas et al are defined with random wavelet expansions and Helin’s hierarchical prior distributions as mixtures of Gaussian measures. In 2010, Stuart [120] applied prior distributions of the type $f(x)\mu(dx)$, where $f \in L^1(\mu)$ and $\mu$ is a Gaussian measure. In the abstract setting, the statistical inverse theory was applied for unknowns described as Souslin space-valued random variables in [99].

It should be mentioned that some combinations of prior informations do not have faithful probabilistic descriptions. In 1987, Backus [4] pointed out that hard constraints, such as the boundedness of an infinite-dimensional random variable $X$ in norm, can lead to troubles if one assumes also isotropy. A well-known example is a Gaussian random variable $X$ that is invariant with respect to rotations (e.g. orthogonal transformations) on an infinite-dimensional separable Hilbert space $H$ but satisfies $\|X\|_H = \infty$ with probability one [9].

1.3.2. Utilization of the posterior distribution. We first look at the history of infinite-dimensional posterior distributions in nonparametric statistics and in statistical inverse problems.

In Bayesian nonparametrics, the both problems, the density estimation and the regression problem, are solved with conditional probability measures. This part of the solution mechanism is exactly the same as in statistical inverse theory. In 1930’s, the rigorous definition of the conditional expectation by Kolmogorov [73] made it possible to define conditional probability measures in the abstract infinite-dimensional setting but it was soon noted that such conditioning did not always
produce a probability measure. The works of Doob [32] and Dieudonné [29, 30] lead to the definition of a regular conditional probability, which is a random probability measure with probability one. The existence of regular versions of all conditional probabilities was verified by applying certain properties of the space of the unknowns in the works of Rohlin [106], Jiřina [60, 61], and Sazonov [110]. Nowadays, one either checks the properties of the space of the unknowns (as in [99]) or checks always the regularity of the acquired conditional measure for the chosen prior distribution (as in [38]). The former is used in theoretical studies for avoiding pathological cases [54, 80, 81, 99] while the latter is convenient in practical solutions where a fixed version is needed [22, 120]. We remark that the non-existence of a regular version is known only for some conditional measures in exemplifying cases (see [10, 103]).

A major step for the statistical inference for stochastic processes was the emergence of the so-called filtering problems in 1940’s by Wiener [138], Kolmogorov [74] and Krein [75, 76]. Especially, Wiener’s [138] straightforward method of solution (by ergodicity and least squares estimation) encouraged others to take later further steps towards the Bayesian nonparametric approach [42, 50, 137]. A good review on developments in the filtering theory is [64]. An interesting work in the filtering theory is [7], where it is shown that the solution of the filtering problem depends continuously on the distribution of the unknown. A nice collection of nonlinear filtering problems with Gaussian noise can be found in [89].

The first deliberate unions of inverse problems and Bayesian statistics were seen in 1960’s in the form of statistical regularization i.e. minimum mean squared error estimation for the Gaussian linear inverse problem

\[ Y = LX + \varepsilon \]

with the finite-dimensional unknown \( X \) and the finite-dimensional observation \( Y \). That is, one pursues after the estimator \( \hat{X}(Y) \) that minimizes \( E[|\hat{X} - X|^2] \) (i.e. the conditional mean). In 1961, motivated by Wiener’s filtering theory, Foster [42] presented a solution to the estimation problem (5). Other motivation for the Gaussian approach arose from the regularization method of Philips [98] generalized later by Twomey for Fredholm integral equations of the first kind [127] and from the Tikhonov regularization method. During 1967-71 Turchin et al (see [126] and references therein), independently with Strand and Westwater [119], replaced the regularization method by a statistical framework that utilized a Gaussian prior distribution. The approach lead to Franklin’s infinite-dimensional description [43] of the minimum mean squared error estimator of a Hilbert space-valued Gaussian unknown whose linear observations were corrupted by an additive Gaussian white noise. The connection between [43] and regularization methods in reproducing Hilbert spaces were studied by Prenter and Vogel [102]. The first work that contained the existence of regular conditional probabilities and an explicit formula for the posterior distribution in a linear infinite-dimensional inverse problem was the seminal paper of Mandelbaum [90] on Hilbert space-valued Gaussian random variables. The value of the result for inverse problems was first recognized by Lehtinen et al [85] who generalized it for the Gaussian (Schwartz) distribution-valued random variables. This work of Lehtinen et al can be considered as the starting point of the infinite-dimensional Bayesian inverse problems. The case of Banach space-valued Gaussian random variables was later considered by Luschgy [87]. In these works, the expression of the posterior mean is obtained by using the equivalence between statistical independence of Gaussian random variables and their orthogonality in
The key factor is the orthogonal random series expansion of the Gaussian observation – a method used by Grenander [50], and even by Poincaré [101]. Cox [24] applied Gaussian separable Banach space-valued unknowns in a linear regression problem with additive Gaussian noise. The approach of Cox differs from that of Mandelbaum since it uses the generalized Bayes formula rather than the special properties of Gaussian random variables (see Proposition 2.1 in [24]). An abstract formulation of Bayesian statistical inverse problems for Souslin space-valued random variables was given by Piiroinen [99], who only required the observation and the unknown to be Souslin space-valued random variables, thus allowing nonlinear direct problems and more complicated noise terms.

Little is known about the form of posterior distributions in infinite-dimensional statistical inverse problems outside the Gaussian linear case [80, 85, 87, 90, 116] and the dominated case with Gaussian noise [54, 81, 147]. When \( F \) and \( G \) are complete separable metric spaces, a result of Macci [88] tells that the Lebesgue decomposition of the posterior distribution with respect to the prior distribution contains a nontrivial singular part in undominated cases. Namely, the Lebesgue decomposition of the posterior distribution with respect to the prior distribution is of the form

\[
\mu(U, y) = \mu^{(ac)}(U, y) + \mu^{(s)}(U, y),
\]

where the absolutely continuous part \( \mu^{(ac)}(U, y) \) is determined by the absolutely continuous part of \( \mu_{Y|X}(\cdot, x) \) with respect to \( \mu_Y \) through the equations

\[
\mu^{(ac)}(U, y) = \int_U d\mu^{(ac)}_{Y|X}(\cdot, x) d\mu_Y(x),
\]

where \( U \in \mathcal{F} \). Moreover, the singular part \( \mu^{(s)}(U, y) \) is determined by the singular part of \( \mu_{Y|X}(\cdot, x) \) with respect to \( \mu_Y \) through the equations

\[
\mu^{(s)}(U, y) = \frac{d\left(\int_U \mu^{(s)}_{Y|X}(\cdot, x) d\mu_X\right)}{d\mu_Y}(y), \quad U \in \mathcal{F},
\]

from which one chooses a regular version. We remark that in such undominated cases, one may expect to meet some surprises. The posterior distribution presents then some things that seemed to be a priori impossible.

The extraction of information from the posterior distribution involves decision theory, including point estimation and hypothesis testing (for the general description of the Bayesian decision theory, see [111]). A decision theoretic view towards Bayesian inversion is given in [35], where one performs the estimation of the (separable Banach space-valued) unknown by first fixing the prior distribution of the unknown \( X \) and choosing a loss function \( \ell : F \times F \rightarrow \mathbb{R} \) that penalizes the inaccuracies in the estimates of the unknown, and then choosing the so-called Bayes estimator \( \hat{X} = X \rightarrow F \), which is a deterministic function that gives the smallest averaged loss \( \mathbb{E}[\ell(\hat{X}(Y), X)] \). This is equivalent to taking as each \( \hat{X}(Y(\omega)) \) the value \( d \) that minimizes the posterior expected loss \( \mathbb{E}[\ell(d, X)\sigma(Y)](\omega) \).

Common point estimators in finite-dimensional statistical inverse problems are the maximum a posteriori (MAP) estimator and the conditional mean (CM) estimator (i.e. the posterior mean) [65]. The CM estimator \( \hat{X}(Y) = \mathbb{E}[X|\sigma(Y)] \) minimizes the posterior risk for the squared error loss function \( \ell(x', x) = |x' - x|^2 \) (when \( X : \Omega \rightarrow \mathbb{R}^n \) is suitably integrable) [65]. Conditional means have appeared also in the framework of infinite-dimensional statistical inverse problems.
The mean of a locally convex Hausdorff topological vector space-valued random variable can arise from different definitions, depending on the space in question. In general, the (weak) mean of a locally convex Hausdorff topological vector space-valued random variable $X$ is a vector $m \in F''$ (or more generally, $m$ in the algebraic dual space of $F'$) such that $\langle m, \phi \rangle_{F'',F'} = \mathbb{E}[\langle X, \phi \rangle_{F,F}]$ for all $\phi \in F' P$-a.s. (see [9]). Such notion of vector-valued integration was developed by Pettis [97] in 1933 for reflexive separable Banach spaces $F$. Gelfand used a similar definition for distribution-valued random variables (see [46]). The Pettis-Gelfand integral was generalized for quasi-complete Souslin space-valued functions by Thomas [125]. For Banach-space valued random variables having integrable norm, a mean can be defined also as the Bochner integral $m = \int_F x d\mu_X(x)$, introduced in early 1930’s by Bochner (see [28] and references therein).

When the posterior distribution $\mu(\cdot, y)$ is known for a given sample $y$ of $Y$, the (weak) conditional mean $m \in F''$ is a vector that satisfies $\langle m, \phi \rangle_{F'',F'} = \int \langle x, \phi \rangle_{F,F} \mu(dx, y)$ for all $\phi \in F'$. When $F$ is a separable reflexive Banach space and $\|X\|$ is integrable, the same posterior mean can also be defined as the Bochner integral (see Proposition V.2.5 in [93]).

We remark that the weak posterior mean $\mathbb{E}[X|\sigma(Y)]$ is a Bayes estimator in a weak sense i.e. it gives the smallest averaged loss for the family of loss functions $\ell_{\phi}(x, x') = |\langle x - x', \phi \rangle|^2$, where $\phi \in F'$. Franklin [43] used such requirement, when he defined the best linear estimator in a Gaussian linear inverse problem in 1970. An earlier approach to the best linear estimator in function-valued Gaussian case was given by Grenander in 1950 (see Chapter 6 in [50]). He considered a Gaussian linear regression problem and identified the best linear estimator (with respect to the pointwise squared error loss $\ell(X(t), \hat{X}(t)) = |X(t) - \hat{X}(t)|^2$) with the posterior mean. Grenander used infinite-dimensional observations, but he made simultaneous inferences on only finitely many values of the unknown function. Moreover, he required, but not proved, the regularity of the conditional probabilities. In this sense, his approach to the posterior means was still far from the description of Mandelbaum from 1984 [90]. Remark, that the technique of estimating the value of $X(t)$ on the basis of infinitely many observations is still the standard in the modern filtering theory [94]. Also in the Bayesian density estimation, the estimation is sometimes carried out in either in the form $\hat{X}(t) = \mathbb{E}[X(t)|Y_1, ..., Y_n]$, where $X$ is the unknown probability density function on, say $[0,1]$ (see [137]), or in the form $\hat{X}(U) = \mathbb{E}[X(U)|Y_1, ..., Y_n]$, where the sets $U \subseteq [0,1]$ are Borel set and $X$ is an unknown random probability measure (see [91] and Proposition 4.2.1 in [47]). The density estimator $\hat{X}$ is a Bayes estimator with respect to the squared error loss function for each $t$ or for each Borel set $U$, respectively. An other option is to use a weighted $L^2$-loss function $\ell(\hat{X}, X) = \int_0^\infty |\hat{X}(t) - X(t)|^2 d\mu(t)$ [38]. The two estimators coincide when $X$ is suitably integrable.

In the works of Mandelbaum [90] and Luschgy [87], the space $F$ is a Hilbert or Banach space, and the posterior mean is defined as a Bochner integral. However, the emphasis is on the Gaussian nature of the prior, and the posterior mean is calculated as $\mathbb{E}[\sum_{i=1}^\infty X_i e_i|\sigma(Y)] = \sum_{i=1}^\infty \mathbb{E}[X_i|\sigma(Y)]e_i$. Similar approach appears in [80, 85] for the distribution space, where the posterior mean is defined in the weak sense.
The weak definition of the posterior mean is used also in [82] for the space \( C([0,1]) \). In [54, 81], the conditional mean of a separable Banach space-valued random variable is defined as a Bochner integral with respect to the posterior distribution. Before Luschgy, Krug [77] determined the posterior mean of a separable Banach space-valued Gaussian unknown in a linear Gaussian case, but he assumed that the given observation was finite-dimensional.

We remark that when \( F \) is a Hilbert space, one can take \( \ell(x',x) = \|x - x'\|^2_F \) as the loss function that gives the CM estimator. As in the finite-dimensional case, the main point is that

\[
E[\|\hat{X}(Y) - X\|^2_F] = E[\|\hat{X}(Y) - E[X|\sigma(Y)]\|^2_F] + E[(\hat{X}(Y) - E[X|\sigma(Y)], E[X|\sigma(Y)] - X)_F] + E[\|X\|^2_F],
\]

and the additional difficulty is just in checking that \( E[(f,X)|\sigma(Y)] = (f, E[X|\sigma(Y)]) \). Such loss functions have been used in the regression problem for the Gaussian mixture priors when \( F = L^2([-1,1]) \) [147]. Instead of an \( L^2 \)-loss function, Abramovich et al [1] used an \( L^1 \)-loss function in a regression problem for a discretized Besov space-valued unknown. We note that a common approach in the regression problem is to present only the Bayes estimates instead of the whole posterior distribution.

Luschgy [87] made an (unproven) remark that for Gaussian posterior distributions the conditional mean is the Bayes estimator for every symmetric quasi-convex (measurable) loss function \( \ell(x,x') = \ell(x - x') \). A proof can be found in [12], where it is derived from the Anderson property of Gaussian measures (for the property, see [86]).

In finite-dimensional spaces, the MAP estimator can be interpreted as a limit of Bayes estimators for the 0-1-valued losses \( \ell_\epsilon(x',x) = 1_{F \setminus \overline{B}(x,\epsilon)}(x') \), where \( \epsilon \to 0 \) [105]. Here \( \overline{B}(x,\epsilon) \) is the closed ball in \( F \) that is centered at \( x \) and has radius \( \epsilon \). Lassas and Siltanen [82] showed that MAP estimates can behave inconsistently as dimensionality of the unknown increases, even though the posterior distributions converge at the same time. In their example, the MAP estimates actually vanish at the limit, regardless of the given observation. Similar result is proved in [55] for a hierarchical edge-preserving prior. The MAP and CM estimates coincide for the finite-dimensional Gaussian priors, and numerical results demonstrate that they can practically coincide for the finite-dimensional approximations of Besov-priors [72]. Cotter et al [22] discussed MAP estimation in the context of infinite-dimensional Bayesian problems. They showed that there exists a minimizer for a penalized log-likelihood function, which has similar form as in the case of finite-dimensional Gaussian unknown. However, the conditions that would relate the penalized log-likelihood function to any posterior density were omitted in [22], which leaves open the question what connections the minimizer has to the infinite-dimensional posterior distributions. Recalling the result of Lassas and Siltanen [82] arises at least some caution. Another attempt towards MAP estimation with infinite-dimensional Gaussian priors is given by Hegland [53]. Unfortunately, the proof of Proposition 1 in [53] is not rigorous, as it involves subtraction of two numbers that are infinitely large with probability 1 (i.e. the Cameron-Martin norms of arbitrary vectors in the space of the unknowns).

In infinite-dimensional statistical inverse problems the hypothesis testing has been largely neglected, although several interesting questions could be raised. For example, Fitzpatrick [39] has made an initiative on testing if the evidence supports
the homogeneity of the unknown diffusion coefficient. Hypothesis testing was proposed also for some nonparametric statistical inverse problems in [8] within the classical framework. However, it was pointed out in [8] that the problems can be similarly handled also by the (finite-dimensional) Bayesian methods but this remark is not elaborated further.

Another approach to exploiting posterior distributions was given by Piiroinen [99]. He interpret the posterior distributions as statistical measurements, which allowed comparisons of information contents of different posterior distributions. The result is especially useful in experimental design [84].

1.3.3. Posterior consistency. The consistency of the posterior distributions (with respect to repeated independent observations) is closely connected to the uniqueness of the deterministic inverse problem of determining \( x \) from \( y = L(x) \). The pioneering work of Doob [33] on martingales touched the question of consistency of the posterior distributions. Doob’s results imply that under model identifiability (i.e. the measures \( \mu_{x+L(x)} \) are different for different \( x \in F \)) the posterior distributions would concentrate (in the weak topology of measures) on the true unknown \( x_0 \) \( \mu_{L(x_0)+\varepsilon} \)-almost surely for \( \mu_X \)-a.e. \( x_0 \) when infinitely many i.i.d. observations would be available. The consistency of the posterior distribution is an important topic because it shows that enough data will guide a Bayesian scientist almost surely to the true answer. The words \( \mu_X \)-a.s. made Doob’s approach slightly impractical as they left open the frequentist case where the observations are not samples of \( L(X)+\varepsilon \) but samples of \( L(x)+\varepsilon \) for some fixed \( x \). Freedman [45] demonstrated that inconsistency could hold on topologically large sets. The problem was approached by Schwartz [113] who described a set of unknowns \( x \) for which consistency holds \( \mu_{x+L(x)} \)-almost everywhere under some decision theoretic conditions and domination (i.e. all measures \( \{\mu_{x+L(x)} : x \in F\} \) are assumed to be absolutely continuous with respect to some common \( \sigma \)-finite measure). The required property is the positive prior probability of all Kullback-Leibler neighborhoods of the unknown \( x \). Consistency has been studied also in other topologies, beside of the weak topology. Barron et al [5] proved consistency in Hellinger distance. Summaries of consistency result in density estimation can be found in [26, 27, 145]. The case of Gaussian regression has been studied in [131], where certain probabilities are shown to converge.

Consistency issues in inverse problems are discussed in [35]. For our statistical inverse problem, the consistency corresponds to observing one sample of \( Y = L(X) + \frac{1}{\sqrt{n}}\varepsilon \), where \( n \) represents the number of i.i.d. observations of \( L(x)+\varepsilon \). The works of Hofinger and Pikkarainen [57, 58], and Neubauer and Pikkarainen [92] on finite-dimensional Gaussian statistical inverse problems concern the question of posterior consistency. They studied the convergence of posterior distributions and the posterior means in linear Gaussian inverse problems for finite-dimensional random variables as the variance of the noise decreases. In particular, it was shown in [57] that the posterior distributions given observed values \( \tilde{Y}_{\delta_n} = Lx + \delta_n\varepsilon(\omega) \) of \( Y_{\delta_n} = LX + \delta_n\varepsilon \) for a sequence \( \delta_n \to 0 \), converge to the point mass on the true value \( x \) in the Ky Fan metric, assuming that also the prior distribution are modified appropriately. Hofinger and Pikkarainen [92] studied posterior convergence rates for finite-dimensional approximations of Hilbert-space-valued random variables when the approximation level increases in certain manner as the noise level \( \delta_n \) approaches to zero. However, the convergence was shown only for unknowns in an \( a \ priori \) zero measurable set (the Cameron-Martin space of the prior...
distribution). Also Florens and Simoni [41] studied the posterior consistency for the infinite-dimensional linear Gaussian inverse problems when the variance of the noise diminishes. They were able to show the posterior consistency if the posterior measures with respect to the weak topology (and give estimates for the speed of convergence of the posterior means) by assuming that the direct theory is regular enough and the prior distribution depends suitably on the noise level.

Another convergence topic that has received more attention in statistical inverse problems is the posterior convergence for approximated unknowns and observations [54, 80, 81, 82, 99, 120]. This case is discussed in Part II. A concept close to the posterior convergence is the so-called discretization invariance, which was first used by Markku Lehtinen in 1990’s (see [81]). It asks that the prior knowledge is consistent at all discretization levels and aims to the stability of posterior knowledge on different discretization levels. Definitions for discretization invariance in statistical inverse problems are given in [81, 82]. In [81], Lassas et al defined a proper linear discretization \( X_n = P_nX \) of a Banach space-valued random variables \( X \), where \( P_n \) are bounded linear operators on the Banach space \( F \) having finite-dimensional ranges and the random variables \( \langle P_nX, \phi \rangle \) converge in distribution to \( \langle X, \phi \rangle \) for all \( \phi \in F' \). Gaussian priors and Besov space priors were shown to be discretization invariant in [81] in the sense that they have proper linear discretization for which the conditional mean estimates converge. An important example was studied by Lassas and Siltanen [82] who showed that the finite dimensional total variation priors converge to a Gaussian measure and the corresponding CM estimates converge to the CM estimate obtained with a Gaussian prior. The total variation priors are not discretization invariant as the finite-dimensional prior distributions lead to unwanted effects. A special method for obtaining stable posterior knowledge was suggested by Kaipio and Somersalo [66], who proposed the approximation error approach for statistical inverse problems. In approximation error approach, the conditioning random variable \( Y = L(X) + \varepsilon \) is written as \( Y = L(X_n) + (L(X) - L(X_n)) + \varepsilon \), where \( L(X) - L(X_n) \) is taken to be an additional noise term \( \tilde{\varepsilon} \). For example, if \( X \) is Gaussian and \( X_n = P_nX \), where \( P_n \) are linear projection operators, the CM estimators take a consistent form \( E[X_n | Y] = P_nE[X | Y] \). The problem becomes computationally more tractable if \( X_n \) and \( \tilde{\varepsilon} \) are statistically independent in which case only the distribution of \( \tilde{\varepsilon} \) needs to be additionally determined. This condition is often forced on \( \tilde{\varepsilon} \) together with a numerically feasible approximated distribution [66, 123].

2. Conditional probabilities and posterior distributions.

2.1. Solution of the statistical inverse problem. We define what we exactly mean by a statistical inverse problem and its solution. We begin by recalling the definition of the conditional expectation.

The conditional expectation of \( f \in L^1(\Omega, \Sigma, P) \) given a sub-\( \sigma \)-algebra \( \Sigma_0 \subset \Sigma \) is a \( \Sigma_0 \)-measurable function \( E[f | \Sigma_0] \) such that

\[
\int_A f dP = \int_A E[f | \Sigma_0] dP
\]

for all \( A \in \Sigma_0 \). Conditional expectations exist due to the Radon-Nikodym theorem as the densities of the (signed) measure \( f dP \) with respect to the measure \( P \) on \( \Sigma_0 \), but they are only defined up to sets \( N \in \Sigma_0 \) of \( P \)-measure zero. We denote \( E[|Y|] = E[|Y^{-1}(\mathcal{G})|] \).
Definition 2.1. Let \((\Omega, \Sigma, P)\) be a complete probability space. Let \(F\) and \(G\) be two Souslin spaces equipped with their Borel \(\sigma\)-algebras \(\mathcal{F}\) and \(\mathcal{G}\), respectively. Let \(X : \Omega \to F\) and \(Y : \Omega \to G\) be measurable mappings. We call a mapping \(\mu : \mathcal{F} \times \mathcal{G} \to [0, 1]\) a solution of the statistical inverse problem of estimating the distribution of the unknown \(X\) given the observation \(Y\) if

1. \(\mu(U, Y(\omega)) = E[1_U(X)Y(\omega)]\) \(P\)-almost surely for every \(U \in \mathcal{F}\),
2. \(y \mapsto \mu(U, y)\) is \(\mu_Y\)-measurable for every \(U \in \mathcal{F}\), and
3. \(U \mapsto \mu(U, y)\) is a probability measure on \((F, \mathcal{F})\) for every \(y \in G\).

The distributions \(\mu(\cdot, y)\) are called posterior distributions of \(X\) given \(Y = y\).

Strictly speaking, the posterior distributions are defined \(a posteriori\) of the observation \(Y(\omega)\) but we feel that there is no harm in calling \(\mu(\cdot, y)\) posterior distributions also for \(y \notin R(Y)\) since \(\mu_Y(G) = 1\).

The solution is just a regular conditional distribution of \(X\) given the sub-\(\sigma\)-algebra \(Y^{-1}(G)\), where the regularity holds in the sense of Doob i.e. the solution \(\mu\) is \(\mu_Y\)-measurable in the second variable (see Remark 10.6.3 in [10] for a further discussion). The nature of the mapping \(\omega \mapsto \mu(U, Y(\omega))\), which need not be \(\sigma(Y)\)-measurable, is verified in the following simple lemma.

Lemma 2.2. Let \((G, \mathcal{G})\) be a measurable space. Let \(Y : \Omega \to G\) be a measurable mapping from a complete probability space \((\Omega, \Sigma, P)\) into \(G\). If \(f : G \to \mathbb{R}\) is a \(\mu_Y\)-measurable function then \(f(Y)\) is a random variable on \((\Omega, \Sigma, P)\), and \(E[f(Y)] = \int f(y) d\mu_Y(y)\). Moreover, if \(\tilde{f} : G \to \mathbb{R}\) is a Borel measurable function such that \(f = f\ \mu_Y\text{-a.s.},\) then \(E[f(Y)] = \tilde{f}(Y(\omega))\) \(P\)-almost surely and \(E[f(Y)|\Sigma_0](\omega) = E[\tilde{f}(Y)|\Sigma_0](\omega)\) \(P\)-almost surely for any sub-\(\sigma\)-algebra \(\Sigma_0 \subset \Sigma\).

Proof. Every \(\mu_Y\)-measurable function has a Borel measurable version (see Proposition 2.1.11 in [10]). Denote with \(\tilde{f}\) a Borel measurable version of \(f\). The set \(N = \{y \in G : f(y) \neq \tilde{f}(y)\} \in \mathcal{G}^{\mu_Y}\) is then \(\mu_Y\)-zero measurable and, by definition, there exists a \(\mu_Y\)-zero measurable Borel set \(B \in \mathcal{G}\) such that \(N \subset B\). Especially, \(Y^{-1}(N) \subset Y^{-1}(B)\), which has \(P\)-measure \(P(Y^{-1}(B)) = \mu_Y(B) = 0\) so that also \(Y^{-1}(N)\) belongs to the complete \(\sigma\)-algebra \(\Sigma\) and has \(P\)-measure zero. Therefore,

\[f(Y(\omega)) = \tilde{f}(Y(\omega))\]

\(P\)-almost surely. By the completeness of \(\Sigma\), also the mapping \(\omega \mapsto f(Y(\omega))\) is \(\Sigma\)-measurable. By the almost sure equivalence of the functions, we get

\[E[f(Y(\omega))] = E[\tilde{f}(Y(\omega))] = \int \tilde{f}(y) d\mu_Y(y) = \int f(y) d\mu_Y(y).

The conditional expectations of equivalent random variables coincide, since they have the same integrals over \(\Sigma_0\)-measurable sets.

From the point of view of the posterior analysis, Condition 3 of Definition 2.1 may give a false sense of security. Any \(\mu\) that satisfies Conditions 1 and 2 but is a probability measure only for \(\mu_Y\text{-a.e.}\) \(y\) can be redefined on a negligible set in such a way that it is a solution. For example, if \(N\) is the \(\mu_Y\)-zero measurable set that contains all \(y\)'s for which \(\mu(\cdot, y)\) is not a probability measure, we may redefine \(\mu(U, y)\) as \(1_U(x_0)\) for some fixed \(x_0 \in F\) and all \(y \in N\). Then \(\mu\) satisfies Conditions 1, 2 and 3, but \(\mu(\cdot, y)\) is not related to the unknown when \(y \in N\).

We briefly compare the solution \(\mu\) with other formulations of Bayesian inverse problems. Clearly, any regular conditional distribution \(\mu\) of \(X\) given \(Y\) (such that
y \mapsto \mu(U, y) is Borel-measurable for any $U \in \mathcal{F}$ qualifies as a solution. Especially, posterior distributions obtained by the Bayes formula (1) on $\mathbb{R}^n$ for positive continuous probability densities form a solution of the form $\mu(U, y) = \int_U \mathbb{D}(\omega|y) dx$ [65]. The Gaussian conditional probabilities in [80, 85, 87, 90] are also solutions that are allowed to be $\mu_Y$-measurable in the sense of Condition 2. Our approach is similar to the work of Piiroinen [99], where a general formulation of the statistical inverse problem for Soussin space-valued random variables first appeared. The difference is that Piiroinen chose the posterior probabilities $\mu(U, y)$ to be universally measurable with respect to the second variable, that is, $m$-measurable for any finite Radon measure $m$ on $(G, \mathcal{G})$ whereas we prefer to take all $\mu_Y$-measurable versions as solutions, since it helps to avoid the somewhat artificial modifications of $\mu_Y$-measurable functions (encountered for example in the Gaussian case [85]) to any universally measurable or $\mathcal{G}$-measurable functions. Lassas et al [81] used a different approach where the posterior distribution was obtained by defining reconstructors.

A mapping $y \mapsto \mathcal{R}(g|y)$ is called a reconstructor of $g \in L^1(\mu_X)$ (more generally, a Bochner integrable $g$) given the observation $Y$ if $\mathcal{R}(g, Y(\omega)) = E[g(X)|Y^{-1}(\mathcal{G})](\omega)$ almost surely [81]. The concept of a reconstructor is more elemental than our solution. However, the reconstructors that were used for solving the statistical inverse problem in [54, 81] were chosen to be more regular. They depend continuously on observations and satisfy also Conditions 1 and 3. Hence, they form a regular conditional distribution and are especially solutions in the sense of Definition 2.1. A common point of the reconstructor and our solution is that both are defined conditional distribution and are especially solutions in the sense of Definition 2.1.

Moreover, the conditional expectation of the measurable function $\omega \mapsto \mathcal{R}(g|y)$ for all $g, \omega \in \mathcal{G}^\nu$ is $\mu_Y$-measurable on $(0,1]$. Take $(\Omega, \Sigma, P)$ to be the completion of $(U, \mathcal{B}_U([0,1]), m|_U)$ and $F = G = [0,1]$ equipped with its Borel $\sigma$-algebra. Let $Y : U \to [0,1]$ be the identity and take $X = Y$. Then the image measure $\mu_Y$ is the Lebesgue measure on $[0,1]$. Moreover, the conditional expectation of the measurable function $\omega \mapsto 1_{[0,1]}(X(\omega))$ given $\sigma(Y)$ is

$$E[1_{[0,1]}(X)|\sigma(Y)|(\omega) = 1_{[0,1]}(Y(\omega))$$

which is equal to $1_U(Y(\omega)) P$-almost surely. However, the reconstructor $\mathcal{R}_1(1_{[0,1]}, \cdot) : y \mapsto 1_U(y)$ is not $\mu_Y$-measurable on $([0,1], \mathcal{B}([0,1]))$, and the two reconstructors $\mathcal{R}_1(1_{[0,1]}, \cdot)$ and $\mathcal{R}_2(1_{[0,1]}, \cdot) : y \mapsto 1_{[0,1]}(y)$ differ on the set $N_0 := [0,1]|U$ which has positive $\mu_Y$-outer measure. Condition 2 helps us to avoid this small shortcoming. If the reconstructors are $\mu_Y$-measurable, the set $N_0 = \{y \in G : \mathcal{R}_1(f, \cdot) \neq \mathcal{R}_2(f, \cdot)\} \in \mathcal{G}^\nu$ and $\{\omega : \mathcal{R}_1(f, Y(\omega)) \neq \mathcal{R}_2(f, Y(\omega))\} = Y^{-1}(N_0)$. Then $N_0$ has zero $\mu_Y$-measure.
A regular conditional distribution is not unique in general because of the non-uniqueness of the conditional expectations. For our theoretical considerations, the following concept (adapted from [10] in context of regular conditional measures) is useful.

**Definition 2.3.** We say that a solution \( \mu \) of the statistical inverse problem of estimating the distribution of \( X \) given the observation \( Y \) is **essentially unique** if for any other solution \( \tilde{\mu} \) of the same statistical inverse problem there exists a set \( C = C(\mu, \tilde{\mu}) \in \mathcal{G}_{\mu Y}^\nu \) with \( \mu_Y(C) = 1 \) such that \( \tilde{\mu} \) agrees with \( \mu \) on \( \mathcal{F} \times C \). Similarly, we say that the posterior distribution \( \mu(\cdot, y) \) is essentially unique if \( \mu \) is essentially unique.

In other words, an essentially unique solution \( \mu \) may be arbitrary on the sets of the form \( \mathcal{F} \times N \), where \( N \subset G \) is a set of \( \mu_Y \)-measure zero. In a sense, this makes the posterior distribution \( \mu(\cdot, Y(\omega)) \) a relevant estimate of the distribution of \( X \) with probability 1.

Next, we recall some results on the existence and essential uniqueness of regular conditional distributions in Souslin spaces. The existence of regular conditional distributions of \( X \) given \( Y \) has been shown in Lemma 4.2 of [99] (by using the definition of the Souslin space and the existence of regular conditional distributions on Polish spaces, leading to a universally measurable kernel \( \mu \)), and in Example 10.7.5 of [10], where also the essential uniqueness has been verified. The present “extension” covers \( \mu_Y \)-measurable solutions. The condensed proof is included only to support the last sentence, which provides some motivation for the main results of this work. Namely, the definition of the conditional expectation may give the impression that we need to specify some random variable \( Y \) among all equivalent random variables for determining the conditional expectation of \( 1_U(X) \) when an observation \( y = Y(\omega_0) \in G \) has occurred. This is not true as a weaker description of \( Y \) and \( X \) suffices.

**Theorem 2.4.** Let \((F, \mathcal{F}) \) and \((G, \mathcal{G}) \) be two measurable spaces. Let \( X \) be an \( F \)-valued random variable and \( Y \) be a \( G \)-valued random variable on a complete probability space \((\Omega, \Sigma, P) \). If \( F \) and \( G \) are Souslin spaces equipped with their Borel \( \sigma \)-algebras, then there exists an essentially unique solution \( \mu : \mathcal{F} \times \mathcal{G} \rightarrow [0, 1] \) of the statistical inverse problem of estimating the distribution of the unknown \( X \) given the observation \( Y \).

The values \( \mu(U, y) \) are determined by the joint distribution \( \mu_{(X,Y)} \) of \( X \) and \( Y \) for all \( U \in \mathcal{F} \) and \( \mu_Y \)-almost every \( y \in G \).

**Proof.** First we show that for each \( U \in \mathcal{F} \) there exists a solution \( \mu_0(U, \cdot) : G \rightarrow [0, 1] \) such that \( y \mapsto \mu_0(U, y) \) is Borel-measurable and \( \omega \mapsto \mu_0(U, Y(\omega)) \) is a conditional expectation of \( 1_U(X) \) given \( \mathcal{G} \).

Consider the measure space \((F \times G, \mathcal{B}(F \times G), \mu_{(X,Y)}) \) and the sub-\( \sigma \)-algebra \( \mathcal{G}_0 = \{\emptyset, F\} \otimes \mathcal{G} \) generated by the canonical projection \( p_2(x, y) = y \) to the second variable. Recall, that the direct products of Souslin spaces are Souslin spaces. Due to the Souslin property of \( F \times G \), there exists a conditional measure \( \mu_0 : \mathcal{B}(F \times G) \times (F \times G) \rightarrow [0, 1] \) such that \( \mu_0(U', \cdot) \) is \( \mathcal{G}_0 \)-measurable for every \( U' \in \mathcal{B}(F \times G) \), the measure \( \mu(\cdot, (x, y)) \) is a probability distribution on \( \mathcal{B}(F \times G) \) for every \((x, y) \in G \), and

\[
\mu_{(X,Y)}(U' \cap V') = \int_{V'} \mu_0(U', (x, y))d\mu_{(X,Y)}(x, y)
\]
for every \( U' \in \mathcal{B}(\mathcal{F} \times G) \) and \( V' \in \mathcal{G}_0 \) by Corollary 10.4.6 in [10]. Let us restrict \( \mu_0(U', (x, y)) \) on sets \( U' \) of the form \( U \times G \), where \( U \in \mathcal{F} \). Since \( \mathcal{G}_0 \) is trivial with respect to the first variable, we may denote the restriction with \( \mu_0(U, y) \) where \( U \in \mathcal{F} \) and \( y \in G \). Especially, \( y \mapsto \mu_0(U, y) \) is \( \mathcal{G} \)-measurable and

\[
P(X \in U \cap Y \in V) = \mu_{(X,Y)}(U \times V) = \int_V \mu_0(U, y) d\mu_Y(y) = \int_{Y^{-1}(V)} \mu_0(U, Y) dP
\]

for every \( U \in \mathcal{F} \) and \( V \in \mathcal{G} \). Therefore, \( \mu_0 : \mathcal{F} \times G \to [0,1] \) is a solution of the statistical inverse problem of estimating the probabilities of the unknown \( X \) given the observation \( Y \).

A solution \( \mu \) is essentially unique since the Borel \( \sigma \)-algebra of a Souslin space is countably generated (see [10]). Indeed, suppose that \( \mu \) and \( \nu \) are two solutions in the sense of Definition 2.1. For \( U \in \mathcal{F} \), we have that \( \mu(U, Y(\omega)) = \mathbb{E}[|_{U}(X)|Y^{-1}(G)](\omega) = \nu(U, Y(\omega)) \) \( P \)-almost surely. Then \( \mu(\cdot, y) = \nu(\cdot, y) \) outside some \( \mu_Y \)-zero measurable set \( N_U \in \mathcal{G}^\nu \), since

\[
0 = \mathbb{E}[|\tilde{\mu}(U, Y) - \tilde{\nu}(U, Y)|] = \int |\mu(U, y) - \nu(U, y)| d\mu_Y(y)
\]

by Lemma 2.2. Every countable algebra \( \mathcal{F}_0 \) that generates the \( \sigma \)-algebra \( \mathcal{F} \) is measure-determining, i.e. measures coinciding on \( \mathcal{F}_0 \) coincide on \( \mathcal{F} \) (e.g. Lemma 1.9.4 in [10]). Hence, the two solutions coincide except for \( y \in \cup_{U \in \mathcal{F}_0} N_U \).

Finally, if \( \mu \) is any solution then the values \( \mu(U, y) \) are determined by the measure \( \mu_{(X,Y)} \) for all \( U \in \mathcal{F} \) and \( \mu_Y \)-almost all \( y \in G \) since \( \mu \) coincides with \( \mu_0 \) on \( \mathcal{F} \times \mathcal{C} \) (by essential uniqueness) and the values of \( \mu_0(U, \cdot) \) are actually versions of the Radon-Nikodym densities of measures \( \mu_{(X,Y)}(U, \cdot) \) with respect to \( \mu_Y \) for \( \mu_Y \)-almost all \( y \). The distribution \( \mu_Y \) is the marginal of \( \mu_{(X,Y)} \).

We have reached the usual starting point of nonparametric Bayesian statistics. In a conventional Bayesian experiment, one specifies only conditional distributions of \( Y \) given \( X = x \) for all values \( x \in G \) – the so-called parametric family of distributions or sampling distributions – and the prior distribution \( \mu_X \) on \( (\mathcal{F}, \mathcal{F}) \) [40, 111], which together determine the joint distribution of \( X \) and \( Y \).

**Remark 1.** The choice of the sample space \( (G, \mathcal{G}) \) of random variable \( Y \) is usually not trivial. One might choose as well a larger (or sometimes even a smaller) space than \( G \). The solutions of the statistical inverse problem could, in principle, depend on the choice of the sample space \( (G, \mathcal{G}) \) since the conditioning \( \sigma \)-algebra \( Y^{-1}(\mathcal{G}) \) depends on the topology of \( G \). But since we are working with the Souslin spaces this is not the case. Indeed, if \( (G_1, \mathcal{G}_1) \) and \( (G_2, \mathcal{G}_2) \) are two Souslin spaces equipped with their Borel \( \sigma \)-algebras and \( i : G_1 \to G_2 \) is a continuous (or just Borel!) injection, then, quite remarkably, \( i^{-1}(\mathcal{G}_2) = \mathcal{G}_1 \). Indeed, \( i^{-1}(\mathcal{G}_2) \subset \mathcal{G}_1 \) by the continuity of \( i \). Moreover, the image of a Borel set under a Borel mapping between Souslin spaces is a Souslin set i.e. a Souslin space with respect to the relative topology (see Theorem 6.7.3 in [10]). Therefore, \( i(G_1), i(B) \) and \( i(G_1 \setminus B) \) are all Souslin sets in \( G_2 \) for any Borel set \( B \in \mathcal{G}_1 \). By injectivity, \( i(G_1 \setminus B) = i(G_1) \setminus i(B) \) i.e. the complement of the Souslin set \( i(B) \) in the subspace \( i(G_1) \) of \( G_2 \) is a Souslin set. By Corollary 6.6.10 in [10], a Souslin set in a Hausdorff space is a Borel set if also its complement is a Souslin set. Therefore, \( i(B) \) is a Borel set in the relative topology of \( i(G_1) \). But each Borel set \( i(B) \) in \( i(G_1) \) is of the form \( i(B) = i(G_1) \cap B' \), where \( B' \in \mathcal{G}_2 \). Therefore, \( B = i^{-1}(B') \) for some \( B' \in \mathcal{G}_2 \) which implies that \( \mathcal{G}_1 = i^{-1}(\mathcal{G}_2) \). Consequently, \( (Y^{-1}(\mathcal{G}_2) = Y^{-1}(\mathcal{G}_1) \) for any \( \mathcal{G}_1 \)-valued random variable.
Y. Therefore, \( \mu_1(\cdot, Y(\omega)) = \mu_2(\cdot, i(Y(\omega))) \) \( P \)-almost surely for any solutions \( \mu_1 \) and \( \mu_2 \) of the inverse problems of estimating the distribution of \( X \) given \( Y : \Omega \to G_1 \) and \( i(Y) \), respectively. If \( \tilde{\mu}_2 \) is a Borel measurable version of \( \mu_2 \), then \( \tilde{\mu}_2(U, i(y)) \) is \( G_1 \)-measurable and \( \mu_2(U, y') = \tilde{\mu}_2(U, y) \), except possibly on some set \( N \) such that \( \mu_i(\cdot)(N) = 0 \) which implies that \( \mu_2(U, i(y)) = \tilde{\mu}_2(U, i(y)) \) except possibly on the set \( i^{-1}(N) \) which has \( \mu_Y(i^{-1}(N)) = 0 \). Therefore, \( \mu_2(\cdot, i(\cdot)) \) is also a solution of the statistical inverse problem of estimating the distribution of \( X \) given \( Y \). We are allowed to mis-specify the Souslin sample space \( G_1 \) by Borel injections without altering the essentially unique solution. In the general case that involves non-Souslin spaces, we only know that \( i^{-1}(G_2) \subset G_1 \), where the inclusion may be strict. As an example, take \( G_1 \) and \( G_2 \) to be the sequence space \( \ell^\infty \) where we take \( G_1 \) to be the usual Borel \( \sigma \)-algebra with respect to the supremum norm topology (which is not separable) and \( G_2 \) to be the Borel \( \sigma \)-algebra with respect to the weak topology, and take \( i \) to be the identity. Then \( i^{-1}(G_2) \neq G_1 \) (see Proposition 2.9 in [130]).

2.2. Partial uniqueness of the solution. From practical point of view, the essential uniqueness is not enough since we are given some fixed observation \( y_0 \in G \) that might belong to the set where arbitrariness of \( \mu \) still rules. Our proposal for removing this deficiency of the posterior distributions is to proceed as in the finite-dimensional case, where \( \mu \) is required to depend continuously on the second variable i.e. the posterior distributions depend continuously on the observations. The following new concept turns out to be useful.

**Definition 2.5.** Let \((\Omega, \Sigma, P)\) be a complete probability space. Let \( F \) and \( G \) be two Souslin spaces equipped with their Borel \( \sigma \)-algebras \( \mathcal{F} \) and \( \mathcal{G} \), respectively. Let \( X : \Omega \to F \) and \( Y : \Omega \to G \) be measurable mappings. Let \( \mu \) be a solution of the statistical inverse problem of estimating the distribution of the unknown \( X \) given the observation \( Y \). Let \( A \subset G \) and let \( \mathcal{F}_0 \subset \mathcal{F} \). We say that a solution \( \mu \) is \( \mathcal{F}_0 \)-continuous on \( A \) if the mapping \( y \mapsto \mu(U, y) \) is continuous on \( A \) with respect to the relative topology for every \( U \in \mathcal{F}_0 \).

Consider a set \( S \subset G \) that contains every point \( y \in G \) whose any open neighborhood has positive \( \mu_Y \)-measure. On Souslin spaces such a set \( S \) is known to coincide with the topological support of \( \mu_Y \), i.e. the smallest closed set \( S \subset G \) such that \( \mu_Y(S) = 1 \).

**Lemma 2.6.** Let \( \nu \) be a Borel probability measure on a Souslin space \( G \). The topological support of \( \nu \) exists and it consists of exactly those \( y \in G \) whose every open neighborhood has positive measure.

**Proof.** See Theorem 2.1 in [96], which generalizes to Souslin spaces, since Souslin spaces are hereditarily Lindelöf by Lemma 6.6.4 of [10].

We obtain partial uniqueness of the posterior distributions by using the continuity of solutions on certain subsets of the topological support of \( \mu_Y \). We denote with \( A^\circ \) the interior points of \( A \).

**Theorem 2.7.** Let \( F \) and \( G \) be Souslin spaces equipped with their Borel \( \sigma \)-algebras \( \mathcal{F} \) and \( \mathcal{G} \) respectively. Let \( X \) be an \( F \)-valued and \( Y \) be a \( G \)-valued random variable on a complete probability space \((\Omega, \Sigma, P)\). Let \( A \subset G^{\mu_Y} \) be a subset of the topological support \( S \) of \( \mu_Y \) such that either \( A \subset \overline{A^\circ} \) or \( \mu_Y(A) = 1 \). Let \( \mathcal{F}_0 \subset \mathcal{F} \) be a measure-determining class.

All solutions of the statistical inverse problem of estimating the probabilities of \( X \) given \( Y \) that are \( \mathcal{F}_0 \)-continuous on \( A \) coincide on \( F \times A \).
Proof. Assume that \( \mu_1 \) and \( \mu_2 \) are two solutions that have the described properties. If \( \mu_1 \neq \mu_2 \) on \( \mathcal{F} \times A \) then there exists \( y_0 \in A \) and \( U_0 \in \mathcal{F} \) such that \( \mu_1(U_0, y_0) \neq \mu_2(U_0, y_0) \), say \( \mu_1(U_0, y_0) > \varepsilon \). Since \( \mu_i(y, y), i = 1, 2, \) are measures, the set \( U_0 \) can be taken to be from the measure-determining class \( \mathcal{F}_0 \).

The function \( f : A \to \mathbb{R} \) defined as \( f(y) := \mu_1(U_0, y) - \mu_2(U_0, y) \) is continuous in the relative topology of \( A \) and positive at \( y_0 \). The set \( f^{-1}((\varepsilon, \infty)) \) is therefore a non-empty open set of \( y_0 \) in the relative topology of \( A \), and there exists a non-empty open set \( V \subset G \) such that \( V \cap A = f^{-1}((\varepsilon, \infty)) \). The set \( V \cap A \) has positive \( \mu_Y \)-measure. Indeed, if \( \mu_Y(A) = 1 \), then \( \mu_Y(V \cap A) = \mu_Y(V) > 0 \) by Lemma 2.6, since \( y_0 \in V \cap A \) belongs also to the support of \( \mu_Y \). On the other hand, if \( A \subset \mathcal{F}^c \), the neighborhood \( V \) of \( y_0 \) contains also points from \( A^c \). It follows that \( V \cap A \) contains a non-empty open set \( V \cap A^c \). By Lemma 2.6, \( \mu_Y(V \cap A) > 0 \).

This implies that \( \mu_1(U_0, y) - \mu_2(U_0, y) > \varepsilon \) on a set \( f^{-1}((\varepsilon, \infty)) \in \mathcal{G}^{-}\mu_V \) of positive \( \mu_Y \)-measure. Therefore, it is impossible that the both mappings \( \mu_1 \) and \( \mu_2 \) satisfy the requirements of Definition 2.1, in particular the property

\[
\int_{f^{-1}((\varepsilon, \infty))} \mu_i(U_0, y) \, d\mu_Y(y) = P(X \in U_0 \cap Y \in f^{-1}((\varepsilon, \infty))), \quad i = 1, 2,
\]

of conditional expectations does not hold. Hence, the two solutions necessarily coincide on \( \mathcal{F} \times A \).

\[\square\]

Remark 2. Recall, that a Borel measure is called strictly positive if it is positive on all non-empty open subsets. Then the topological support of the measure is the whole space. When \( \mu_Y \) is strictly positive, the partial uniqueness holds on \( \mathcal{F} \times G \) for the solutions that are \( \mathcal{F}_0 \)-continuous on \( G \). If \( \mu_Y \) is strictly positive and the solution \( \mu \) is \( \mathcal{F}_0 \) continuous on some non-empty open subset \( A \) of \( G \), we get similarly the uniqueness on \( \mathcal{F} \times A \). This situation is often encountered in finite-dimensional statistical inverse problems, where one usually excludes those \( y \in G \) for which the continuous probability density function of \( Y \) vanishes.

Remark 3. The partial uniqueness of the solution is obtained by fixing the topology of the space of observations. However, the topology of a Souslin space is a slightly ambiguous concept in measure theoretical sense. Namely, it is well-known that different topologies can generate the same Borel sets. For example, any Borel measurable function on a Souslin space is continuous with respect to some stronger topology that makes the space Souslin and generates the same Borel sets as the original topology (see Exercise 6.10.62 in [10] for the proof). If \( \mu \) is a \( \mathcal{F} \)-continuous on a Souslin space and \( \tilde{\mu} \) is its Borel-measurable version that is not continuous, then the both are continuous with respect to some stronger topology that makes also \( \tilde{\mu} \) continuous. We remark that although the essentially unique solutions are invariant under injective continuous mappings between Souslin spaces (see Remark 1), the strengthening of the topology can affect the partial uniqueness of the solution e.g. by diminishing the topological support.

Due to the properties of the conditional expectation, the prior distribution \( \mu_X(U) \) is the mixture \( \int \mu(U, y) \, d\mu_Y(y) \) of all posterior distributions so that the prior probability of \( U \) vanishes exactly when \( \mu_Y \)-almost all posterior probabilities of \( U \) vanish. When \( \mu \) is regular enough, we get the following converse result, which contrasts nicely with the well-known representation theorem considered in the next section.

Theorem 2.8. Let \( F \) and \( G \) be Souslin spaces equipped with their Borel \( \sigma \)-algebras \( \mathcal{F} \) and \( \mathcal{G} \) respectively. Let \( X \) be an \( F \)-valued and \( Y \) be a \( G \)-valued random variable on
a complete probability space \((\Omega, \Sigma, P)\). Let \(A \in \mathcal{G}^{\mu_Y}\) be any subset of the topological support \(S\) of \(\mu_Y\) such that either \(A \subset \overline{A}\) or \(\mu_Y(A) = 1\). If \(\mu\) is a solution of the statistical inverse problem of estimating probabilities of \(X\) given \(Y\) that is \(\mathcal{F}\)-continuous on \(A\), the posterior distribution \(\hat{\mu}(\cdot, y)\) at any \(y \in A\) is absolutely continuous with respect to the prior distribution.

**Proof.** Assume that \(\hat{\mu}_X(U) = 0\) for some \(U \subset \mathcal{F}\). According to the definition of conditional expectation,

\[
\int \mu(U, y) d\mu_Y(y) = P(X \in U) = \hat{\mu}_X(U),
\]

which now vanishes. Since the solution is non-negative, we get that \(\mu(U, y) = 0\) \(\mu_Y\)-almost surely on \(\overline{G}\). Since \(y \mapsto \mu(U, y)\) is continuous on \(A\), the set \(\overline{V} = \{y \in A : \mu(U, y) > 0\}\) is a relatively open set i.e. there exist an open set \(V \subset G\) such that \(V \cap A = \overline{V}\). Suppose \(\overline{V}\) is non-empty. Similarly as in the proof of Theorem 2.7, \(\mu_Y(\overline{V}) = \mu_Y(V) > 0\) when \(A\) has full measure and \(\mu_Y(\overline{V}) > 0\) when \(A \subset \overline{A}\). But this contradicts (6) because \(\hat{\mu}_X(U) = 0\). Thus the set \(\overline{V}\) is empty and \(\mu(U, y) = 0\) for all \(y\) from \(A\).

When Theorem 2.8 holds for \(A = S\), every Borel set \(B\) with full prior probability has also full posterior probability \(\mu(B, y) = 1\) for all \(y \in S\). Our posterior perception of the unknown appears to be inline with our prior insight in this aspect.

**Remark 4.** According to a result of Macci [88], the absolute continuity of the posterior distributions \(\mu(\cdot, y)\) with respect to the prior distribution for \(\mu_Y\)-almost every \(y\) implies that the conditional distribution of \(Y\) given \(X = x\) is absolutely continuous with respect to (=dominated by) \(\mu_Y\) for \(\mu_X\)-a.e. \(x \in F\). By Theorem 2.8, the posterior probabilities of Borel sets may depend continuously on the observations \(y \in G\) only in the dominated cases i.e. the conditional distribution of \(Y\) given \(X = x\) has to be absolutely continuous with respect to some \(\sigma\)-finite measure for \(\mu_X\)-a.e. \(x \in F\). The same conclusion holds even if the space \(G\) is replaced with some subset \(A \in \mathcal{G}^{\mu_Y}\) having full \(\mu_Y\)-measure. In the undominated cases, the posterior distribution has necessarily a large amount of discontinuities – the set of all discontinuity points must have positive \(\mu_Y\)-measure.

A partial converse to Theorem 2.8 holds in complete separable metric spaces.

**Theorem 2.9.** Let \(F\) and \(G\) be complete separable metric spaces equipped with their Borel \(\sigma\)-algebras \(\mathcal{F}\) and \(\mathcal{G}\), respectively. Let \(X\) be an \(F\)-valued and \(Y\) be a \(G\) random variable on a complete probability space \((\Omega, \Sigma, P)\). Let \(\mu\) be a solution of the statistical inverse problem of estimating the distribution of \(X\) given \(Y\). If the family of the posterior distributions \(\{\mu(\cdot, y) : y \in G\}\) is dominated by a Borel measure \(\nu\) on \(\mathcal{F}\), then for every \(\epsilon > 0\) there exists a compact set \(K = K(\epsilon, \mu) \in \mathcal{G}\) such that \(\mu_Y(G\setminus K) < \epsilon\) and \(\mu\) is \(\mathcal{F}_0\)-continuous on \(K\) for some family \(\mathcal{F}_0\) of measurable determining sets.

**Proof.** Equip the space \(M\) of all probability measures on \((F, \mathcal{F})\) with the topology of the weak convergence (i.e. convergence of integrals of all bounded continuous functions). It is well-known that this space is a complete separable metric space whenever \(F\) is a complete separable metric space (see Theorem 8.9.5 in [10]). Equip \(M\) with the Borel \(\sigma\)-algebra \(\mathcal{M}\) with respect to the weak topology i.e. the cylinder set \(\sigma\)-algebra of the sets of the type

\[
\{\nu \in \mathcal{M} : (\nu(f_1), \nu(f_2), \ldots) \in B\},
\]
where \( f_i, i \in \mathbb{N} \) are continuous bounded functions on \( F \) and \( B \in \mathcal{B}(\mathbb{R}^\infty) \) (with respect to the coordinate-wise convergence). By Condition 2 of Definition 2.1, the solutions \( y \mapsto \mu(y) \) are \( \mu_Y \)-measurable mappings from \( G \) to \( M \) since \( y \mapsto \mu(f, y) \) is \( \mu_Y \)-measurable as a pointwise limit of integrals of simple functions. By the Lusin theorem (see Theorem 7.1.13 in [9]), there exists a family of compact sets \( \mathcal{K} \subset \mathcal{G} \) such that given any \( \epsilon > 0 \), the probability \( \mu_Y(K^C) < \epsilon \) for some \( K \in \mathcal{K} \) and the measure-valued random variable \( y \mapsto \mu(\cdot, y) \) is continuous on \( K \) in the weak topology, of measures implying that \( \lim_{i \to \infty} \mu(f, y_i) = \mu(f, y) \) whenever \( \lim_{i \to \infty} y_i = y \) in \( K \).

Especially, the mappings \( y \mapsto \mu(\cdot, y) \) are \( \mathcal{F}_0 \)-continuous on \( K \), where \( \mathcal{F}_0 \) consists of all Borel sets \( U \) whose boundary satisfies \( \mu(\partial U, y) = 0 \) for all \( y \in K \). This follows from the fact that \( \lim_{i \to \infty} \mu(U, y_i) = \mu(U, y) \) whenever \( \lim_{i \to \infty} y_i = y \) by the weak convergence (see Corollary 8.2.10 in [10]). If the family of posterior distributions \( \{\mu(\cdot, y) : y \in K\} \) is dominated by some Borel measure, then \( \mathcal{F}_0 \) is a measure-determining set (see Lemma 1.9.4 and Proposition 8.2.8 in [10]).

3. The representation of posterior distributions. In this section, we consider a known representation formula (see Section 1.2.2 in [40], Theorem 1.31 in [111], or pp. 231–232 in [115]) for solution of the statistical inverse problem of estimating the distribution of \( X \) given the observations \( Y \) that generalizes the finite-dimensional formula

\[
D(x|y) = CD(y|x)D_{pr}(x).
\]

For readers convenience, the proofs of Lemma 3.1, Lemma 3.2 and Theorem 3.3 are included, although they are special cases of more general known results.

Throughout the section, we assume that \( F \) and \( G \) are locally convex Souslin topological vector spaces equipped with their Borel \( \sigma \)-algebras \( \mathcal{F} \) and \( \mathcal{G} \), respectively, and \( X \) is taken to be an \( F \)-valued random variable and \( \varepsilon \) is taken to be a \( G \)-valued random variable statistically independent from \( X \). All the random variables are defined on the same complete probability space \( (\Omega, \Sigma, \mathbb{P}) \). The mapping \( L : F \to G \) is assumed to be continuous. We denote \( Y = L(X) + \varepsilon \).

First, we check that \( Y \) is indeed a random variable as a combination of Borel measurable mappings. The product space \( F \times G \) is equipped with the usual product \( \sigma \)-algebra \( \mathcal{F} \otimes \mathcal{G} \) generated by rectangles \( U \times V \), where \( U \in \mathcal{F} \) and \( V \in \mathcal{G} \).

**Lemma 3.1.** The mapping \( T : (x, z) \mapsto L(x) + z \) is Borel measurable from \( F \times G \) to \( G \).

**Proof.** As the addition is just \( \mathcal{B}(G \times G) \)-measurable by continuity, there is the question whether the Borel \( \sigma \)-algebra \( \mathcal{B}(G \times G) \) of the topological product space coincides with the product \( \sigma \)-algebra \( \mathcal{G} \otimes \mathcal{G} \) generated by the rectangles \( V \times W \) where \( V, W \in \mathcal{G} \).

Certainly, \( \mathcal{G} \otimes \mathcal{G} \subset \mathcal{B}(G \times G) \), since the products of open sets \( V, W \subset G \) form a basis of topology for \( G \times G \).

Due to the Souslin property, the space \( G \times G \) is hereditarily Lindelöf ([10], Lemma 6.6.4 and Lemma 6.6.5). Any open set in \( G \times G \) can therefore be expressed as a countable union of sets of the form \( V \times W \), where \( V, W \in \mathcal{G} \) are open. Hence \( \mathcal{B}(G \times G) \subset \mathcal{G} \otimes \mathcal{G} \).

We verify now that for any \( \mu_Y \)-integrable \( f : G \to \mathbb{R} \), the conditional expectation of \( f(Y) \) given \( X \) is the random variable

\[
\mathbb{E}[f(Y)|X](\omega) = \int_G f(z) d\mu_{z + L(X(\omega))}(z).
\]
Here the measure $\mu_{\varepsilon + L(X(\omega))}$ is the image measure of the random variable $\omega' \mapsto \varepsilon(\omega') + L(X(\omega))$, where $X(\omega)$ is treated as a constant. We apply the following more general claim, for which we failed to find a reference.

**Lemma 3.2.** Let $Z_1$ be an $F$-valued and $Z_2$ be a $G$-valued random variable that are statistically independent. Denote $Z_3 = T(Z_1, Z_1)$, where $T : F \times G \to G$ is a Borel measurable mapping. For any $\mu_{Z_3}$-integrable function $f : G \to \mathbb{R}$, it holds that

$$E[f(Z_3)|Z_1](\omega) = \int_G f(z)d\mu_T(Z_1(\omega), Z_2)(z)$$

$P$-almost surely, and $\int_G f(z)d\mu_T(Z_1(\omega), Z_2)(z)$ is a version of the conditional expectation of $f(Z_3)$ given $\sigma(Z_1)$.

**Proof.** We show that the claim holds for a Borel measurable version of $f$, which exists by Proposition 2.1.11 in [10]. The generalization for $\mu_{Z_3}$-measurable functions follows then from Lemma 2.2.

Remark that $f \circ T : F \times G \to \mathbb{R}$ is then a Borel measurable function. We will show that $E[g(Z_1, Z_2)|Z_2](\omega) = \int_G g(Z_1(\omega), z_2)d\mu_{Z_t}(z_2)$ holds for all Borel measurable simple functions $g$ on $F \times G$. The usual approximation of Borel measurable functions with simple functions implies then for $g = f \circ T$ that

$$E[f(T(Z_1, Z_2))|Z_1](\omega) = \int_G f(T(Z_1(\omega), Z_2))d\mu_{Z_2}(z_2) = E[f(T(Z_1(\omega), Z_2))|Z_1](\omega)$$

$$= \int_G f(z)d\mu_T(Z_1(\omega), Z_2)(z).$$

Take now $g = 1_{C}$, where $C \in \mathcal{B}(F \times G)$. We need to determine the conditional expectation $E[1_{C}(Z_1, Z_2)|Z_2]$ i.e. the conditional distribution of $(Z_1, Z_2)$ given $\sigma(Z_2)$. Since $F$ and $G$ are Souslin spaces, a regular conditional measure exists (by Corollary 10.4.6 in [10]) and is determined by values on any measure-determining sets. In Souslin spaces, the rectangular sets $C = B_1 \times B_2$, where $B_1 \in \mathcal{F}$ and $B_2 \in \mathcal{G}$ are measure-determining sets, since $\mathcal{B}(F \times G) = \mathcal{F} \otimes \mathcal{G}$ (see the proof of Lemma 3.1 and Lemma 1.9.4 in [10])). By the properties of the conditional expectation,

$$E[1_{B_1 \times B_2}(Z_1, Z_2)|Z_1](\omega) = 1_{B_1}(Z_1(\omega)) \int 1_{B_2}(z_2)d\mu_{Z_2}(z_2)$$

$$= \int 1_{B_1 \times B_2}(Z_1(\omega), z_2)d\mu_{Z_2}(z_2).$$

Here is the description of the solutions $\mu(U, z)$ modulo $\mu_Y$-zero measurable sets. The result is a special case of Kallianpur-Striebel formula [68].

**Theorem 3.3.** Let $\mu_{\varepsilon + L(x)}$ be absolutely continuous with respect to a $\sigma$-finite measure $\nu$ for $\mu_X$-a.e. $x \in F$. Set

$$\rho(x, z) = \begin{cases} \frac{d\mu_{\varepsilon + L(x)}}{d\nu}(z) & \text{when } \mu_{\varepsilon + L(x)} \ll \nu \\ 0 & \text{otherwise.} \end{cases}$$

If $\rho(x, z)$ is a non-negative $\mu_X \times \nu$-measurable function on $F \times G$, then there is an essentially unique solution $\mu$ of the statistical inverse problem of estimating the distribution of $X$ given $Y = L(X) + \varepsilon$ such that

$$\mu(U, z) = \frac{\int 1_{U}(x)\rho(x, z)d\mu_{X}(x)}{\int \rho(x, z)d\mu_{X}(x)}$$

\[7\]
for all \( z \in G \setminus N_0 \), where the set
\[
N_0 = \{ z \in G : \int \rho(x,z) d\mu_X(x) = 0 \text{ or } \infty \}
\]
has \( \mu_Y \)-measure zero.

If \( \mu_Y(N) = 0 \) then \( N \) is also \( \mu_{\tau} - \text{zero measurable} \) for \( \mu_X \)-almost every \( x_0 \in F \). If additionally \( \mu_{\tau} - \text{zero measurable} \) for \( \mu_X \)-almost everywhere, then \( \mu_{\tau} - \text{zero measurable} \) for \( \mu_X \)-almost every \( x_0 \in F \).

**Proof.** Let \( \mu(U,z) \) be defined by (7). If \( z \in N_0 \), we set \( \mu(U,z) = 1_{U}(x_0) \) for some fixed \( x_0 \in F \). We prove that \( \mu \) is a solution.

Let \( U \in \mathcal{F} \) and \( V \in \mathcal{G} \). By Theorem 2.4 there exists an essentially unique solution, which we denote here with \( \tilde{\mu} \). We write two expressions for \( P(X \in U \cap Y \in V) \) using Lemma 3.2. The first is
\[
E[1_U(X)E[1_V(Y)|X]] = \int 1_U(x) \left( \int 1_V(z) d\mu_{\tau}(z) \right) d\mu_X(x)
\]
\[
= \int 1_U(x) \left( \int 1_V(z) \rho(x,z) d\nu(z) \right) d\mu_X(x)
\]
\[
= \int 1_V(z) \left( \int 1_U(x) \rho(x,z) d\mu_X(x) \right) d\nu(z)
\]
and the second expression is
\[
E[1_V(Y)E[1_U(X)|Y]] = E[1_V(Y)\tilde{\mu}(U,Y)]
\]
\[
= E[E[1_V(Y)\tilde{\mu}(U,Y)|X]]
\]
\[
= \int \int 1_V(z) \rho(x,z) \tilde{\mu}(U,z) d\nu(z) d\mu_X(x)
\]
\[
= \int 1_V(z) \left( \int \rho(x,z) d\mu_X(x) \right) \tilde{\mu}(U,z) d\nu(z).
\]
The measurability of \( \rho \) is used in changing the order of integrations by the Fubini theorem. The integrability of \( \rho \) follows automatically from the finiteness of the left-hand side of (9) for \( U = F \) and \( V = G \).

Since the equivalence of (8) and (9) holds for all \( V \in \mathcal{G} \), we obtain
\[
\tilde{\mu}(U,z) \int \rho(x,z) d\mu_X(x) = \int 1_U(x) \rho(x,z) d\mu_X(x)
\]
for \( \nu \)-almost every \( z \). Hence, \( \tilde{\mu}(U,z) = \mu(U,z) \) for \( \nu \)-almost every \( y \) such that \( 0 < \int \rho(x,z) d\mu_X(x) < \infty \).

The denominator in (7) may vanish only on a set \( A \) of \( \mu_Y \)-measure zero since the choice \( U = F \), \( V = A \) gives \( \mu_Y(A) = 0 \) in (8). The same consideration implies that also the measure \( \mu_Y \) is absolutely continuous with respect to \( \nu \). Similarly, the denominator is finite \( \nu \)-almost surely, which implies \( \mu_Y \)-almost surely. We conclude that \( N_0 \) has \( \mu_Y \)-measure zero and \( \tilde{\mu}(U,z) = \mu(U,z) \) \( \mu_Y \)-almost surely. Then \( \mu(U,y) \) satisfies Condition 2 of Definition 2.1. By Lemma 2.2, \( \mu \) satisfies Condition 1. By the integrability of \( \rho \), \( \mu \) satisfies Condition 3 of Definition 2.1.

We proceed to the last claim. Taking \( V = N \) and \( U \in \mathcal{F} \) in (8) implies that \( \mu_{\tau} - \text{zero measurable} \) for \( \mu_X \)-almost all \( x \in F \). When \( \rho \) is a.e. positive, also \( \nu(N) \) has to vanish. We obtain \( \mu_{\tau}(x_0) = 0 \) for all \( x_0 \in F \) by using the absolute continuity.
The last statement of the above theorem is added to show how small the zero measurable set for a given unknown is. The representation formula does not improve the essential uniqueness of solutions, because the Radon-Nikodym density $z \mapsto d\mu_{x+L(x)} + \varepsilon/d\nu(z)$ is only determined up to $\nu$-equivalence. It should be noted that under the domination assumptions on $\mu_{x+L(x)}$ in Theorem 3.3, there always exists versions of the Radon-Nikodym densities that are jointly measurable. In [68], this claim is proved assuming that $Y^{-1}(G)$ is countably generated. In Souslin spaces, the Borel $\sigma$-algebras are countably generated by Corollary 6.7.5 in [10].

It is easy to see, that the prior distribution $\mu_X$ and the posterior distribution $\mu(\cdot, z)$ are equivalent if (7) holds and $\rho(\cdot, z) > 0 \mu_X$-almost everywhere.

**Remark 5.** The existence of $\nu$ is a delicate matter. For example, the measure $\mu_{x+L(x)}$ may not be almost surely absolutely continuous with respect to $\mu_Y$, although

$$
\mu_Y(U) = E[1_U(Y)] = E[E[1_U(Y)|X]] = \int \mu_{x+L(x)}(U)d\mu_X(x)
$$

by Lemma 3.2. We can only conclude that $\mu_{x+L(x)}(U)$ vanishes and $\mu_X$-a.s. whenever $\mu_Y(U)$ vanishes and the $\mu_X$-zero measurable set may depend on $U$. A Gaussian example in Remark 9 of Section 4 shows that this is indeed the case. In general, the Halmos-Savage theorem (Lemma 7 in [51]), states that from a dominated family of finite measures, which in our case is $\{\mu_{x+L(x)} : x \in M\}$ where $\mu_X(M) = 1$, one can pick out countably many measures $\mu_{x+L(x)}$ such that the measure $\nu := \sum a_i\mu_{x+L(x)}$, where $\sum a_i = 1$ and all $a_i > 0$, is not only a dominating measure but also equivalent to the family $\{\mu_{x+L(x)} : x \in M\}$ (i.e. the measures in the family vanish on the same subsets as $\nu$). Especially, this gives a necessary and sufficient condition for the domination of the probability measures $\mu_{x+L(x)}$. In Section 4 we concentrate on special cases where $\mu_x$ can be taken as $\nu$. In these examples, we require that $\mu_{x+L(x)}(U) = 0$ whenever $\mu_{x}(U) = 0$ i.e. $\mu_x$ is quasi-invariant with respect to translations with $L(x)$, where $x \in F$. This allows the use of any prior distribution on $F$. However, Remark 6 in Section 4 demonstrates that in dominated cases it is not always possible to choose $\mu_x$ as $\nu$.

We return to the question of partial uniqueness (Theorem 2.7). The conditions in the next theorems allow easier validation of the measurability and guarantee some continuity for the solutions. However, under the stronger assumption that the function $(x, z) \mapsto \rho(x, z)$ is jointly continuous and bounded, the solution $\mu$ is always $F$-continuous on $G$ (see Theorem 7.14.8 in [10]). Recall, that the class of all Souslin sets is quite large since all Borel subsets of a Souslin space are Souslin sets by Corollary 6.7.7 in [10].

**Theorem 3.4.** Let $\mu_{x+L(x)}$ be absolutely continuous with respect to a $\sigma$-finite measure $\nu$ for $\mu_X$-almost every $x \in F$. If

$$
\rho(x, z) = \begin{cases} 
d\mu_{x+L(x)}(z) & \text{when } \mu_{x+L(x)} \ll \nu \\
0 & \text{otherwise.}
\end{cases}
$$

is a separately continuous function on some $F_0 \times A$, where $F_0$ is a Souslin subset of $F$ with full $\mu_X$-measure and $A$ is a Souslin subset of $G$ such that $\nu(A^C) = 0$, then $\rho$ is $\mu_X \times \nu$-measurable.
If additionally \( \sup_{x \in K} \rho(x, z) \in L^1(\mu_X) \) for all compact sets \( K \subset G \) then

\[
(10) \quad z \mapsto \mu(U, z) = \frac{\int 1_U(x) \rho(x, z) d\mu_X(x)}{\int \rho(x, z) d\mu_X(x)}
\]
is \( F \)-continuous on \( K \cap \{ z \in A : 0 < \int \rho(x, z) d\mu_X(x) < \infty \} \) for every compact set \( K \subset G \).

Proof. Assume that \( \rho \) is separately continuous. Since \( F_0 \) is a Souslin space, there exists a continuous surjection \( R \) from some complete separable metric space \( M \) onto \( F_0 \). We consider first the function \((m, z) \mapsto \rho(R(m), z)\) on \( M \times A \). This function is a pointwise limit of continuous functions due to a theorem of W. Rudin [107]. Hence the function \((m, z) \mapsto \rho(R(m), z)\) is \( \mathcal{B}(M \times A) = \mathcal{B}(M) \otimes \mathcal{B}(A) \)-measurable. We compose it with a \( \mu_X \times \nu \)-measurable mapping \((R^{-1}, I)\) where the inverse comes from the measurable choice theorem (see Theorem 6.9.1 in [10], note that Souslin sets are universally measurable i.e. measurable with respect to any finite Radon measure by Theorem 7.4.1 in [10]). Then we see that \( 1_A(z)1_{F_0}(x)\rho(x, z) \), together with its equivalent mapping \( \rho(x, z) \), is \( \mu_X \times \nu \)-measurable.

By the Lebesgue dominated convergence theorem, we obtain the sequential continuity of the marginals. On Souslin spaces, the compact sets are metrizable (see Corollary 6.7.8 in [10]). In metrizable spaces sequential continuity coincides with continuity.

The above Theorem 3.4 shows \( \mathcal{F} \)-continuity of the solution \( \mu \) on \( \{ z \in A : 0 < \int \rho(x, z) d\mu_X(x) < \infty \} \) when \( G \) is e.g. a \( k \)-space i.e. a subset \( C \) of \( G \) is closed if and only if \( C \cap K \) is closed for every compact \( K \subset G \) (see Definition 43.8 in [140]). Indeed, it is well-known that a function \( f \) is continuous on a \( k \)-space if and only if it is continuous on every compact subset. In particular, this holds for all first-countable spaces, like metric spaces. Note, that the space of tempered distributions \( S'(\mathbb{R}^n) \) is a \( k \)-space when equipped with its strong topology but not with its weak topology, while the distribution space \( \mathcal{D}'(U) \) is not a \( k \)-space with respect to either topology [52]. But \( \mathcal{D}'(U) \) is a Lusin space [112] – i.e. a Hausdorff space that is a continuous injective image of a complete metric space – and can be equipped with a stronger metrizable topology inherited from the metric space. However, this topology depends on the chosen metric space and has all the drawbacks indicated in Remark 3.

We combine Theorem 2.7 and Theorem 3.4 in a simple case.

**Corollary 1.** Let \( G \) be a \( k \)-space. Let \( \mu_{x \in L(x)} \) be equivalent with a probability measure \( \nu \) for every \( x \in F \). Denote \( S_\nu \) the topological support of \( \nu \). If

\[
\rho(x, z) = \frac{d\mu_{x \in L(x)}}{d\nu}(z)
\]
is a separately continuous function on \( F \times S_\nu \) and if \( \sup_{x \in K} \rho(x, z) \in L^1(\mu_X) \) for all compact subsets \( K \subset G \) then all solutions of the statistical inverse problem of estimating the distribution of \( X \) given \( Y \) that are \( \mathcal{F} \)-continuous on \( \{ z \in S_\nu : 0 < \int \rho(x, z) d\mu_X(x) < \infty \} \) coincide with

\[
(11) \quad z \mapsto \mu(U, z) = \frac{\int 1_U(x) \rho(x, z) d\mu_X(x)}{\int \rho(x, z) d\mu_X(x)}
\]
on \( \mathcal{F} \times \{ z \in S_\nu : 0 < \int \rho(x, z) d\mu_X(x) < \infty \} \).
The proof is an immediate consequence of the following lemma, where we characterize the topological support of $\mu_Y$ in more convenient terms.

**Lemma 3.5.** Let $Y = L(X) + \varepsilon$, where $X$ and $\varepsilon$ are statistically independent. The topological support of $\mu_Y$ is the smallest closed set $S \subseteq G$ such that $\mu_{x+L(x)}(S) = 1$ for $\mu_X$-almost every $x \in F$. Moreover, if $\mu_{x+L(x)}$ is equivalent with a probability measure $\nu$ for every $x \in F$, then the topological supports of $\mu_Y$ and $\nu$ coincide.

**Proof.** The first claim follows from the convolution $\mu_Y = \int_G \mu_{x+L(x)}(S)d\mu_X(x)$. For the second claim, we note that $\mu_{x+L(x)}(S) = \nu(S)$ for every closed set $S \subseteq G$ with full $\nu$-measure (or full $\mu_{x+L(x)}$-measure).

4. **Examples of noise.** Below, some cases are presented, where the Radon-Nikodym derivatives

$$\frac{d\mu_{x+L(x)}}{d\nu}$$

exist with respect to some $\sigma$-finite measure $\nu$. Two first cases, where the noise term is finite-dimensional or Gaussian, are well-known. For these cases, we apply the results of previous sections. The next four cases demonstrate that the approach taken in this paper applies also for more general noise models.

4.1. **Finite-dimensional noise with a probability density.** Let $G$ be the Euclidian space $\mathbb{R}^k$, let $F$ be a locally convex Souslin space, and let $L : F \rightarrow G$ be a continuous mapping. Consider the statistical inverse problem of estimating the distribution of an $F$-valued random variable $X$ given a sample $y$ of a $G$-valued random variable $Y = L(X) + \varepsilon$, where the $G$-valued random variable $\varepsilon$ is statistically independent from $X$. In order to use the representation formula of Theorem 3.3 for the essentially unique posterior distribution of $X$ given a sample $y_0$ of $Y$, we need the required $\sigma$-finite measure $\nu$. A natural choice is to take the Lebesgue measure as $\nu$, when possible.

Assume that the noise $\varepsilon$ is $\mathbb{R}^k$-valued random vector whose image measure $\mu_\varepsilon$ is absolutely continuous with respect to the Lebesgue measure, say $\mu_\varepsilon(dx) = D_\varepsilon(x)dx$, with the property that $D_\varepsilon > 0$ almost everywhere. Especially, $\mu_\varepsilon$ is then equivalent to the Lebesgue measure.

In Theorem 3.3, the Radon-Nikodym derivative of $\mu_{x+L(x)}$ with respect to the Lebesgue measure, i.e. $(x, y) \mapsto D_\varepsilon(y - L(x))$, is required to be jointly measurable. Since $D_\varepsilon(x)$ is measurable, and the addition is measurable, the continuity of $L$ suffices here. We obtain an essentially unique solution $\mu$ of the statistical inverse problem of estimating the distribution of $X$ given a sample $y_0$ of $Y$ that satisfies

$$\mu(U, y_0) = \frac{\int_U D_\varepsilon(y_0 - L(x))d\mu_X(x)}{\int D_\varepsilon(y_0 - L(x))d\mu_X(x)}$$

for all $U \in \mathcal{F}$ and all $y_0$ such that $0 < \int D_\varepsilon(y_0 - L(x))d\mu_X(x) < \infty$. Here $D_\varepsilon(y_0 - L(x))$ is often called the likelihood function. If $D_\varepsilon$ is continuous and bounded, we may drop out the word “essentially”, as the solution is the unique continuous solution by Corollary 1 (the topological support of $\mu_Y$ is the whole space by Lemma 3.5 since $\mu_{x+L(x)}$ is equivalent with the Lebesgue measure).

When $X$ is an $\mathbb{R}^m$-valued random variable with a density $D_{pr}(x)$ with respect to the Lebesgue measure, we get the familiar expression

$$\mu(U, y) = \frac{\int_U D_\varepsilon(y - L(x))D_{pr}(x)dx}{\int D_\varepsilon(y - L(x))D_{pr}(x)dx}$$
for all $y$ such that $0 < \int D_\varepsilon(y - L(x)) D_{pr}(x) dx < \infty$.

**Remark 6.** When $D_\varepsilon \geq 0$ almost everywhere, $\mu_\varepsilon$ need not be equivalent to the Lebesgue measure. Moreover, the translated measure $\mu_{\varepsilon + L(x)}$ need not be absolutely continuous with respect to $\mu_\varepsilon$.

### 4.2. Infinite-dimensional Gaussian noise.

The finite-dimensional Gaussian noise model is often chosen because of its relatively straightforward justification – if the total noise is produced by many identical independent noise sources (having finite mean and variance), the sum is nearly Gaussian by the central limit theorem. For instance, this applies to the origin of thermal noise in electrical circuits, where heat motion of the charge carriers disturbs the analog signal. The usual model of thermal noise is white Gaussian noise, which is an acceptable approximation on usual frequencies.

We first recall a method for constructing infinite-dimensional Gaussian random vectors by a procedure linked to abstract Wiener spaces [9].

#### 4.2.1. Basics of Hilbert space-valued Gaussian random variables.

Let $H$ be a separable Hilbert space. We define $Z$ as a random sum

$$Z = \sum_{i=1}^{\infty} Z_i e_i,$$

where $Z_i$ are independent standard normal random variables on $(\Omega, \Sigma, P)$ and $\{e_i\}$ is an orthonormal basis of $H$. Clearly, the sum does not converge a.s. in $H$. Instead, we take a larger Hilbert space $G$ into which $H$ can be imbedded with an injective Hilbert-Schmidt operator $j$. When the range of the imbedding is dense, the triple $(j, H, G)$ is a special case of an abstract Wiener space [9]. However, we do not require the range to be dense. Let $G'$ denote the dual space of $G$ and $\langle \cdot, \cdot \rangle$ the duality between $G$ and $G'$.

A sufficient condition for the a.s. convergence of the random sums $\sum_{i=1}^{\infty} Z_i e_i$ in $G$ is that the series

$$\sum_{i=1}^{\infty} E[\|Z_i e_i\|_G^2] = \sum_{i=1}^{\infty} \|e_i\|_G^2.$$

is convergent [63]. But this follows from the Hilbert-Schmidt property of the inclusion map $j$. Since $G$ is a separable Fréchet space (more generally, a locally convex Souslin space [112]), its Borel $\sigma$-algebras with respect to the weak and the original topology coincide. The benefit of the weak topology is that the measurability of the limit $Z = \lim_{n \to \infty} \sum_{i=1}^{n} Z_i e_i$ can be checked similarly as in the case of real-valued functions with sets of the type $\bigcap_{i=1}^{k} \{|(Z, \phi_i) - a_i| < c_i\}$. We conclude that the a.s. limit $Z$ of the random sums defines a measurable mapping from $(\Omega, \Sigma, P)$ to $(G, G')$. Its image measure $\mu_Z = P \circ Z^{-1}$ can be viewed also as a countably additive cylinder set measure.

In general, the mean of a random variable $Z$ is the vector $m \in G'' = G$ such that $E[\langle Z, \phi \rangle] = \langle m, \phi \rangle$ for all $\phi \in G'$ and the covariance operator of $Z$ is the mapping $C : G' \to G$ such that

$$\langle C\phi, \psi \rangle = E[(\langle Z, \phi \rangle - \langle m, \phi \rangle)(\langle Z, \psi \rangle - \langle m, \psi \rangle)]$$

for all $\phi, \psi \in G'$ [9].
Since limits of Gaussian random variables are Gaussian, the random variable
\[ \langle Z + m, \phi \rangle = \langle m, \phi \rangle + \sum_{i=1}^{\infty} Z_i \langle e_i, \phi \rangle, \]
where \( m \in G \), has a characteristic function
\[ e^{i \langle m, \phi \rangle - \frac{1}{2} \langle C \phi, \phi \rangle} = E[e^{i \langle Z, \phi \rangle}] \tag{13} \]
for all \( \phi \in G' \). In our case, the random variable \( Z \) has mean \( m = 0 \) and covariance \( \langle C \phi, \phi \rangle = \sum_{i=1}^{\infty} \langle e_i, \phi \rangle^2 \). The covariance \( \langle C \phi, \phi \rangle \) is the squared norm of \( \phi \) in the strong dual space \( H' \) of \( H \). Indeed, the linear form \( \langle j, \phi \rangle_{G, G'} \) is continuous on \( H \) so it belongs to \( H' \) and its norm is \( \langle C \phi, \phi \rangle \). For short, we denote \( \phi \in H' \). The covariance \( \langle C \phi, \phi \rangle \) for any \( \phi \in G' \) is finite, since \( H \rightarrow G \) implies that \( G' \rightarrow H' \) continuously. The dual space \( G' \) is actually dense in \( H' \) as a consequence of the Hahn-Banach theorem. Indeed, if \( h_0' \in H' \setminus j'(G') \neq \emptyset \) then there would exist \( h \in H'' = H \) such that \( \langle h, h_0' \rangle = 1 \) and \( \langle h, h' \rangle = 0 \) for every \( h' \in j'(G') \). But \( j'(G') \) separates the points in \( H \) because of the injectivity of \( j \). Therefore, \( h = 0 \) and hence \( j'(G') \) is dense in \( H' \).

The mapping \( G' \ni \phi \mapsto \langle C \phi, \phi \rangle \) has an extension \( H' \ni g \mapsto \langle Cg, g \rangle := \|g\|_{H'}^2 \). By the polarization equality, \( C \) is the isometric isomorphism between \( H' \) and \( H \) defined by the Riesz representation theorem. We continue to denote \( C \) with \( C \).

**Remark 7.** It is well-known that the sample space \( G \) of \( Z \) can be replaced with any bigger locally convex Souslin vector space \( G_0 \) into which \( G \) can be continuously and injectively embedded. For example, \( G_0 \) may be the distribution space \( D'(U) \), where \( U \subset \mathbb{R}^n \) is open, equipped with the usual weak topology.

Measures having characteristic functions of the above form (13) are called Gaussian measures. Especially, the image measure \( \mu_Z = P \circ Z^{-1} \) is Gaussian. Random variables, whose image measures are Gaussian, are called Gaussian random variables. The space \( H \) is the so-called *Cameron-Martin space of \( \mu_Z \).

By Theorems 3.2.3, 3.2.7 and 3.5.1 in [9] any zero-mean Gaussian random variable on a locally convex Souslin space is equivalent with a random variable of the form (12). More details on Gaussian measures can be found in [9, 48, 78].

### 4.2.2. Inverse problems with Gaussian noise

We consider the statistical inverse problem of estimating the distribution of \( X \) given a sample of \( Y = L(X) + \varepsilon \), where \( \varepsilon \) is a zero mean Gaussian random variable that has values in a separable Hilbert space \( G \).

We denote with \( H_{\mu_{\varepsilon}} \) the Cameron-Martin space of \( \mu_{\varepsilon} \) and with \( C_{\varepsilon} : H_{\mu_{\varepsilon}} \rightarrow H_{\mu_{\varepsilon}} \) the covariance operator of \( \varepsilon \). The unknown random variable \( X \) has values in some locally convex Souslin topological vector space \( F \). The random variables \( \varepsilon \) and \( X \) are taken to be independent. The direct theory \( L : F \rightarrow G \) is a continuous mapping that satisfies the folloing additional restrictive conditions: \( L : F \rightarrow H_{\mu_{\varepsilon}} \) is continuous, the range of the combined mapping \( C_{\varepsilon}^{-1}L \) belongs to \( G' \) where \( G' \) is the strong dual of \( G \), and the mapping \( C_{\varepsilon}^{-1}L : F \rightarrow G' \) is continuous. Furthermore, let \( \exp(a\|C_{\varepsilon}^{-1}L(X)\|_{G'}) \in L^1(P) \) for all \( a > 0 \).

According to the famous Cameron-Martin formula (see Corollary 2.4.3 and Theorem 3.2.3 in [9]), the Gaussian measures \( \mu_{\varepsilon} \) and \( \mu_{\varepsilon + L(x)} \) are equivalent when \( L(x) \in H_{\mu_{\varepsilon}} \). The corresponding Radon-Nikodym density is
\[ \rho(x, z) := \frac{d\mu_{\varepsilon + L(x)}}{d\mu_{\varepsilon}}(z) = \exp \left( \langle z, C_{\varepsilon}^{-1}L(x) \rangle - \frac{1}{2} \|L(x)\|_{H_{\mu_{\varepsilon}}}^2 \right), \quad z \in G. \]
Theorem 3.3, we may choose $G$ to the space $G_L$ should hold also for some $x$ by the formula for a Gaussian random variable on $(G, \langle \cdot \rangle_{G,G'})$. Namely, the vector $C_x^{-1}L(x)$ is not so simple in the situation described in Remark 7. 1. In the light of Corollary 1 and the discussion preceding it, the partial uniqueness of the solution $\mu$ is the only $\mu_Y$-measurable. In Theorem 3.3, we may choose $\mu = \mu_x$ and take

$$\mu(U,y) := \frac{\int_U \exp \left( \langle y, C_x^{-1}L(x) \rangle - \frac{1}{2} \|L(x)\|^2_{H_x} \right) d\mu_X(x)}{\int \exp \left( \langle y, C_x^{-1}L(x) \rangle - \frac{1}{2} \|L(x)\|^2_{H_x} \right) d\mu_X(x)}$$

as an essentially unique solution for all $y \in G$. Note, that our assumptions guarantee that

$$0 < \exp \left( \langle y, C_x^{-1}L(x) \rangle - \frac{1}{2} \|L(x)\|^2_{H_x} \right) \leq \exp \left( \|y\|_{G'} \|C_x^{-1}L(x)\|_{G'} \right) \in L^1(\mu_X).$$

We consider next the partial uniqueness of the solution $\mu$ on $\mathcal{F} \otimes G$. Denote with $S_{\mu_x}$ the support of $\mu_x$ on $G$, which coincides with the closure of the Cameron-Martin space $H_{\mu_x}$ in $G$ by Theorem 3.6.1 in [9]. The measure $\mu_{x+L(x)}$ is equivalent with $\mu_x$ by the Cameron-Martin formula. Hence, the measure $\mu_Y$ has the same topological support as the measure $\mu_x$ by Lemma 3.5. We conclude that $S_{\mu_Y} = \overline{H_{\mu_x}}$. Since $\sup_{z \in K} \rho(x,z) \leq \exp(C \|C_x^{-1}L(x)\|_{G'})$, the solution $\mu$ is $\mathcal{F}$-continuous on $G \cap \overline{H_{\mu_x}}$ by Theorem 3.4. Hence, $\mu$ is the only $\mathcal{F}$-continuous solution on $G \cap \overline{H_{\mu_x}}$ by Corollary 1. In the light of Corollary 1 and the discussion preceding it, the partial uniqueness is not so simple in the situation described in Remark 7.

Remark 9. In general, the measure $\mu_Y = \mu_{L(X)+\epsilon}$ does not satisfy $\mu_{x+L(x)} \ll \mu_Y$ for $\mu_X$-almost every $x$. Indeed, take $X$ and $\epsilon$ to be independent Gaussian random variables with the same Cameron-Martin space $L^2(I)$, where $I$ is the unit interval $(0,1)$. Let $L$ be the identity. If $\mu_Y(U) = 0$, then $\mu_{x+\epsilon}(U) = 0$ for $\mu_X$-almost every $x$ by the formula

$$\mu_Y(U) = E[|1_{Y}(U)|X] = \int \mu_{x+\epsilon}(U) d\mu_X(x).$$

Suppose that $\mu_{x+\epsilon} \ll \mu_Y$ for $\mu_X$-a.e. $x$, say, for all $x \in M$ such that $\mu_X(M) = 1$. The random variable $Y = X + \epsilon$ is also Gaussian, and any two Gaussian measures on the same locally convex space are either equivalent or singular. But then $\mu_{x+\epsilon}$ is equivalent to $\mu_Y$ and $\mu_{x_1+\epsilon}$ is equivalent to $\mu_Y$ for any $x_1, x_2 \in M$ so also the two measures $\mu_{x_1+\epsilon}$ and $\mu_{x_2+\epsilon}$ are equivalent. But $P(X \in L^2(I)) = 0$ so equivalence should hold also for some $x_1, x_2 \notin L^2(I)$, which is impossible by the Cameron-Martin theorem. The $\mu_X$-zero measurable set in (14) necessarily depends on $U$ in this case.

4.3. Gaussian dominated noise. We consider a simple modification of Gaussian noise. Suppose that the assumptions in Section 4.2.2 hold except that instead of $Y = L(X) + \epsilon$ we are observing $\tilde{Y} = L(X) + \tilde{\epsilon}$, where $\tilde{\epsilon}$ is dominated by the
Gaussian measure $\mu_\varepsilon$ i.e.

$$\frac{d\mu_\varepsilon}{d\mu}(y) = f(y)$$

for some $f \in L^1(\mu)$. The translation of $\mu_\varepsilon$ by $L(x)$ has the form

$$\mu_{\varepsilon + L(x)}(V) = \int_V 1_Y(y + L(x))d\mu_\varepsilon(y)$$

$$= \int_V 1_Y(y + L(x))f(y)d\mu_\varepsilon(y)$$

$$= \int_V f(y - L(x))d\mu_{\varepsilon + L(x)}(y)$$

$$= \int_V f(y - L(x))\exp \left(\langle y, C_\varepsilon^{-1}L(x) \rangle - \frac{1}{2}\|L(x)\|_{H_{\mu_\varepsilon}}^2\right) d\mu_\varepsilon(y).$$

The integrand is a $\mu_X \times \mu_\varepsilon$-measurable functions as a product of two $\mu_X \times \mu_\varepsilon$-measurable functions. By Theorem 3.3, the posterior distribution of $X$ given a sample $y$ of $\tilde{Y} = L(X) + \tilde{\varepsilon}$ can be taken to be

$$\mu(U, y) = \frac{\int_U f(y - L(x))\exp \left(\langle y, C_\varepsilon^{-1}L(x) \rangle - \frac{1}{2}\|L(x)\|_{H_{\mu_\varepsilon}}^2\right) d\mu_X(x)}{\int f(y - L(x))\exp \left(\langle y, C_\varepsilon^{-1}L(x) \rangle - \frac{1}{2}\|L(x)\|_{H_{\mu_\varepsilon}}^2\right) d\mu_X(x)}$$

whenever the denominator is positive and finite.

When $f > 0$ is a bounded continuous function, then the topological supports of $\mu_{\varepsilon + L(x)}$ and $\mu_\varepsilon$ coincide for all $x \in F$, and by Lemma 3.5, the topological supports of $\mu_{\tilde{\varepsilon}}$ and $\mu_\varepsilon$ coincide. Hence, the topological support of $\mu_{\tilde{\varepsilon}}$ is the closure of $H_{\mu_\varepsilon}$ in $G$. The solution $\mu$ is the only $F$-continuous solution on $G \cap H_{\mu_\varepsilon}$ by Corollary 1.

As a second example, let $\varepsilon$ to be a restriction of $\varepsilon$ to some open set $K \in \mathcal{G}$ that has positive $\mu_\varepsilon$-measure. This means that the noise $\tilde{\varepsilon} = \varepsilon|_K$ has the distribution

$$\mu_{\varepsilon}(V) = \frac{P(\varepsilon \in K \cap V)}{P(\varepsilon \in K)}$$

for all $V \in \mathcal{G}$ i.e. we consider conditional probabilities

$$\mu_{\varepsilon}(V) = \mu_{\varepsilon}(V|K).$$

Note that as a Borel set, $K$ is of the form $K = \{y \in G : (y, \phi_1), (y, \phi_2), \cdots \in E\}$, where $\phi_i \in G'$ separate the points in $G$ and $E \in \mathcal{B}(\mathbb{R}^\infty)$. The Radon-Nikodim density of $\mu_{\varepsilon}$ with respect to $\mu_\varepsilon$ is by (16)

$$f(y) = \frac{d\mu_{\varepsilon}}{d\mu_\varepsilon}(y) = \frac{1}{\mu_\varepsilon(K)}1_K(y).$$

By Theorem 3.3, an essentially unique posterior distribution of $X$ given a sample $y_0$ of $\tilde{Y} = L(X) + \tilde{\varepsilon}$ can be represented as

$$\mu(U, y_0) = \frac{\int_U 1_K(y_0 - L(x))\exp \left(\langle y_0, C_\varepsilon^{-1}L(x) \rangle - \frac{1}{2}\|L(x)\|_{H_{\mu_\varepsilon}}^2\right) d\mu_X(x)}{\int 1_K(y_0 - L(x))\exp \left(\langle y_0, C_\varepsilon^{-1}L(x) \rangle - \frac{1}{2}\|L(x)\|_{H_{\mu_\varepsilon}}^2\right) d\mu_X(x)}$$

whenever the denominator is positive. We see that when we can exclude noise patterns, the posterior distribution will concentrate more on the true value $x_0$ (when $L$ is injective). The partial uniqueness with respect to the topology of $G$ remains an open question.
Another example arises from the Girsanov formula. We equip \( G = C([0, T]) \), where \( T > 0 \), with the usual supremum norm. The space \( G \) is then complete separable Banach space and its dual space \( G' \) is the space of Radon measures on \([0, T]\). We assume that the observation is of the form \( Y_t = L(X)_t + \tilde{\varepsilon}_t \) for \( 0 \leq t \leq T \), where \( F \)-valued \( X \) and \( C([0, T]) \)-valued \( \tilde{\varepsilon} \) are statistically independent and \( L : F \to C([0, T]) \) is a continuous mapping. More precisely, we assume a stronger condition that \( L : F \to C^2_{\tilde{G}}(0, T) \) is continuous.

Suppose that the noise \( \tilde{\varepsilon} \in G \) is of the form

\[
\tilde{\varepsilon}_t = \varepsilon_t + \int_0^t a(s; \varepsilon_s)ds
\]

where \( \varepsilon_t \) is an ordinary Brownian motion on \([0, T] \) and \( a : [0, T] \times \mathbb{R} \to \mathbb{R} \) is continuous. Note that \( \varepsilon_t \) indeed is a \( C([0, T]) \)-valued random variable since the continuous functionals \( \{\delta_t : t \in Q \cap [0, T]\} \) separate the points in \( G \) and, therefore, also generate the \( \sigma \)-algebra of \( G \).

It is well-known that the Cameron-Martin space of the Brownian motion on \([0, T] \) is the separable Hilbert space \( \{ f \in H^1(0, T) : f(0) = 0 \} \) equipped with the norm \( \|f\|_{L^2} \), the covariance operator \( C_\varepsilon \) has kernel \( \min(t, s) \) and \( C_\varepsilon^{-1} = \frac{d^2}{ds^2} \) on \( \{ f \in H^2(0, T) : f(0) = 0, f'(T) = 0 \} \) (see [9]). By the Cameron-Martin theorem

\[
\frac{d\mu_{\varepsilon + L(x)}}{d\mu_{\varepsilon}}(y) = \exp \left( \int_0^T y_s \frac{d^2L(x)}{ds^2} ds - \frac{1}{2} \left\| \frac{dL(x)}{ds} \right\|_{L^2(0, T)}^2 \right).
\]

The Girsanov formula

\[
\frac{d\mu_{\varepsilon}}{d\mu}(y) = \exp \left( \int_0^T a(s, \varepsilon_s)dy_s - \frac{1}{2} \int_0^T |a(s, \varepsilon_s)|^2 ds \right),
\]

where the first integral is a sample of the corresponding stochastic integral, holds when the Novikov’s condition

\[
\mathbb{E} \left[ \exp \left( \int_0^T |a(s; \varepsilon_s)|^2 ds \right) \right] < \infty
\]

is satisfied (see [94]). For example, if \( |a(s, x)| \leq C(1 + |x|) \) for some \( C > 0 \), then (17) holds since

\[
\mathbb{E}[e^{C \int_0^T \varepsilon^2 ds}] \leq e^{C^2} \mathbb{E}[e^{\|\varepsilon\|_{L^2(0, T)}^2}] < \infty
\]

by the Fernique theorem (see Corollary 2.8.6 in [9]). For instance, take \( a(s, x) = \frac{2x}{1 + x^2} \). By the Itô formula, we see that the mapping

\[
\varepsilon \mapsto \int_0^T a(\varepsilon_s) d\varepsilon_s = \ln(1 + \varepsilon_T^2) - \int_0^T \frac{1 - \varepsilon_s^2}{(1 + \varepsilon_s^2)^2} ds
\]

extends to a continuous functional on \( C([0, T]) \). Thus \( y \mapsto \frac{d\mu_{\varepsilon}}{d\mu}(y) \) has a continuous version

\[
\frac{d\mu_{\varepsilon}}{d\mu}(y) = (1 + y_T^2) \exp \left( - \int_0^T \frac{1 - y_s^2}{(1 + y_s^2)^2} ds - \frac{1}{2} \int_0^T \frac{2y_s}{1 + y_s^2} ds \right)
\]

on \( C([0, T]) \). As in (15), we obtain an explicit solution

\[
\mu(U, y) = \frac{\int_U d\mu_{\varepsilon}(y - L(x)) d\mu_{\varepsilon + L(x)}(y) d\mu_X(x)}{\int d\mu_{\varepsilon}(y - L(x)) d\mu_{\varepsilon + L(x)}(y) d\mu_X(x)}
\]
of the statistical inverse problem of estimating the distribution of $X$ given the observation $Y_t = L(X)_t + \varepsilon_t + \int_0^t a(s, \varepsilon_s)ds$ on $[0, T]$.

In general, any $G$-valued random variable $\varepsilon$ whose image measure is absolutely continuous with respect to a zero mean Gaussian measure $\mu$ satisfies

$$\varepsilon = \varepsilon + T(\varepsilon)$$

in distribution for some mapping $T : G \to H_{\mu}$ (see Corollary 4.2 in [11]).

4.4. **Spherically invariant noise.** Let $F$ and $G$ be locally convex Souslin topological vector spaces. We say that $\varepsilon$ is a spherically invariant $G$-valued random variable if $\varepsilon = \gamma Z$, where $Z$ is a zero-mean Gaussian $G$-valued random variable whose Cameron-Martin space is infinite-dimensional, and $Z$ is statistically independent from a non-negative real-valued random variable $\gamma$ whose distribution has no atom at zero.

The expression “spherically invariant random process (SIRP)” is used in the engineering literature [141] while the more descriptive but little used expression “$H_{\mu Z}$-spherically symmetric measure” appears in the mathematical literature (see Definition 7.4.1 in [9]). The latter has emphasis on the fact that the measure is only invariant with respect to orthogonal operators on $H_{\mu Z}$ (see Theorem 7.4.2 in [9]).

In order to study the posterior measure of $X$ given $Y = L(X) + \gamma Z$, we apply an averaging principle together with the following lemma.

**Lemma 4.1.** Let $F$ and $G$ be locally convex Souslin topological vector spaces. Let $Z$ be a zero-mean Gaussian $G$-valued random variable whose Cameron-Martin space is infinite-dimensional. Let $X$ be an $F$-valued random variable, and let $\gamma$ be a non-negative random variable whose distribution has no atom at zero. Suppose that $\gamma$, $X$ and $Z$ are statistically independent.

Let $L : F \to G$ be a continuous mapping such that $L(F) \subset H_{\mu Z}$, where $H_{\mu Z}$ is the Cameron-Martin space of $\mu Z$. Let $\{e_i\}_{i=1}^\infty$ be an orthonormal basis of $H_{\mu Z}$ such that $C_{Z}^{-1}e_i \in G'$, where $C_Z$ is the covariance operator of $Z$. Set $Y = L(X) + \gamma Z$.

For any $f \in L^1(\mu_{Y, \gamma})$, the conditional expectation

$$E[f(Y, \gamma) | \sigma(Y)](\omega) = f(Y(\omega), \gamma Y(\omega))$$

for $P$-almost every $\omega \in \Omega$, where $y \mapsto \gamma_y$ is a $G$-measurable function on $G$ that satisfies

$$\gamma_y = \left( \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \langle y, C_Z^{-1}e_i \rangle^2 \right)^{\frac{1}{2}}$$

whenever a finite limit exists and $\gamma_y = 0$ otherwise.

**Proof.** The mapping $y \mapsto \gamma_y$ is indeed measurable since the set

$$N = \{ y \in G : \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^n \langle y, C_Z^{-1}e_i \rangle^2 \beta \}$$

is a Borel set (see Lemma 2.1.7 in [10]). We show in a moment that $\gamma = \gamma Y$ $P$-almost surely. Then the conditional expectations of $f(Y, \gamma)$ and $f(Y, \gamma Y)$ coincide since the two random variables coincide almost surely. In order to conclude the claim, we note that $\gamma Y(\omega)$ is $Y^{-1}(G)$-measurable as a combination of two measurable functions.

The random variables $\gamma$, $X$ and $Z$ are statistically independent, which implies that their image measure $\mu_{(\gamma, X, Z)}$ is a product measure on the product space $\mathbb{R}_+ \times F \times G$. 
Since \( Z \) has a Gaussian distribution, the random variables \((Z, C_Z^{-1} e_i)\) are statistically independent standard normal random variables. The same holds for the random variables \((t, x, z) \rightarrow (z, C_Z^{-1} e_i)\) on the measure space \((\mathbb{R}_+ \times F \times G, \mathcal{B}(\mathbb{R}_+ \times F \times G), \mu_\gamma \otimes \mu_X \otimes \mu_Z)\). The random variable
\[
(t, x, z) \mapsto L(x) + tz
\]
has the following property. The law of large numbers implies that
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (L(x) + tz, C_Z^{-1} e_i)^2 = \lim_{n \to \infty} \frac{t^2}{n} \sum_{i=1}^{n} (z, C_Z^{-1} e_i)^2 = t^2
\]
for any \( t \in \mathbb{R}_+, x \in F, \) and \( \mu_Z\text{-a.e. } z \in G. \) Since the image measure has the product structure, this also holds for \( \mu_{(\gamma, X, Z)}\)-almost every \((t, x, z).\) Hence,
\[
(\gamma, X, Z)^{-1} \{ (t, x, z) : \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (L(x) + tz, C_Z^{-1} e_i)^2 = t^2 \}
\]
has full \( P \)-measure i.e.
\[
\lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} (L(X) + \gamma Z, C_Z^{-1} e_i)^2 = \gamma^2
\]
P-almost surely. \(\square\)

The averaging principle for the posterior distributions is given in the following lemma. Note, that also topological products of Souslin spaces are Souslin spaces.

**Lemma 4.2.** Let the assumptions of the Lemma 4.1 hold.

A solution \( \mu(\cdot, y) \) of the statistical inverse problem of estimating the distribution of \( X \) given a sample \( y \) of \( Y = L(X) + \gamma Z \) coincides \( \mu_Y \)-almost surely with a Borel measurable solution \( \tilde{\mu}(\cdot, (y, \gamma y)) \) of the statistical inverse problem of estimating the distribution of \( X \) given \( (Y, \gamma) = (y, \gamma y) \), where \( \gamma y \) is defined by (18).

**Proof.** The \( \sigma \)-algebra \( \sigma(Y) \) generated by \( Y = L(X) + \gamma Z \) is a sub-\( \sigma \)-algebra of the \( \sigma \)-algebra \( \sigma((Y, \gamma)) \) generated by the \( G \times \mathbb{R}_+ \)-valued random variable \((Y, \gamma) = (\gamma Z + L(X), \gamma) \). By Lemma 4.1 and a property of conditional expectations, the solutions satisfy
\[
\mu(U, Y) = \mathbb{E}[1_U(X)|\sigma(Y)] = \mathbb{E}[\mathbb{E}[1_U(X)|\sigma(Y, \gamma)]|\sigma(Y)]
\]
\[
= \mathbb{E}[\tilde{\mu}(U, (Y, \gamma))|\sigma(Y)] = \tilde{\mu}(U, (Y, \gamma y))
\]
almost surely for a fixed \( U \in \mathcal{F} \). It is easy to see that \( y \mapsto \tilde{\mu}(U, (y, \gamma y)) \) is Borel-measurable. By the Souslin property, it is enough to consider only countably many \( U \in \mathcal{F} \) in order to identify the two measures. Hence, \((U, y) \mapsto \tilde{\mu}(U, (y, \gamma y))\) is a solution of the statistical inverse problem of estimating the distribution of \( X \) given a sample \( y \) of \( Y = L(X) + \gamma Z \). \(\square\)

**Theorem 4.3.** Let the assumptions of the Lemma 4.1 hold. The essentially unique posterior distribution of \( X \) given a sample \( y \) of \( Y = L(X) + \gamma Z \) has a version
\[
\mu(A, y) = \frac{\int_A \exp \left( \langle y, \gamma y^{-2} C_Z^{-1} L(x) \rangle - \frac{1}{2\gamma^2} \| L(x) \|_{\mu_{\gamma Z}}^2 \right) d\mu_X(x)}{\int_F \exp \left( \langle y, \gamma y^{-2} C_Z^{-1} L(x) \rangle - \frac{1}{2\gamma^2} \| L(x) \|_{\mu_{\gamma Z}}^2 \right) d\mu_X(x)},
\]
for any \( A \subset F \).

}\textit{Inverse Problems and Imaging} Volume 6, No. 2 (2012), 215–266
for all \( y \in G \) such that the limit
\[
\gamma_y = \left( \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \langle y, C_Z^{-1} e_i \rangle^2 \right)^{\frac{1}{2}}
\]
exists and does not vanish.

Proof. Let us calculate the posterior distribution of \( X \) given \((Y, \gamma)\).

The conditional distribution of \((Y, \gamma)\) given a sample \( x \) of \( X \) is \( \mu_{\gamma Z + L(x), \gamma}(C \times B) \), where \( C \in \mathcal{G} \) and \( B \in \mathcal{B}(\mathbb{R}_+) \) by Lemma 3.2. Furthermore, the conditional distribution of \( L(x) + \gamma Z \) given \( \sigma((\gamma, X)) \) is \( \mu_{(\omega_0)Z + L(x)} \). Taking conditional expectations inside the integral gives
\[
\mu_{(\gamma Z + L(x), \gamma)}(C \times B) = P(\gamma Z + L(x) \in C \cap \gamma \in B)
\]
\[
= E[1_C(\gamma Z + L(x))1_B(\gamma)]
\]
\[
= E[1_C(\gamma Z + L(x))|\sigma(\gamma)]1_B(\gamma)]
\]
\[
= \int \mu_{aZ + L(x)}(C)1_B(a)d\mu_\gamma(a).
\]

We may now use the absolute continuity of the translated measures \( \mu_{aZ + L(x)} \) with respect to \( \mu_{aZ} \) which follows from the Cameron-Martin theorem. We obtain
\[
\mu_{(\gamma Z + L(x), \gamma)}(C \times B) = \int_B \int_C e^{\langle y, a^{-2}C_Z^{-1}L(x) \rangle} - \frac{1}{2a^2} \| L(x) \|^2_{H_{\mu Z}} d\mu_{\gamma}(y) d\mu_\gamma(a)
\]
\[
= \int_{C \times B} e^{\langle y, a^{-2}C_Z^{-1}L(x) \rangle} - \frac{1}{2a^2} \| L(x) \|^2_{H_{\mu Z}} d\mu_{\gamma}(y) d\mu_\gamma(a)
\]
\[
= E[1_C \times B(\gamma Z, \gamma)e^{\langle \gamma Z, \gamma^{-2}C_Z^{-1}L(x) \rangle} - \frac{1}{2a^2} \| L(x) \|^2_{H_{\mu Z}} |\sigma(\gamma)]]
\]
\[
= \int_{C \times B} e^{\langle y, a^{-2}C_Z^{-1}L(x) \rangle} - \frac{1}{2a^2} \| L(x) \|^2_{H_{\mu Z}} d\mu_{(\gamma Z, \gamma)}(y, a).
\]

Hence, the Radon-Nikodym derivative of \( \mu_{(\gamma Z + L(x), \gamma)} \) with respect to \( \mu_{(\gamma Z, \gamma)} \) is
\[
\frac{d\mu_{(\gamma Z + L(x), \gamma)}}{d\mu_{(\gamma Z, \gamma)}}(y, a) = \exp \left( \langle y, a^{-2}C_Z^{-1}L(x) \rangle - \frac{1}{2a^2} \| L(x) \|^2_{H_{\mu Z}} \right).
\]

The posterior distribution of \( X \) given a sample \((y, a)\) of \((Y, \gamma)\) has a version
\[
\mu(U, (y, a)) = \int_U \exp \left( \langle y, a^{-2}C_Z^{-1}L(x) \rangle - \frac{1}{2a^2} \| L(x) \|^2_{H_{\mu Z}} \right) d\mu_X(x)
\]
\[
= \int_U \exp \left( \langle y, a^{-2}C_Z^{-1}L(x) \rangle - \frac{1}{2a^2} \| L(x) \|^2_{H_{\mu Z}} \right) d\mu_X(x)
\]
for all \( y \in G \) and \( a \neq 0 \). We obtain the required result by Lemma 4.2. \( \square \)

Remark 10. The posterior distribution (19) does not depend on the distribution of \( \gamma \). Especially, \( \gamma \) does not necessarily have finite moments.

Remark 11. If the sample \( y \in H_{\mu Z} \), then the estimated random number \( \gamma_y = 0 \). Consequently, we can not apply Theorem 2.7 for the solution (19) on any measurable linear subspace of \( G \) of full \( \mu_Z \)-measure, since it contains the Cameron-Martin space \( H_{\mu Z} \). Besides the Lusin theorem, nothing seems to be known about the continuity of the measurable function \( y \mapsto \gamma_y \). Even though the continuity of the posterior distribution as a function of observations remains an open question, we can anticipate from the form of the posterior distribution that the prior distribution will have a good regularizing effect on the corresponding ill-posed inverse problem.
Following [108], we call \( \varepsilon = \gamma Z \) a symmetric \( \alpha \)-stable sub-Gaussian \( G \)-valued random variable if \( \gamma = \sqrt{\Gamma} \), where the non-negative random variable \( \Gamma \) satisfies

\[
E[e^{-t\Gamma}] = e^{-t^{\alpha/2}}, \quad t > 0,
\]

for some \( 0 < \alpha < 2 \), and \( Z \) is a zero mean \( G \)-valued Gaussian random variable.

For instance, \( \alpha \)-stable random variables are used as approximative models for ambient noise. An example of ambient noise is the acoustic noise in oceans originating from e.g. shipping, rain fall, waves, animal activity, bubbles, cracking of ice and geological processes [56, 129]. It disturbs acoustic communication and active acoustic remote sensing in underwater environments [17, 79]. The finite-dimensional distributions of ambient noise are thought to originate from many disturbances occurring in natural environments: typically few strong and a large number of weak disturbances of different orders. The variances of individual disturbances are often such that Lindeberg’s condition, which is a sufficient condition (and in some cases also necessary) for the applicability of the classical central limit theorem, does not hold [135]. A generalized central limit theorem states that a.s. converging sums of independent random variables necessarily have stable distributions (see Definition 1.1.5 in [108]). Non-Gaussian stable distributions exhibit heavy tails, which explains why the Gaussian distributions are not the best ones for modeling ambient noise. Symmetric \( \alpha \)-stable sub-Gaussian random variables are perhaps the most simple subclass of stable distributions.

Sub-Gaussian noise is encountered also in fMRI (functional magnetic resonance imaging), where it models physiological noise, e.g. disturbances originating from breathing and heartbeat [14].

Spherically symmetric noise models are used also as approximative models in high resolution radar imaging for describing the ground-clutter (i.e. unwanted echoes of the transmitted radar signal from the ground), and also sea-clutter (i.e. echoes from the surface of the sea) [19, 20, 21]. It should be noted that the modeling of radar clutter and underwater noise is not yet a mature field of science. Beside of spherically symmetric models also other models have been developed and better models are pursued after.

Noise is usually rougher than the signal by rule of thumb. In the above applications, it is not verified whether this holds for the noise \( \varepsilon \) and signals \( L(x) \), where \( x \in F \). For radar imaging this is not a critical point since the reflected signal acquires some regularity from the transmitted signal.

4.5. Subordinated noise. We consider another generalization of Gaussian noise that is similar to spherically symmetric noise.

Let \( B_t \) be a Brownian motion on \( \mathbb{R}_+ \) satisfying \( B_0 = 0 \) almost surely. Subordinated noise is here defined as a time-changed process

\[
\varepsilon_t = B_{\alpha_t},
\]

where \( \alpha_t \) is a strictly increasing stochastic process that is statistically independent from the Brownian motion \( B_t \). We assume that \( \alpha_t \) has bi-Lipschitz-continuous sample paths and satisfies \( \alpha_0 = 0 \). For example, \( \alpha_t \) can be an integral function of some statistically independent Gamma process starting from a non-zero value. Such a distribution of \( \alpha \) can reflect inaccuracies that are believed to be present in the covariance operator of the noise \( \varepsilon \).

**Lemma 4.4.** The random function \( \varepsilon \) on \([0, 1]\) is a \( C([0, 1]) \)-valued random variable.
Proof. The sample paths of $\varepsilon$ are continuous functions as compositions of continuous functions. Moreover, the space $C([0,1])$ is a separable Fréchet space, which implies that its Borel $\sigma$-algebra is generated by the cylinder sets

$$A = \{ f \in C([0,1]) : f(t_i) \in U_i \forall i \in I \}$$

where $U_i \in \mathcal{B}(\mathbb{R})$, $I \subset \mathbb{N}$ are finite sets, and $\bigcup_{i=1}^{\infty} t_i$ is a dense subset of $[0,1]$. (see Theorem A.3.7 in [9]). It is enough to check that the mapping

$$\omega \mapsto B_{\alpha_i,\omega}(\omega)$$

is a random variable for any $t_i \in [0,1]$. But this follows from the joint measurability of the Brownian motion from $[0,1] \times \Omega$ into $\mathbb{R}$. \hfill $\square$

We take $G = C([0,1])$, $\mathcal{G} = \mathcal{B}(C([0,1]))$, and denote $\mu_\varepsilon(A) = P(\varepsilon \in A)$ for any Borel set $A \subset C([0,1])$.

**Lemma 4.5.** The Gaussian measure $\mu_{B_n(\omega)+L(x)}$ that has mean $L(x)$ and the covariance operator with kernel $C_{\alpha}(\omega)(t,s) = \min(\alpha_i(\omega), \alpha_j(\omega))$ on $[0,1] \times [0,1]$ is a version of the conditional probability $V \mapsto \mathbf{E}[1_{V(B_n+L(x))}|\sigma(\alpha)](\omega)$ on $\mathcal{G}$.

Proof. By defining $B_t = 1_{t \geq 0} B_t$ on $\mathbb{R}$, the Brownian motion extends to a $C(\mathbb{R})$-valued random variable, where $C(\mathbb{R})$ is equipped with the Borel $\sigma$-algebra with respect to the locally convex topology given by the family of seminorms

$$\rho_i(f) = \sup_{t \in K_i} |f(t)|,$$

where $K_i = [-i,i]$ and $i \in \mathbb{N}$ (i.e. the topology of uniform convergence on compact sets). The space $C(\mathbb{R})$ is then a locally convex Souslin space, since its topology is metrizable by a complete metric

$$d(f_1, f_2) = \sum_{i=1}^{\infty} 2^{-i} \frac{\rho_i(f_1 - f_2)}{1 + \rho_i(f_1 - f_2)}$$

and the polynomials with rational coefficients form a dense set by the Stone-Weierstrass theorem. Moreover, $\alpha$ is a $C([0,1])$-valued random variable.

Recalling Lemma 3.2, we need to check that the composition mapping $(f, g) \mapsto f \circ g + L(x)$ is Borel measurable from $C(\mathbb{R}) \times C([0,1])$ into $C([0,1])$. Since point evaluations generate the Borel $\sigma$-algebra of $C([0,1])$, it is enough to show that functionals $(f, g) \mapsto f \circ g(t) + L(x)t$ are Borel measurable for a fixed $t \in [0,1]$. We show that this function is actually continuous. Since the both spaces are metric spaces, it is enough to check the sequential continuity on the product space $C(\mathbb{R}) \times C([0,1])$, which is metrizable.

Let $\lim_{i \to \infty} (f_i, g_i) = (f, g)$ in $C(\mathbb{R}) \times C([0,1])$, which implies that $\lim_{i \to \infty} f_i = f$ and $\lim_{i \to \infty} g_i = g$ in corresponding spaces. Then $K = \{ g_i(t) \in \mathbb{R} : i \in \mathbb{N} \}$ is compact for the fixed $t \in [0,1]$ and

$$|f_i(g_i(t)) - f(g(t))| = |f_i(g_i(t)) - f(g_i(t)) + f(g_i(t)) - f(g(t))|$$

$$< \sup_{i \in K} |f_i(t) - f(t)| + |f(g_i(t)) - f(g(t))| \to 0$$

as $i \to \infty$ by the convergence of $(f_i, g_i)$ and the continuity of $f$. \hfill $\square$

**Theorem 4.6.** Let $F$ be a locally convex Souslin topological vector space equipped with its Borel $\sigma$-algebra $\mathcal{F}$ and let $L : F \to H^1([0,1])$ be a continuous mapping that satisfies $L(x)|_{t=0} = 0$ for all $x \in F$. Let $B_t$ be a Brownian motion on $[0,1]$ starting
from zero. Let $\alpha_t$ be a strictly increasing stochastic process that is statistically independent from the Brownian motion $B_t$ and that has bi-Lipschitz continuous sample paths satisfying $\alpha(0) = 0$ almost surely. Let $X$ be an $F$-valued random variable that is statistically independent from the Brownian motion $B_t$ and the stochastic process $\alpha_t$.

The essentially unique solution of estimating the distribution of $X$ given a sample path $y : [0, 1] \to \mathbb{R}$ of $Y_t = L(X)_t + B_{\alpha_t}$ has a version $\mu$ such that

$$\mu(U, y) = \frac{\int_U \exp \left( \langle y, C_{B_{\alpha_t}}^{-1} L(x) \rangle - \frac{1}{2} \|L(x)\|_{H_{B_{\alpha_t}}}^2 \right) d\mu_X(x)}{\int \exp \left( \langle f, C_{B_{\alpha_t}}^{-1} L(x) \rangle - \frac{1}{2} \|L(x)\|_{H_{B_{\alpha_t}}}^2 \right) d\mu_X(x)}$$

for any $U \in \mathcal{F}$ and for any $y \in C([0, 1])$ such that its quadratic variation $[y]$ satisfies $0 < [y]_t < \infty$ for all $t \in (0, 1]$.

Proof. Let $g$ be some sample of $\alpha$ on $[0, 1]$. The mapping

$$T_g : f \mapsto f \circ g$$

is linear and measurable from $C(\mathbb{R}_+)$ to $C([0, 1])$. Hence, the Cameron-Martin space of $T_g B = B_g$ coincides with $T(H^1_0(\mathbb{R}_+))$ as a vector space (see Theorems 3.7.3 and 3.7.6 in [9]; choose $X = C(\mathbb{R}_+) \times C([0, 1])$ in order to generalize the claim to the present situation). Since $g$ is bi-Lipschitz continuous, the mapping $f \circ g$ is in $H^1(0, 1)$ whenever $f \in H^1(g(0, 1))$ (e.g. Theorem 2.2.2 in [148]), and the mapping is actually onto the subspace $H = \{ f \in H^1(0, 1) : f(0) = 0 \}$. Especially, the vector $L(x) \in H_{B_{\alpha_t}}$ by the assumption, so by Lemma 4.5

$$\frac{d\mu_{T_g B_{\alpha_t} + L(X), \alpha}}{d\mu_{B_{\alpha_t}, \alpha}}(z, \alpha) = \exp \left( \langle z, C_{\alpha_t}^{-1} L(x) \rangle - \frac{1}{2} \|L(x)\|_{H_{B_{\alpha_t}}}^2 \right)$$

as in Theorem 4.3. It is well-known that for continuous time-changes $\alpha_t$ the quadratic variation of $B_{\alpha_t}$ coincides with $\alpha_t$ (see Chapter 5: Proposition 1.5 in [104]). Therefore, $\alpha_t$ is a measurable function of the sample paths of $B_{\alpha_t}$ (the quadratic variation is obtained by taking a limit in probability and we need to pick up a subsequence in order to get the a.s. convergence). Since $L(X) \in H^1(0, 1)$, it has finite variation, which implies that its quadratic variation vanishes. Also $\mu_Y$-almost every sample path of $L(X)_t + B_{\alpha_t}$ has $\alpha_t$ as its quadratic variation. We obtain the claim similarly as in Lemma 4.2.

4.6. Decomposable additive noise. Let $F$ and $G$ be locally convex Souslin topological vector spaces. We say that $G$-valued random noise $\varepsilon$ is decomposable if it is of the form

$$\varepsilon = \sum_{i=1}^{\infty} \varepsilon_i f_i,$$

where $\varepsilon_i$ are independent random variables with a.e. positive probability density functions $\rho_i$ with respect to the Lebesgue measure and $f_i \in G$ are some non-zero vectors.

Remark 12. If $\varepsilon_i$ are random variables and $\varepsilon = \sum_{i=1}^{\infty} \varepsilon_i f_i$ a.s. for some vectors $f_i \in G$, then $\varepsilon$ is a $G$-valued random variable. Indeed, since $G$ is a Souslin topological vector space, the mapping $\mathbb{R} \times G \ni (a, f) \mapsto af =: T(a, f)$ is continuous, therefore also $B(\mathbb{R} \times G) = B(\mathbb{R}) \otimes \mathcal{G}$ measurable. The composition of the measurable mapping $(\omega, f) \mapsto (\varepsilon_i(\omega), f)$ with $T$ gives a $G$-valued random variable $T(\varepsilon_i, f) = \varepsilon_i f$.
the sum of two $G$-valued random variables is a $G$-valued random variable and limits of locally convex Souslin space-valued random variables are random variables (since the cylinder sets generate the Borel $\sigma$-algebra by Theorem 6.8.9 in [10]).

If all possible signals $L(x), x \in F$ are sparse in the sense that they belong to the linear span of $\{f_i : i \in \mathbb{N}\}$ and the noise $\varepsilon$ is decomposable, then the measures $\mu_{\varepsilon + L(x)}$ are absolutely continuous with respect to $\mu_{\varepsilon}$ [109].

Moreover, if $\{f_i\}_{i=1}^{\infty}$ is a basis of the closed subspace $\text{span}(\{f_i : i \in \mathbb{N}\})$, the proof in [109] gives, with minor additional work, an explicit formula for the Radon-Nikodym density. For simplicity, we take $G = \text{span}(\{f_i : i \in \mathbb{N}\})$.

**Theorem 4.7.** Let $G$ be a locally convex Souslin topological vector space equipped with the Borel $\sigma$-algebra $\mathcal{G}$ and a basis $\{f_i\}_{i=1}^{\infty}$ such that the unique coefficients $y_i$ in $y = \sum_{i=1}^{\infty} y_i f_i$ depend measurably on $y \in G$. Let a $G$-valued random variable $\varepsilon$ be of the form $\varepsilon = \sum_{i=1}^{\infty} \varepsilon_i f_i$, where the random variables $\varepsilon_i$ are statistically independent and have probability density functions $\rho_i$ that are a.e. positive. If $L_n(x) = \sum_{i=1}^{n} a_i(x)f_i$, then

$$\frac{d\mu_{\varepsilon + L_n(x)}}{d\mu_{\varepsilon}}(y) = \frac{\prod_{i=1}^{n} \rho_{y_i}(y_i - a_i(x))}{\prod_{i=1}^{n} \rho_{y_i}(y_i)}$$

for almost every $y = \sum_{i=1}^{\infty} y_i f_i$.

**Proof.** Let $A \in \mathcal{G}$. By possibly rearranging finitely many vectors, we may suppose that $L_n(x) = \sum_{i=1}^{n} a_i f_i$. We consider the probability

$$\mu_{\varepsilon + L_n(x)}(A) = P(\varepsilon + L_n(x) \in A) = E \left[1_A \left( \sum_{i=1}^{\infty} \varepsilon_i f_i + \sum_{i=1}^{n} a_i f_i \right) \right].$$

Denote $Z = \sum_{i=n+1}^{\infty} \varepsilon_i f_i$. Following [109], we calculate the conditional expectation of $1_A(\varepsilon + L(X))$ given $\sum_{i=1}^{n} (\varepsilon_i + a_i) f_i$ and, by Lemma 3.2, obtain with straightforward calculations

$$E \left[1_A \left( \sum_{i=1}^{n} \varepsilon_i f_i + \sum_{i=1}^{n} a_i f_i \right) \right] = E \left[ E \left[1_A \left( \sum_{i=1}^{\infty} \varepsilon_i f_i + \sum_{i=1}^{n} (\varepsilon_i + a_i) f_i \right) \right] \right]$$

$$= \int_{\mathbb{R}^{n}} \mu_{Z + \sum_{i=1}^{n} (\varepsilon_i + a_i) f_i} (A) dP(\omega)$$

$$= \int_{\mathbb{R}^{n}} \mu_{Z + \sum_{i=1}^{n} (\varepsilon_i + a_i) f_i} (A) \left( \prod_{i=1}^{n} \rho_i(y_i) \right) dy_1 \cdots dy_n$$

$$= \int_{\mathbb{R}^{n}} \mu_{Z + \sum_{i=1}^{n} a_i f_i} (A) \left( \prod_{i=1}^{n} \rho_i(y_i - a_i) \right) dy_1 \cdots dy_n.$$
At this point, the proof differs from [109]. Namely, we multiply and divide with the positive densities of \( \varepsilon_i \), and obtain

\[
E\left[ 1_A \left( \sum_{i=1}^{\infty} \varepsilon_i f_i + \sum_{i=1}^{n} a_i f_i \right) \right] = \int \mu_{Z+\sum_{i=1}^{n} y_i f_i}(A) \prod_{i=1}^{n} \frac{\rho_i(y_i - a_i)}{\rho_i(y_i)} \rho_i(y_i) dy_1 \cdots dy_n
\]

\[
= \int \mu_{Z+\sum_{i=1}^{n} \varepsilon_i(\omega) f_i}(A) \prod_{i=1}^{n} \frac{\rho_i(\varepsilon_i(\omega) - a_i)}{\rho_i(\varepsilon_i(\omega))} dP(\omega)
\]

\[
= E \left[ E \left[ 1_A \left( \sum_{i=1}^{\infty} \varepsilon_i f_i \right) \sum_{i=1}^{n} \varepsilon_i f_i \prod_{i=1}^{n} \frac{\rho_i(\varepsilon_i - a_i)}{\rho_i(\varepsilon_i)} \right] \right]
\]

\[
= E \left[ 1_A(\varepsilon) \prod_{i=1}^{n} \frac{\rho_i(\varepsilon_i - a_i)}{\rho_i(\varepsilon_i)} \right].
\]

Since the unique coefficients \( \varepsilon_i \) depend measurably on \( \varepsilon \), we may write

\[
\mu_{\varepsilon+L_n(x)}(A) = E \left[ 1_A(\varepsilon) \prod_{i=1}^{n} \frac{\rho_i(\varepsilon_i - a_i)}{\rho_i(\varepsilon_i)} \right]
\]

\[
= \int 1_A(y) \prod_{i=1}^{n} \frac{\rho_i(y_i - a_i)}{\rho_i(y_i)} d\mu_{\varepsilon}(y).
\]

The above theorem verifies the intuitive picture that for sparse signals we may as well study the posterior of \( X \) given the finite-dimensional data

\[
Y_n = \sum_{i=1}^{n} (a_i(X) f_{k_i} + \varepsilon_i f_{k_i}).
\]

The following theorem gives a significant enlargement of applicable noise models in statistical inverse problems.

**Corollary 2** (Generalized Cameron-Martin formula). Let \( G \) be a locally convex Souslin topological vector space equipped with the Borel \( \sigma \)-algebra \( \mathcal{G} \) and a basis \( \{ f_i \}_{i=1}^{\infty} \) such that the unique coefficients \( y_i \) in \( y = \sum_{i=1}^{\infty} y_i f_i \) depend measurably on \( y \in G \). Let a \( G \)-valued random variable \( \varepsilon \) be of the form \( \varepsilon = \sum_{i=1}^{\infty} \varepsilon_i f_i \), where the random variables \( \varepsilon_i \) are statistically independent and have probability density functions \( \rho_i \) that are a.e. positive. If \( L(x) = \sum_{i=1}^{\infty} a_i(x) f_i \) for all \( x \in F \), and densities

\[
\frac{d\mu_{\varepsilon+L_n(x)}}{d\mu_{\varepsilon}}(y) = \prod_{i=1}^{n} \frac{\rho_i(y_i - a_i(x))}{\prod_{i=1}^{n} \rho_i(y_i)}
\]

are uniformly integrable with respect to \( \mu_{\varepsilon} \) and convergent \( \mu_{\varepsilon} \)-almost everywhere, then

\[
\frac{d\mu_{\varepsilon+L(x)}}{d\mu_{\varepsilon}}(y) = \prod_{i=1}^{\infty} \frac{\rho_i(y_i - a_i(x))}{\rho_i(y_i)}
\]

for \( \mu_{\varepsilon} \)-almost every \( y = \sum_{i=1}^{\infty} y_i f_i \).

**Proof.** See Proposition 9.9.10 in [10], which says that if \( \lim_n T_n = T \), where \( T_n \) and \( T \) are measurable mappings on a completely regular space, and the distributions of all \( T_n \) have uniformly integrable Radon-Nikodym densities \( \rho_n \) with respect to the same Radon probability measure \( \nu \), then the distribution of \( T \) has Radon-Nikodym density \( \rho \) with respect to the same Radon probability measure as well.
and $\rho$ is the limit of $\rho_n$ in the weak topology of $L^1(\nu)$. This result is especially applicable to the random variables $T_n(x, z) = L_n(x) + z$ and $T(x, z) = L(x) + z$ on $(F \times G, \mathcal{F} \otimes \mathcal{G}, \mu_X \otimes \mu_Z)$ and the measure $\nu = \mu_Z$. The integrals of the densities over any Borel set converge. By Theorem 4.5.6 and Corollary 4.5.7 in [10] the weak limit coincides with the almost sure limit.

In Corollary 2, the Radon-Nikodym density $\frac{d\mu_{X+L_n(x)}}{d\mu_X}(y)$ has a form similar to Radon-Nikodym densities appearing in the Kakutani dichotomy theorem, which addresses the equivalence and singularity of infinite product measures on $\mathbb{R}^\infty$ [67]. Also Umemura [128] has given conditions for the absolute continuity of measures on abstract spaces when the corresponding finite-dimensional distributions are absolutely continuous. In our case, Umemura’s conditions ask to be a Cauchy sequence in $L^2(\mu_Z)$. We feel that the uniform integrability of the Radon-Nikodym densities is easier to validate than Umemura’s conditions.

**Corollary 3.** Let the assumptions of Corollary 2 hold. The essentially unique posterior distribution of $X$ given a sample $y$ of $Y = L(X) + \varepsilon$ has a version $\mu(\cdot, y)$ such that

$$
\mu(U, y) = \frac{\int_U \prod_{i=1}^\infty \frac{\rho_i(y_i - L(x_i))}{\rho_i(y_i)} d\mu_X(x)}{\int \prod_{i=1}^\infty \frac{\rho_i(y_i - L(x_i))}{\rho_i(y_i)} d\mu_X(x)}
$$

whenever $0 < \int \prod_{i=1}^\infty \frac{\rho_i(y_i - L(x_i))}{\rho_i(y_i)} d\mu_X(x) < \infty$.

**Remark 13.** Signals in non-Gaussian noise may appear in model approximations. Let the true model be

$$
Y = L(X) + \varepsilon
$$

where $\varepsilon = \sum_{i=1}^\infty \varepsilon_i f_i$ and all $\varepsilon_i$ are statistically independent. When the model $L$ is numerically very complicated, the common practice is to replace $L$ with some simpler approximation $L_n$. For example, $L_n$ may have the form $L_n(X) = \sum_{i=1}^n a_i(X) f_i$. Though the true model $L$ and the approximated model $L_n$ are known, the model error $\tilde{\varepsilon} = L(X) - L_n(X)$ is sometimes replaced with a $G$-valued random variable $\varepsilon'$ that has the same distribution as $\tilde{\varepsilon}$ but is statistically independent from $X$ [123]. We note that the observation model is then

$$
Y = L_n(X) + \varepsilon' + \varepsilon,
$$

where $\varepsilon$ represents the uncertainties in the forward model $L_n$. Beside of physical noise, the distribution of the noise may represent our prior beliefs about the uncertainties in the forward model, which do not necessarily have Gaussian distributions.

4.7. Periodic signals in decomposable Laplace noise. In this section, we study an example case of the generalized Cameron-Martin formula for a non-Gaussian noise distribution. A similar distribution has been constructed before by Shimomura [114] who gave conditions under which certain translates of the distribution were equivalent to the original distribution. However, we use the methods of Section 4.6.

One class of inverse problems that involves periodic signals are the inverse scattering problems – the far-field pattern of the scattered wave in the 2D fixed energy inverse acoustic or potential scattering problem is a function on the torus. The measured far-field pattern is possibly contaminated by instrumental noise, far-fields
of other unknown incoming fields, contributions from other scatterers, and the near-field and plane wave approximation errors. Although the random model below is oversimplified to fully cover this case, it shows how periodicity can be utilized in Bayesian inverse problems.

Suppose that $L(x) \in C^\alpha(S^1)$ for all $x \in F$ and some $\alpha > 1$. Then the Fourier coefficients

$$
\hat{L}(x)_k = \frac{1}{2\pi} \int_0^{2\pi} L(x; t)e^{-ikt} dt
$$

are $\ell_1$-summable and the corresponding Fourier series converges to the limit

$$
L(x; t) = \sum_{k=-\infty}^{\infty} \hat{L}(x)_k e^{ikt}
$$

in $C(S^1)$ (equipped with the usual supremum norm).

Let $\varepsilon_k$ be mutually statistically independent random variables whose probability density functions with respect to the Lebesgue measure are

$$
\rho_k(t) = \frac{1}{2b} e^{-|t|/b}
$$

for all $k \in \mathbb{Z}$ and some common $b > 0$ i.e. they are zero mean Laplace random variables. The relation of the normal distribution to the Laplace distribution is that a conditionally normal random variable $\tilde{\varepsilon}_k|\sigma \sim N(0, \sigma^2)$ with a Rayleigh distributed variance has a Laplace distribution. In statistical inverse problems, one interpretation of the Laplace distribution is that we do not know the error variance exactly and are lead to describe our lack of knowledge in the form of a probability distribution.

**Lemma 4.8.** The random sum

$$
\varepsilon = \sum_{k=-\infty}^{\infty} \varepsilon_k e^{ikt}
$$

converges in $H^{-1}(S^1)$ and

$$
\frac{1}{2\pi} \langle \varepsilon, e^{-ikt} \rangle_{H^{-1}(S^1), H^1(S^1)} = \varepsilon_k
$$

**Proof.** The space $H^{-1}(S^1)$ is a Hilbert space, and the random variables $\varepsilon_k$ have zero mean so it suffices to prove that $\sum_k E[\|\varepsilon_k e^{ikt}\|_{H^{-1}(S^1)}^2] < \infty$ (see Theorem 2 in Chapter 3.2 in [63]). The sequence $\{e^{ikt}\}_{k=-\infty}^{\infty}$ forms an orthonormal basis of $L^2(S^1)$. The imbedding of $L^2(S^1)$ into $H^{-1}(S^1)$ is Hilbert-Schmidt by Maurin’s theorem, which implies that

$$
\sum_k E[\|e^{ikt}\|_{H^{-1}(S^1)}^2] < \infty
$$

and therefore

$$
E \sum_{k=-\infty}^{\infty} |\varepsilon_k|^2 \|e^{ikt}\|_{H^{-1}(S^1)}^2 \leq C \sum_k \|e^{ikt}\|_{H^{-1}}^2 < \infty.
$$

Here we used the fact that the variance of the Laplace random variable $\varepsilon_k$ is $2b^2$. $\square$
Next, we quickly check that $\varepsilon$ is a non-Gaussian random variable. The characteristic function of the random variable $\varepsilon$ is

$$\hat{\mu}_\varepsilon(\phi) = \mathbb{E}[e^{i\langle \varepsilon, \phi \rangle}] = \prod_{k=-\infty}^{\infty} \frac{1}{1 + b^2(2\pi \hat{\phi}_k)^2},$$

where $\hat{\phi}_k$ is the Fourier coefficient $\frac{1}{2\pi} \int_0^{2\pi} \phi(t)e^{-ikt}dt$ of $\phi \in H^1(S^1)$. Note, that when $\phi_j \to \psi \in L^2(S^1)$ as $j \to \infty$, then

$$\lim_{j \to \infty} \hat{\mu}_\varepsilon(\phi_j) = \prod_{k=-\infty}^{\infty} \frac{1}{1 + b^2(2\pi \hat{\psi}_k)^2} =: \hat{\mu}_\varepsilon(\psi),$$

i.e. their distributions converge weakly. Especially, when $\hat{\psi}_k = \frac{1}{\pi(\frac{|k|}{2} - \frac{1}{2})}$, $k \neq 0$ and $\hat{\psi}_0 = 0$, then

$$\hat{\mu}_\varepsilon(t\psi) = \frac{1}{\cosh^2(2\pi bt)}.$$

Since the weak limits of zero mean Gaussian distributions are always zero mean Gaussian distributions, this shows that $\varepsilon$ is indeed a non-Gaussian random variable.

We wish to study the statistical inverse problem of estimating the probability distribution of $X$ when a sample of $Y = L(X) + \varepsilon$ is known. This means that the inexact observations of the Fourier coefficients of $L(X)$ are assumed to be similarly inaccurate and some components are allowed to have high inaccuracies. Since Laplace distribution has heavier tails than the Gaussian distribution, it protects against outliers better than the normal distribution.

Consider first finite sums

$$L_n(x; t) = \sum_{|k| \leq n} \hat{L}(x)_k e^{ikt}.$$

By Lemma 4.8 and Theorem 4.7, the Radon-Nikodym derivative of the translated measure $\mu_{\varepsilon+L_n(x)}$ with respect to the measure $\mu_\varepsilon$ on $H^{-1}(S^1)$ is

$$\frac{d\mu_{\varepsilon+L_n(x)}}{d\mu_\varepsilon}(y) = \prod_{k=-n}^{n} \frac{\rho_k(\hat{y}_k - \hat{L}(x)_k)}{\rho_k(\hat{y}_k)} = \prod_{k=-n}^{n} e^{-b^{-1}|\hat{\phi}_k - \hat{L}(x)_k| + b^{-1}|\hat{\phi}_k|},$$

By the triangle inequality, we obtain that

$$||\hat{y}_k| - |\hat{y}_k - \hat{L}(x)_k|| \leq |\hat{L}(x)_k|,$$

which are summable. Therefore, the limit

$$\sum_{k=-\infty}^{\infty} (||\hat{y}_k| - |\hat{y}_k - \hat{L}(x)_k||)$$

exists.

Random variables $\varepsilon + L_n(x)$ converge almost surely to $\varepsilon + L(x)$. Therefore, corresponding measures converge weakly i.e. for all Borel sets $A$ whose boundary
is $\mu_{\varepsilon+L(x)}$-zero measurable it holds that
\[
\mu_{\varepsilon+L(x)}(A) = \lim_{n \to \infty} \mu_{\varepsilon+L_n(x)}(A)
\]
\[
= \lim_{n \to \infty} \int_A e^{b^{-1} \sum_{k=-\infty}^{\infty} (|\hat{y}_k|-|\hat{y}_k-L(x)_k|)} d\mu_{\varepsilon}(y)
\]
\[
= \int_A e^{b^{-1} \sum_{k=-\infty}^{\infty} (|\hat{y}_k|-|\hat{y}_k-L(x)_k|)} d\mu_{\varepsilon}(y)
\]
by the Lebesgue dominated convergence theorem. The exponential function is the Radon-Nikodym derivative $\frac{d\mu_{\varepsilon+L(x)}}{d\mu_{\varepsilon}}(y)$ and $1 \leq \frac{d\mu_{\varepsilon+L(x)}}{d\mu_{\varepsilon}}(y) \leq e^{b^{-1} \sum_{k=1}^{\infty} |L(x)_k|}$.

We have shown the following theorem.

**Theorem 4.9.** Let $F$ be a locally convex Souslin topological vector space equipped with its Borel $\sigma$-algebra. Let $X$ be an $F$-valued random variable. Let
\[
\varepsilon = \sum_k \varepsilon_k e^{ikt}
\]
be a $H^{-1}(S^1)$-valued random variable such that all $\varepsilon_k$, where $k \in \mathbb{Z}$, are mutually statistically independent random variables with probability density functions
\[
\rho_k(t) = \frac{1}{2b} e^{-|t|/b}
\]
with respect to the Lebesgue measure for some $b > 0$. Let $L : F \to C^\alpha(S^1)$ be a continuous mapping for some $\alpha > 1$ such that $E[e^{b^{-1} \sum_{k=1}^{\infty} |L(X)_k|}] < \infty$.

The solution of the statistical inverse problem of estimating the distribution of $X$ given a sample $y \in H^{-1}(S^1)$ of $Y = L(X) + \varepsilon$ is essentially unique and has a version $\mu$ such that
\[
\mu(U, y) = \frac{\int_{U} e^{b^{-1} \sum_{k=-\infty}^{\infty} (|\hat{y}_k|-|\hat{y}_k-L(x)_k|)} d\mu_X(x)}{\int e^{b^{-1} \sum_{k=-\infty}^{\infty} (|\hat{y}_k|-|\hat{y}_k-L(x)_k|)} d\mu_X(x)}
\]
for all $U \in \mathcal{F}$ and for all $y \in H^{-1}(S^1)$.

If $E[\exp \left( b^{-1} \|L(X)\|_\rho \right)] \leq C$ for some $C > 0$, the posterior distributions depend continuously on the observations. Indeed, the posterior distributions are sequentially continuous by the Lebesgue dominated convergence theorem and the continuity of $y \mapsto \hat{y}_k$, and sequentially continuous functions on Hilbert spaces are continuous. The topological support of $\mu_Y$ coincides with the topological support of $\mu_{\varepsilon}$ by Lemma 3.5. The topological support of $\mu_{\varepsilon}$ coincides with the closure of the linear span of $\{e^{ikt}\}_k$ in $H^{-1}(S^1)$, that is, $H^{-1}(S^1)$ (see [109]). By Theorem 2.7, $\mu$ is the only posterior distribution that depends continuously on the observations.

5. **Conclusions.** The explicit expressions of posterior distributions derived in Section 4 for simple non-Gaussian noise models may serve as model cases for further studies on the effects of non-Gaussianity of the noise distribution. In Example 4.6, a Kakutani type generalization of the Cameron-Martin formula was derived. We anticipate that this generalization opens a way for a wide class of non-Gaussian noise models in statistical inverse problems, especially when used in connection with the wavelet expansions. Example 4.4 demonstrates the surprising fact that the posterior distribution given an infinite-dimensional observation can have significantly simpler expression than the posterior distribution given a corresponding finite-dimensional
observation. This suggest that in some cases the infinite-dimensional model could provide new numerical approximations schemes.

It is well-known that the generalized Bayes formula holds when the measures $\mu_{x + L(x)}$ are $\mu_X$-almost surely dominated i.e. absolutely continuous with respect to some $\sigma$-finite measure for $\mu_X$-almost every $x \in F$. We showed that there is a curious interplay between the continuity of posterior distributions with respect to the observations and the $\mu_X$-a.s. domination of $\mu_{x + L(x)}$ (cf. Theorem 2.8 and Remark 4). The continuity of the posterior distribution with respect to observations is only possible in the dominated case, which means that in the undominated cases some posterior distributions have discontinuities. Moreover, the regularizing effect of the prior distributions on an ill-posed inverse problem could be of limited power.

Continuity of the posterior distributions with respect to the observations has also other roles in statistical inverse problems. It helps to reduce the nonuniqueness of posterior distributions in quite general cases (cf. Theorem 2.7).

Acknowledgments. This work was supported by the Academy of Finland (application number 213476, Finnish Programme for Centres of Excellence in Research 2006-2011).

REFERENCES

[1] F. Abramovich, T. Sapatinas and B. W. Silverman, Wavelet thresholding via a Bayesian approach, J. R. Stat. Soc. Ser. B Stat. Methodol., 60 (1998), 725–749.
[2] F. Abramovich and B. W. Silverman, Wavelet decomposition approaches to statistical inverse problems, Biometrika, 85 (1998), 115–129.
[3] F. Abramovich, T. Sapatinas and B. W. Silverman, Stochastic expansions in an overcomplete wavelet dictionary, Probab. Theory Related Fields, 117 (2000), 133–144.
[4] G. Backus, Isotropic probability measures in infinite-dimensional spaces, Proc. Nat. Acad. Sci. U.S.A., 84 (1987), 8755–8757.
[5] A. Barron, M. J. Schervish and L. Wasserman, The consistency of posterior distributions in nonparametric problems, Ann. Statist., 27 (1999), 536–561.
[6] J.-M. Bernardo and A. F. M. Smith, “Bayesian Theory,” Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics, John Wiley & Sons, Chichester, 1994.
[7] A. B. Bhatt, G. Kallianpur and R. L. Karandikar, Robustness of the nonlinear filter, Stochastic Process. Appl., 81 (1999), 247–254.
[8] N. Bissantz and H. Holzmann, Statistical inference for inverse problems, Inverse Problems, 24 (2008), 034009, 17 pp.
[9] V. I. Bogachev, “Gaussian Measures,” Mathematical Surveys and Monographs, 62, American Mathematical Society, Providence, RI, 1998.
[10] V. I. Bogachev, “Measure Theory,” Vol. I, II, Springer-Verlag, Berlin, 2007.
[11] V. I. Bogachev, A. V. Kolesnikov and K. V. Medvedev, Triangular transformations of measures, Sb. Math., 196 (2005), 309–335.
[12] V. V. Buldygin, On invariant Bayesian estimators for generalized random variables, Theor. Probability Appl., 22 (1977), 172–175.
[13] J. P. Burgess and R. D. Mauldin, Conditional distributions and orthogonal measures, Ann. Probab., 9 (1981), 902–906.
[14] V. D. Calhoun and T. Adali, Unmizing fMRI with independent component analysis, IEEE Engineering in Medicine and Biology Magazine, 25 (2006), 79–90.
[15] L. Cavalier, Inverse problems with non-compact operators, J. Statist. Plann. Inference, 136 (2006), 390–400.
[16] L. Cavalier, Nonparametric statistical inverse problems, Inverse Problems, 24 (2008), 034004, 19 pp.
[17] M. A. Chitre, J. R. Potter and Ong Sim-Heng, Optimal and near-optimal signal detection in snapping shrimp dominated ambient noise, IEEE J. Ocean. Eng., 31 (2006), 497–503.
[18] N. Choudhuri, S. Ghosal and A. Roy, *Bayesian methods for function estimation*, “Bayesian Thinking: Modeling and Computation” (eds. D. K. Dey, et al), Handbook of Statist., 25, Elsevier/North-Holland, Amsterdam, (2005), 373–414.

[19] E. Conte and M. Longo, *Characterisation of radar clutter as a spherically invariant random process*, IEE Proc. Part F, 134 (1987), 191–197.

[20] E. Conte, M. Longo and M. Lops, *Modelling and simulation of non-Rayleigh radar clutter*, IEE Proc. Part F, 138 (1991), 121–130.

[21] E. Conte and A. De Maio, *Mitigation techniques for non-Gaussian sea clutter*, IEEE J. Ocean. Eng., 29 (2004), 284–302.

[22] S. L. Cotter, M. Dashti, J. C. Robinson and A. M. Stuart, *Bayesian inverse problems for functions and applications to fluid mechanics*, Inverse Problems, 25 (2009), 115008, 43 pp.

[23] S. L. Cotter, M. Dashti and A. M. Stuart, *Approximations of Bayesian inverse problems for PDEs*, SIAM J. Numer. Anal., 48 (2010), 322–345.

[24] D. D. Cox, *An analysis of Bayesian inference for nonparametric regression*, Ann. Statist., 21 (1993), 903–923.

[25] W. B. Davenport, Jr. and W. L. Root, “An Introduction to the Theory of Random Signals and Noise,” McGraw-Hill Book Company, Inc., New York-Toronto-London, 1958.

[26] P. Diaconis and D. Freedman, *On the consistency of Bayes estimates*, Ann. Statist., 14 (1986), 1–67.

[27] P. Diaconis and D. Freedman, *Consistency of Bayes estimates for nonparametric regression: Normal theory*, Bernoulli, 4 (1998), 411–444.

[28] J. Diestel and J. J. Uhl, Jr., “Vector Measures,” With a foreword by B. J. Pettis, Mathematical Surveys, No. 15, American Mathematical Society, Providence, RI, 1977.

[29] J. Dieudonné, *Un exemple d’espace normal non susceptible d’une structure uniforme d’espace complet*, C. R. Acad. Sci. Paris, 209 (1939), 145–147.

[30] J. Dieudonné, *Sur le théorème de Lebesgue-Nikodým. III*, Ann. Univ. Grenoble. Sect. Sci. Math. Phys. (N.S.), 23 (1948), 25–53.

[31] J. L. Doob, *Stochastic processes depending on a continuous parameter*, Trans. Amer. Math. Soc., 42 (1937), 107–140.

[32] J. L. Doob, *Stochastic processes with an integral-valued parameter*, Trans. Amer. Math. Soc., 44 (1938), 87–150.

[33] J. L. Doob, *Application of the theory of martingales*, in “Le Calcul des Probabilités et ses Applications,” Colloques Internationaux du Centre National de la Recherche Scientifique, no. 13, Centre National de la Recherche Scientifique, Paris, (1949), 23–27.

[34] M. Fréchet, *On two new chapters in the theory of probability*, Math. Mag., 22 (1948), 1–12.

[35] D. Freedman, *On the asymptotic behavior of Bayes’ estimates in the discrete case*, Ann. Math. Statist., 34 (1963), 1386–1403.

[36] I. M. Gel’fand and N. Ya. Vilenkin, “Generalized Functions. Vol. 4: Applications of Harmonic Analysis,” Academic Press, New York-London, 1964.
Inverse Problems and Imaging Volume 6, No. 2 (2012), 215–266

[47] J. K. Ghosh and R. V. Ramamoorthi, “Bayesian Nonparametrics,” Springer Series in Statistics, Springer-Verlag, New York, 2003.
[48] I. I. Gihman and A. V. Skorohod, “The Theory of Stochastic Processes. I,” Die Grundlehren der mathematischen Wissenschaften, Band 210, Springer-Verlag, New York-Heidelberg, 1974.
[49] P. Gravel, G. Beaudoin and J. A. De Guise, A method for modeling noise in medical images, IEEE Trans Med Imaging., 23 (2004), 1211–1232.
[50] U. Grenander, Stochastic processes and statistical inference, Ark. Mat., 1 (1950), 195–277.
[51] P. R. Halmos and L. J. Savage, Application of the Radon-Nikodym theorem to the theory of sufficient statistics, Ann. Math. Statist., 20 (1949), 225–241.
[52] K. Harada and H. Saigo, The space of tempered distributions as a k-space, preprint, arXiv:1009.1429.
[53] M. Hegland, Approximate maximum a posteriori with Gaussian process priors, Constr. Approx., 26 (2007), 205–224.
[54] T. Helin, On infinite-dimensional hierarchical probability models in statistical inverse problems, Inverse Probl. Imaging., 3 (2009), 567–597.
[55] T. Helin and M. Lassas, Hierarchical models in statistical inverse problems and the Mumford-Shah functional, Inverse problems, 27 (2011), 015008, 32 pp.
[56] J. A. Hildebrand, The theory of integration in a space of an infinite number of dimensions, Acta Math., 63 (1934), 249–323.
[57] M. Jiřina, On regular conditional probabilities, Czechoslovak Math. J., 9 (1959), 445–451.
[58] M. Jiřina, Conditional probabilities on σ-algebras with countable basis, in “Select. Transl. Math. Statist. and Probability,” Vol. 2, American Mathematical Society, Providence, RI, (1962), 79–86.
[59] I. M. Johnstone and B. W. Silverman, Speed of estimation in positron emission tomography and related inverse problems, Ann. Statist., 18 (1990), 251–280.
[60] G. Kallianpur and C. Striebel, Estimation of stochastic systems: Arbitrary system process with additive white noise observation errors, Ann. Math. Statist., 39 (1968), 785–801.
[61] K. Karhunen, Über lineare Methoden in der Wahrscheinlichkeitsrechnung, Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys., 1947 (1947), 79 pp.
[62] E. J. Kelly, I. S. Reed and W. L. Root, The detection of radar echoes in noise. I, II, J. Soc. Indust. Appl. Math., 8 (1960), 309–341, 481–507.
[63] G. S. Kimeldorf and G. Wahba, A correspondence between Bayesian estimation on stochastic processes and smoothing by splines, Ann. Math. Statist., 41 (1970), 495–502.
[64] V. Kolehmainen, M. Lassas, K. Niinimäki and S. Siltanen, Sparsity-promoting Bayesian inversion, preprint, 2011.
[65] A. Kolmogoroff, “Grundbegriffe der Wahrscheinlichkeitsrechnung,” Springer, Berlin, 1933.
[66] A. Kolmogorov, Stationary sequences in Hilbert’s space (Russian), Bolletin Moskovskogo Gosudarstvenogo Universiteta, Matematika, 2 (1941), 40 pp.
[67] M. Krein, On a generalization of some investigations of G. Szegő, V. Smirnoff and A. Kolmogoroff, C. R. (Doklady) Acad. Sci. URSS (N.S.), 46 (1945), 91–94.
[68] M. Krein, On a problem of extrapolation of A. N. Kolmogoroff, C. R. (Doklady) Acad. Sci. URSS (N. S.), 46 (1945), 306–309.

Inverse Problems and Imaging Volume 6, No. 2 (2012), 215–266
[77] P. Krug, *The conditional expectation as estimator of normally distributed random variables with values in infinitely-dimensional Banach spaces*, J. Multivariate Anal., 38 (1991), 1–14.

[78] H. H. Kuo, “Gaussian Measures in Banach Spaces,” Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin-New York, 1975.

[79] E. E. Kuruoglu, W. J. Fitzgerald and P. J. W. Rayner, *Near optimal detection of signals in impulsive noise modeled with a symmetric α-stable distribution*, IEEE Communications Letters, 2 (1998), 282–284.

[80] S. Lasanen, “Discretizations of Generalized Random Variables With Applications to Inverse Problems,” Dissertation, Ann. Acad. Sci. Fenn. Math. Diss., No. 130, University of Oulu, Oulu, 2002.

[81] M. Lassas, E. Saksman and S. Siltanen, *Discretization-invariant Bayesian inversion and Besov space priors*, Inverse Probl. Imaging, 3 (2009), 87–122.

[82] M. Lassas and S. Siltanen, *Can one use total variation prior for edge-preserving Bayesian inversion?*, Inverse Problems, 20 (2004), 1537–1563.

[83] M. Ledoux and M. Talagrand, “Probability in Banach Spaces. Isoperimetry and Processes,” Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], 23, Springer-Verlag, Berlin, 1991.

[84] M. Lehtinen, B. Damtie, P. Piiroinen and M. Orispää, *Perfect and almost perfect pulse compression codes for range spread radar targets*, Inverse problems, 3 (2009), 465–486.

[85] M. Lehtinen, L. Päivärinta and E. Somersalo, *Linear inverse problems for generalised random variables*, Inverse Problems, 5 (1989), 599–612.

[86] M. Lewandowski, M. Ryznar, and T. Žak, *Anderson inequality is strict for Gaussian and stable measures*, Proc. Amer. Math. Soc., 123 (1995), 3875–3880.

[87] H. Luschgy, *Linear estimators and Radonifying operators*, Theory Probab. Appl., 40 (1995), 167–175.

[88] C. Macci, *On the Lebesgue decomposition of the posterior distribution with respect to the prior in regular Bayesian experiments*, Statist. Probab. Lett., 26 (1996), 147–152.

[89] P. K. Mandal and V. Mandrekar, *A Bayes formula for Gaussian noise processes and its applications*, SIAM J. Control Optim., 39 (2000), 852–871.

[90] A. Mandelbaum, *Linear estimators and measurable linear transformations on a Hilbert space*, Z. Wahrsch. Verw. Gebiete, 65 (1984), 385–397.

[91] P. Müller and F. A. Quintana, *Nonparametric Bayesian data analysis*, Statist. Sci., 19 (2004), 95–110.

[92] A. Neubauer and H. K. Pikkarainen, *Convergence results for the Bayesian inversion theory*, J. Inverse Ill-Posed Probl., 16 (2008), 601–613.

[93] J. Neveu, “Discrete-Parameter Martingales,” Revised edition, North Holland Mathematical Library, Vol. 10, North-Holland Publishing Co., Amsterdam-Oxford, American Elsevier Publishing Co., Inc., New York, 1975.

[94] B. Øksendal, “Stochastic Differential Equations. An Introduction with Applications,” Sixth edition, Universitext, Springer-Verlag, Berlin, 2003.

[95] F. O’Sullivan, *A statistical perspective on ill-posed inverse problems*, Statist. Sci., 1 (1986), 502–527.

[96] K. R. Parthasarathy, “Probability Measures on Metric Spaces,” Reprint of the 1967 original, AMS Chelsea Publishing, Providence, RI, 2005.

[97] B. J. Pettis, *On integration in vector spaces*, Trans. Amer. Math. Soc., 44 (1938), 277–304.

[98] D. L. Phillips, *A technique for the numerical solution of certain integral equations of the first kind*, Journal of the ACM, 9 (1962), 84–97.

[99] P. Piirinen, “Statistical Measurements, Experiments and Applications,” Dissertation, Ann. Acad. Sci. Fenn. Math. Diss., No. 143, University of Helsinki, Helsinki, 2005.

[100] H. Poincaré, “Science and Hypothesis,” Walter Scott Publishing, London, 1905.

[101] H. Poincaré, “Calcul des Probabilités,” Reprint of the second (1912) edition, Les Grands Classiques Gauthier-Villars, Éditions Jacques Gabay, Sceaux, 1987.

[102] P. M. Prenter and C. R. Vogel, *Stochastic inversion of linear first kind integral equations. I. Continuous theory and the stochastic generalized inverse*, J. Math. Anal. Appl., 106 (1985), 202–218.

[103] D. Ramachandran, *A note on regular conditional probabilities in Doob’s sense*, Ann. Probab., 9 (1981), 907–908.
Inverse Problems and Imaging Volume 6, No. 2 (2012), 215–266

[129] R. J. Urick, “Ambient Noise in the Sea,” Undersea Warfare Technology Office, Naval Sea Systems Command, Dept. of the Navy, Washington, D.C., 1984.

[130] N. N. Vakhania, V. I. Tarieladze and S. A. Chobanyan, “Probability Distributions on Banach Spaces,” Reidel Publishing Co., Dordrecht, 1987.

[131] A. W. van der Vaart and J. H. van Zanten, Rates of contraction of posterior distributions based on Gaussian process priors, Ann. Statist., 36 (2008), 1435–1463.

[132] V. S. Varadarajan, “Measures on Topological Spaces,” Amer. Math. Soc. Transl., 2 (1965), 161–220.

[133] G. Wahba, Improper priors, spline smoothing and the problem of guarding against model errors in regression, J. Roy. Statist. Soc. Ser. B, 40 (1978), 364–372.

[134] S. G. Walker, P. Damien, P. W. Laud and A. F. M. Smith, Bayesian nonparametric inference for random distributions and related functions, With discussion and a reply by the authors, J. R. Stat. Soc. Ser. B Stat. Methodol., 61 (1999), 485–527.

[135] R. J. Webster, Ambient noise statistics, IEEE Trans. Signal Process., 41 (1993), 2249–2253.

[136] P. Whittle, Curve and periodogram smoothing, J. Roy. Statist. Soc. Ser. B, 19 (1957), 38–47.

[137] P. Whittle, On the smoothing of probability density functions, J. Roy. Statist. Soc. Ser. B, 20 (1958), 334–343.

[138] N. Wiener, “Extrapolation, Interpolation, and Smoothing of Stationary Time Series. With Engineering Applications,” Chapman & Hall, Ltd., London, 1949.

[139] N. Wiener, “Collected Works. Vol. I” (ed. P. Masani), MIT Press, Cambridge, Mass.-London, 1976.

[140] S. Willard, “General Topology,” Dover Publications Inc., Mineola, NY, 2004.

[141] G. Wise and N. Gallagher, On spherically invariant random processes, IEEE Trans. Information theory, 24 (1978), 118–120.

[142] R. L. Wolpert and K. Ickstadt, Reflecting uncertainty in inverse problems: A Bayesian solution using Lévy processes, Inverse Problems, 20 (2004), 1759–1771.

[143] R. L. Wolpert, K. Ickstadt and M. B. Hansen, A nonparametric Bayesian approach to inverse problems, in “Bayesian Statistics,” 7, Oxford Univ. Press, New York, (2003), 403–417.

[144] D. X. Xia, “Measure and Integration Theory on Infinite-Dimensional Spaces,” Academic Press, New York-London, 1972.

[145] Y. Xing and B. Ranneby, Sufficient conditions for Bayesian consistency, J. Statist. Plann. Inference, 139 (2009), 2479–2489.

[146] Y. Yamasaki, “Measures on Infinite-Dimensional Spaces,” World Scientific Publishing Co., Singapore, 1985.

[147] L. H. Zhao, Bayesian aspects of some nonparametric problems, Ann. Statist., 28 (2000), 532–552.

[148] W. P. Ziemer, “Weakly Differentiable Functions. Sobolev Spaces and Functions of Bounded Variation,” Springer-Verlag, New York, 1989.

Received June 2009; revised January 2012.

E-mail address: sari.lasanen@oulu.fi