Bayesian inference for Gaussian models: Inverse problems and evolution equations

Yan, D.

Chapter 8

Inverse Problems with Discrete Observations in Smoothness Scales

In this chapter, we continue the study in Chapter 6. In practice one usually does not have access to a 'continuous' observation Y, but only records noisy samples of the unknown function $\mathcal{A}f$ at a finite number of locations in its domain. This is the situation that we consider in the present chapter. To cope with discrete observations, we consider the operator $\mathcal{A}$ introduced in Section 5.2 with a specified domain. Assume that $\mathcal{A}: H \to G$ is a linear mapping from a Hilbert space H into a pre-Hilbert space $G = \mathcal{L}^2(\mathfrak{D})$ of square-integrable functions $g: \mathfrak{D} \to \mathbb{R}^d$ on a bounded domain $\mathfrak{D} \subset \mathbb{R}^d$. Then, the observation is formed as follows. For a given set of design points

$$\mathfrak{D}_n := \{x_1, \ldots, x_n\} \subset \mathfrak{D}. \tag{8.1}$$

we observe the vector $Y^n = (Y_1, \ldots, Y_n)$ defined by

$$Y_i = \mathcal{A}f(x_i) + Z_i, \qquad i = 1, \cdots, n, \tag{8.2}$$

with Z_i i.i.d. standard normal random variables. We wish to estimate f from the observations $(x_i, Y_i)_{1 \le i \le n}$.

This chapter extends results for the white noise model obtained in Chapter 6 to the case of discrete observations. Although we repeat some necessary definitions, we refer to the mentioned chapter for further examples and discussion. This chapter is organized as follows. Section 8.1 demonstrates a procedure to reconstruct continuous signals from discrete observations. With the help of the just mentioned reconstruction procedure and the projection method from Section 5.3, we present a general contraction theorem for the regression model in Section 8.2. Then, the same priors considered in Chapter 6 are studied: series priors and Gaussian priors in Section 8.3 and Section 8.4, respectively. In addition, Gaussian mixture priors are introduced to obtain adaptation in Section 8.5. In the end, Section 8.6 contains the proofs.

8.1 Signal Reconstruction

The sampling scheme (8.2) collects discrete data, but the operator $\mathcal{A}$ acts on a continuous function, which we wish to estimate on its full domain. In this section we describe an interpolation technique that maps discrete signals to the continuous domain. A similar technique has been used in the context of proving asymptotic equivalence of the white noise model and nonparametric regression; see [82] and the references therein. The assumptions are also inspired by the theory from the field of numerical analysis, see [12].

The range space G of the operator $\mathcal{A}: H \to G$ is a collection of functions $g: \mathcal{D} \to \mathbb{R}$, equipped with a pre-inner product $\langle \cdot, \cdot \rangle$. The design points (8.1) give rise to a "discrete" semi-inner product and semi-norm on G given by

$$\langle g, h \rangle_n := \frac{1}{n} \sum_{i=1}^{n} g(x_i) h(x_i) \qquad \|g\|_n := \sqrt{\frac{1}{n} \sum_{i=1}^{n} g^2(x_i)}.$$

(The notation $\|\cdot\|_n$ clashes with the notation $\|\cdot\|_s$ for the norms of the smoothness scales, but this should not lead to confusion as n will never appear as smoothness level.)

For every $n \in \mathbb{N}$ we fix an n-dimensional subspace $\widetilde{W}_n \subset G$ with two properties.

Assumption 8.1 (Interpolation)**.**

(i) There exist constants $0 < C_1 < C_2 < \infty$, independent of n, such that

$$C_1 \|w\| \leq \|w\|_n \leq C_2 \|w\|, \qquad w \in \widetilde{W}_n. \tag{8.3}$$

(ii) For every $g \in G$ the unique element $\mathcal{I}_n g$ of $\widetilde{W}_n$ that interpolates g at the design points, i.e. $\mathcal{I}_n g(x_i) = g(x_i)$, for every $i = 1, \ldots, n$, satisfies, for every s in some interval (s_d, S_d),

$$\|\mathcal{I}_n g - g\| \lesssim \delta_d(n, s) \|g\|_s. \tag{8.4}$$

Condition (8.3) requires that the discrete and continuous norms be equivalent on the subspace $\widetilde{W}_n$, whereas (8.4) ensures that the subspaces $\widetilde{W}_n$ have good approximation properties under discretization for smooth functions. In this condition $\|\cdot\|_s$ are the norms of a smoothness scale $(G_s)_{s \in \mathbb{R}}$ as in Definition 2.1, in which the space G is embedded as $G = G_0$, and the approximation numbers $\delta_d(n, s)$ will often be the same as the approximation rates $\delta(n, s)$ in Assumption 2.3. However, the approximation (8.4) is typically not true for every smoothness level $s > 0$, but only for s in a range (s_d, S_d). For instance, for Sobolev scales the lower bound s_d is typically equal to $d/2$ for d the dimension of the domain $\mathcal{D}$ of the functions in G, and the upper bound S_d is the regularity of the basis elements used to define the scale.

Condition (8.3) implies that the set $\widetilde{W}_n$ is also n-dimensional over the design points, so that the interpolation $\mathcal{I}_n g$ indeed exists and is unique. In Lemma 8.3 it will be seen to be also the orthogonal projection of $g \in G$ onto $\widetilde{W}_n \subset G$ relative to the *discrete* inner product $\langle \cdot, \cdot \rangle_n$.

Fix an arbitrary orthonormal basis $e_{1,n}, \ldots, e_{n,n}$ of $\widetilde{W}_n$ relative to the discrete inner product $\langle \cdot, \cdot \rangle_n$, and given the discrete data $(Y_1, \ldots, Y_n)$ as in (8.2), define

$$Y^{(n)} = \sum_{i=1}^{n} \frac{1}{n} \sum_{j=1}^{n} Y_j e_{i,n}(x_j)\, e_{i,n}. \tag{8.5}$$

This embeds the discrete data as a 'continuous signal' into the space $\widetilde{W}_n \subset G$. If the observations satisfy (8.2), then the continuous observation $Y^{(n)}$ can be decomposed as

$$Y^{(n)} = \sum_{i=1}^{n} \langle Af, e_{i,n} \rangle_n e_{i,n} + \frac{1}{n} \sum_{j=1}^{n} Z_j \sum_{i=1}^{n} e_{i,n}(x_j)\, e_{i,n} = \mathcal{I}_n Af + \frac{1}{\sqrt{n}} \xi^{(n)},$$

where $\xi^{(n)}$ is a Gaussian random variable with values in the space in $\widetilde{W}_n$. Loosely speaking, as $n \to \infty$ the operators $\mathcal{I}_n$ should tend to the identity operator, and the mean of the signal $Y^{(n)}$ should become more representative of the full signal Af. As shown in the following lemma the noise $\xi^{(n)}$ remains bounded as $n \to \infty$.

Lemma 8.2. *The variable $\xi^{(n)}$ defined in the preceding display is a Gaussian random element in $\widetilde{W}_n \subset G$ with mean zero. Under (8.3) its covariance operator is up to multiplicative constants that do not depend on n bounded below and above by the orthogonal projection $\tilde{Q}_n : G \to \widetilde{W}_n$ relative to the continuous inner product $\langle \cdot, \cdot \rangle$.*

Proof. For $g \in G$ we can write $\langle \xi^{(n)}, g \rangle = n^{-1/2} \sum_{j=1}^{n} Z_j \sum_{i=1}^{n} e_{i,n}(x_j) \langle e_{i,n}, g \rangle$. Clearly the expectation of this variable vanishes, while the variance is given by

$$\begin{aligned}
\mathbb{V}\mathrm{ar} \langle \xi^{(n)}, g \rangle &= \frac{1}{n} \sum_{j=1}^{n} \Bigl(\sum_{i=1}^{n} e_{i,n}(x_j) \langle e_{i,n}, g \rangle \Bigr)^2 \\
&= \frac{1}{n} \sum_{k=1}^{n} \sum_{l=1}^{n} \sum_{j=1}^{n} e_{k,n}(x_j) \langle e_{k,n}, g \rangle e_{l,n}(x_j) \langle e_{l,n}, g \rangle \\
&= \sum_{k=1}^{n} \sum_{l=1}^{n} \langle e_{k,n}, e_{l,n} \rangle_n \langle e_{k,n}, g \rangle \langle e_{l,n}, g \rangle = \Bigl\| \sum_{k=1}^{n} \langle e_{k,n}, g \rangle e_{k,n} \Bigr\|_n^2.
\end{aligned}$$

The right side is the square of the norm $\|g_n\|$ of the vector $g_n = (g_{1,n}, \ldots, g_{n,n})$ of continuous coefficients $g_{j,n} = \langle g, e_{k,n} \rangle$ of g relative to the discrete basis. The orthogonal projection $\tilde{Q}_n g$ of g onto $\widetilde{W}_n$ relative to the continuous inner product can be written in terms of the discrete basis $e_{1,n}, \ldots, e_{n,n}$ as $\tilde{Q}_n g = \sum_{i=1}^{n} \alpha_i e_{i,n}$, for $\alpha = (\alpha_1, \ldots, \alpha_n)^T = \Sigma_n^{-1} g$ and Σ_n the Gram matrix $(\langle e_{i,n}, e_{j,n} \rangle)$. Hence $\|\tilde{Q}_n g\|^2 = \alpha^T \Sigma_n \alpha = g_n^T \Sigma_n^{-1} g_n$. Because Σ_n is bounded above and below by a multiple of the identity, by Lemma 8.3 below, it follows that $\|\tilde{Q}_n g\| \simeq \|g_n\|$. □

Lemma 8.3. *Suppose that (8.3) holds. Then* *(i) For every* $g \in G$ *the function* $\mathcal{I}_n g$ *is the orthogonal projection of* g *onto* $\widetilde{W}_n \subset G$ *relative to the* discrete *inner product* $\langle\cdot,\cdot\rangle_n$.

(ii) $C_1\|\mathcal{I}_n g\| \le \|g\|_n \le C_2\|\mathcal{I}_n g\|$, *for every* $g \in G$.

(iii) The Gram matrix $(\langle e_{i,n}, e_{j,n}\rangle)_{i,j=1..n}$ *of any basis* $e_{1,n},\ldots,e_{n,n}$ *of* $\widetilde{W}_n$ *that is orthonormal relative to the discrete inner product* $\langle\cdot,\cdot\rangle_n$ *is bounded below and above by the identity, up to multiplicative constants that do not depend on* n.

Proof. For an arbitrary orthonormal basis $e_{1,n},\ldots,e_{n,n}$ of $\widetilde{W}_n$ relative to the discrete inner product $\langle\cdot,\cdot\rangle_n$, the matrix $(e_{j,n}(x_i))_{i,j=1..n}$ has orthogonal columns of Euclidean length $\sqrt{n}$ and hence can be represented as $\sqrt{n}\,O$ for an orthogonal matrix O. The interpolation in $\widetilde{W}_n$ of a function g at the design points is the function $\sum_{j=1}^n \alpha_j e_{j,n}$ with the coefficients $\alpha = (\alpha_1,\ldots,\alpha_n)^T$ satisfying $\sqrt{n}O\alpha = (g(x_i))_{i=1..n}$. The unique solution to the latter equation is $\alpha = (1/\sqrt{n})O^T(g(x_i))_{i=1..n}$, which can be seen to be equal to $(\langle g, e_{i,n}\rangle_n)_{i=1..n}$. Thus the interpolation is indeed the orthogonal projection $\mathcal{I}_n g = \sum_{i=1}^n \langle g, e_{i,n}\rangle_n e_{i,n}$.

As the functions g and $\mathcal{I}_n g$ coincide at the design points, clearly $\|g\|_n = \|\mathcal{I}_n g\|_n$, for any $g \in G$, and this is equivalent to the continuous norm $\|\mathcal{I}_n g\|_n$, by (8.3).

To prove the third statement note that the square discrete and continuous norms of $\sum_{i=1}^n \alpha_i e_{i,n}$ are given by $\alpha^T\alpha$ and $\alpha^T\Sigma\alpha$, respectively, for Σ the Gram matrix of the basis functions $e_{i,n}$ relative to the continuous inner product. By (8.3) these norms are proportional and hence the eigenvalues of Σ are bounded from below and above. ☐

The following two examples exhibit suitable discretization spaces, both with equidistant design points.

Example 8.4 (Trigonometric Polynomials)**.** This example is adapted from Section 2.3 in [82]. Let $\mathcal{D} = \mathfrak{T}^d = (0,1]^d$, for $d \in \mathbb{N}$, and consider the set of $n = m^d$ design points $\mathcal{D}_n = \{k/m\}_{k\in\{1,\cdots,m\}^d}$, for a given odd natural number m. In this case, the Fourier system with $\mathsf{i} = \sqrt{-1}$,

$$e_k(x) = e^{\mathsf{i}2\pi\langle k,x\rangle_{\mathbb{R}^d}}, \qquad k = (k_1,\cdots,k_d) \in \mathbb{Z}^d,$$

is not only orthonormal in the continuous space $L^2(\mathfrak{T}^d)$, but also with respect to the discrete inner product $\langle\cdot,\cdot\rangle_n$, i.e.

$$\langle e_j, e_k\rangle_n = \begin{cases} 1, & \text{if } j_l \equiv k_l \mod m, \forall l \in \{1,\cdots,d\}, \\ 0, & \text{otherwise.} \end{cases} \tag{8.6}$$

The scale of isotropic Sobolev spaces $H_s(\mathfrak{T}^d)$ is defined in terms of the Fourier coefficients $f_k = \int_{\mathfrak{T}^d} f(x)e_k(x)\,dx$ of functions $f \in L^2(\mathfrak{T}^d)$, as (for $|k|$ any norm on $\mathbb{R}^d$)

$$H_s(\mathfrak{T}^d) := \Big\{ f \in L^2(\mathfrak{T}^d) :\ \|f\|_{H_s}^2 := \sum_{k\in\mathbb{Z}^d} (1+|k|)^{2s}|f_k|^2 < \infty \Big\}.$$

For smoothness levels $s \in \mathbb{N}$, this norm is equivalent to the canonical Sobolev norm $\sum_{|l|_1 \leq s} \|D^l f\|_{L^2(\mathfrak{T}^d)}$.

The spaces V_j obtained as the linear span of the basis elements, ordered suitably, satisfy Assumption 2.3. Due to (8.6), the space $\widetilde{W}_n = \mathsf{Span}\{e_k : |k|_\infty \leq (m-1)/2\}$ satisfies (8.3) with $C_1 = C_2 = 1$.

As noted in [82], the following estimates hold for $f \in H_s(\mathfrak{T}^D)$,

$$\begin{aligned} &\|f - Q_n f\|_{L^2} \lesssim n^{-s/d} \|f\|_{H_s}, \\ &\|Q_n f - \mathcal{I}_n f\| \lesssim_d n^{-s/d} \|f\|_{H_s}. \qquad \text{if } s > d/2. \end{aligned}$$

Here Q_n is the orthogonal projection on $\widetilde{W}_n$. Consequently (8.4) is fulfilled for $s > s_d = d/2$.

Example 8.5 (Wavelets)**.** This example is adapted from Section 3.3 in [82]. Let $\mathcal{D} = \mathfrak{T}^d = (0,1]^d$, for $d \in \mathbb{N}$, and consider the design points $\mathcal{D}_n = \{k2^{-j}\}_{k \in \{1,\cdots,2^j\}^d}$, where $n = 2^{jd}$ for some $j \in \mathbb{N}$. We consider a multiresolution analysis $\{V_j\}_{j\geq 0}$ on $L^2(\mathfrak{T}^d)$ obtained by periodization and tensor products. Let $\widetilde{\phi}$ be a standard orthonormal scaling function of an S-regular multiresolution analysis for $L^2(\mathbb{R})$, with compact support in $[S-1, S]$. In particular, the polynomial exactness condition is satisfied: $\sum_{k\in\mathbb{Z}} k^q \widetilde{\phi}(x-k) - x^q$ is a polynomial of maximal degree $q-1$ for $q \in [0, S-1]$. As shown in [82], the functions

$$e_{j,k}(x_1, \ldots, x_d) = \sum_{m \in \mathbb{Z}^d} 2^{jd/2} \prod_{i=1}^{d} \widetilde{\phi}(2^j x_i - k_i + 2^j m_i),$$

are well defined and form an orthonormal basis in $L^2(\mathfrak{T}^d)$. Furthermore, for $\widetilde{W}_n := V_j = \mathsf{Span}\{e_{j,k} \mid k \in \{1, \cdots, 2^j\}^d\}$ with $n = 2^{jd} \geq 2S - 1$, conditions (8.3) is satisfied with constants C_1, C_2 that depend only on $\widetilde{\phi}$. Moreover, for the functions $e_{j,k}$ belong to the Besov space $B^s_{2,2}(\mathbb{T})$, for $s < S$, and, for every f in this Besov space and $d/2 < s < S$,

$$\|f - \mathcal{I}_n f\|_{L^2} \lesssim n^{-s/d} \|f\|_{B^s_{2,2}}.$$

Thus (8.4) is satisfied, with the smoothness scale $(H_s)_{s\in\mathbb{R}}$ taken equal to the canonical Sobolev spaces (i.e. Besov spaces $B^s_{2,2}$) on $\mathfrak{T}^d$.

Other examples of suitable discretization spaces are provided by orthogonal polynomials, for instance the systems of Legendre, Chebyshev, or Jacobi polynomials, etc., for suitably chosen design points. First, $(H_s(\mathfrak{T}))_{s\in\mathbb{R}}$ being canonical Sobolev spaces on $\mathfrak{T} = (0,1]$ satisfies Assumption 2.3 This is due to the standard Sturm-Liouville theory (see 5.2 in [12]): the polynomials form infinitely differentiable orthogonal bases in $L^2(\mathfrak{T})$. Second Assumption 8.1 is satisfied with Gaussian quadrature points as design points (Section 5.3 in [12]). These results can be extended to the multivariate domains by using tensor products. See Chapter 5 in [12] for more information.

8.2 General Contraction Rates

In this section we present a general theorem on posterior contraction. We form the posterior distribution $\Pi_n(\cdot \mid Y^n)$ as in (4.1), given a prior Π on the space $H = H_0$ and an observation $Y^n = (Y_1, \ldots, Y_n)$, whose conditional distribution given f is determined by the model (8.2). We study this random distribution under the assumption that Y^n follows the model (8.2) for a given 'true' function $f = f_0$, which we assume to be an element of H_β in a given smoothness scale $(H_s)_{s\in\mathbb{R}}$, as in Definition 2.1.

The theorem is stated in terms of the Galerkin solution to the continuous inverse problem, which is defined as follows. (See e.g., [57] for a general introduction to the Galerkin method and Section 5.3 for a self-contained derivation of the necessary inequalities, exactly in our framework.) Let $W_j = AV_j \subset G$ be the image under the operator $\mathcal{A}$ of a finite-dimensional approximation space V_j linked to the smoothness scale $(H_s)_{s\in\mathbb{R}}$ as in Assumption 2.3, and let $Q_j : G \to W_j$ be the orthogonal projection onto W_j. If $A : H \to G$ is injective, then A is a bijection between the finite-dimensional vector spaces V_j and W_j, and hence for every $f \in H$ there exists $f^{(j)} \in V_j$ such that $Af^{(j)} = Q_j Af$. The element $f^{(j)}$ is called the *Galerkin solution* to Af in V_j, and is an approximation to f that is more accurate, but also more complex, for larger j.

In our current setting we have no access to the continuous function Af, but must reconstruct f from the discrete approximation to $\mathcal{A}_n f$, for $\mathcal{A}_n = \mathcal{I}_n A$, and $\mathcal{I}_n$ the interpolation operator defined in Section 8.1. Thus we shall use the Galerkin solution $f^{(j,n)} = A^{-1} Q_j \mathcal{A}_n f$ to the interpolation $\mathcal{A}_n f$ of the discrete signal. This *discrete Galerkin solution* is illustrated in the following diagram

$$\begin{array}{ccc} H \ni f & \xrightarrow{\mathcal{A}_n = \mathcal{I}_n A} & \mathcal{A}_n f \in \widetilde{W}_n \subset G \\ & & \Big\downarrow{\scriptstyle Q_j} \\ H \supset V_j \ni f^{(j,n)} & \xleftarrow{A^{-1}} & Q_j \mathcal{A}_n f \in W_j \subset G \end{array}$$

In this scheme the space $\widetilde{W}_n$, used to construct the continuous interpolation, may or may not be equal to $W_n = AV_n$. Setting it equal to W_n simplifies the scheme, but then the interpolation properties in Assumption 8.1 must be verified for AV_n.

Theorem 8.6. *For smoothness classes $(H_s)_{s\in\mathbb{R}}$ as in Definition 2.1, assume that the opeartor $\mathcal{A} : H_0 \to G$ satisfies $\|\mathcal{A}f\| \simeq \|f\|_{-\gamma}$, for some $\gamma > 0$. Let $f^{(j,n)}$ denote the discrete Galerkin solution to $\mathcal{A}_n f = \mathcal{I}_n Af$ relative to linear subspaces V_j associated to $(H_s)_{s\in\mathbb{R}}$ as in Assumption 2.3 and interpolation spaces $\widetilde{W}_n$ satisfying (8.3)-(8.4) from Assumption 8.1. Let $f_0 \in H_\beta$ and $Af_0 \in G_{\beta+\gamma}$ for some $\beta \in (s_d - \gamma, S_d - \gamma)$, and for $\eta_n \geq \varepsilon_n \downarrow 0$ such that $n\varepsilon_n^2 \to \infty$, and $j_n \in \mathbb{N}$ such that $j_n \to \infty$, and some*

$c > 0$, assume

$$j_n \leq cn\varepsilon_n^2, \tag{8.7}$$

$$\eta_n \geq \frac{\varepsilon_n}{\delta(j_n, \gamma)}, \tag{8.8}$$

$$\eta_n \geq \delta(j_n, \beta) \vee \frac{\delta_d(n, \beta + \gamma)}{\delta(j_n, \gamma)}. \tag{8.9}$$

Consider prior probability distributions Π on H_0 satisfying

$$\Pi(f \in H : \|\mathcal{A}f - \mathcal{A}f_0\|_n < \varepsilon_n) \gtrsim e^{-n\varepsilon_n^2}, \tag{8.10}$$

$$\Pi(f \in H : \|f^{(j_n,n)} - f\| > \eta_n) \lesssim e^{-4n\varepsilon_n^2}. \tag{8.11}$$

Then the posterior distribution in the model (8.2) contracts at the rate η_n at f_0, i.e. for a sufficiently large constant M we have $\Pi_n(f : \|f - f_0\| > M\eta_n \mid Y_1, \dots, Y_n) \to 0$, in probability if $Y_1, \dots, Y_n$ follow (8.2) with $f = f_0$.

Proof. The Kullback-Leibler divergence and variation between the (multivariate-normal) distributions of $(Y_1, \dots, Y_n)$ under two functions f and f_0 are given by $n\|\mathcal{A}f - \mathcal{A}f_0\|_n^2/2$ and twice this quantity, respectively. Therefore the neighbourhoods $B_{n,2}(f_0, \varepsilon_n)$ in (8.19) of [35] contain the balls $\{f \in H : \|\mathcal{A}f - \mathcal{A}f_0\|_n \leq \varepsilon_n\}$. By assumption (8.10) this has prior mass at least $\exp(-n\varepsilon_n^2)$.

Because the quotient of the left sides of (8.11) and (8.10) is $o(\exp(-2n\varepsilon_n^2))$, the posterior probability of the set $\{f : \|f^{(j_n,n)} - f\| > \eta_n\}$ tends to zero, by Theorem 8.20 in [35].

By a variation of Theorem 8.22 in [35] it is now sufficient to show the existence of tests τ_n such that, for some $M > 0$,

$$P_{f_0}^{(n)}\tau_n \to 0, \qquad \sup_{\substack{\|f - f_0\| > M\eta_n, \\ \|f^{(j_n,n)} - f\| \leq \eta_n}} P_f^{(n)}(1 - \tau_n) \leq e^{-4n\varepsilon_n^2}.$$

Define the operator $\mathcal{R}_j : G \mapsto V_j$ by $\mathcal{R}_j = \mathcal{A}^{-1}Q_j$, where $Q_j : G \to W_j$ is the orthogonal projection onto $W_j = \mathcal{A}V_j$ and $\mathcal{A}^{-1}$ is the inverse of $\mathcal{A}$ (restricted to W_j). Then we shall employ the tests

$$\tau_n = 1\{\|\mathcal{R}_{j_n}Y^{(n)} - f_0\| \geq M_0\eta_n\}, \tag{8.12}$$

where M_0 is a given constant, to be determined.

By definition $f^{(j,n)} = \mathcal{R}_j\mathcal{A}_nf$ is equal to the discrete Galerkin solution to $\mathcal{A}_nf$. For $Y^{(n)}$ defined in (8.5) we have

$$\mathcal{R}_jY^{(n)} = \mathcal{R}_j\mathcal{A}_nf + \frac{1}{\sqrt{n}}\mathcal{R}_j\xi^{(n)} = f^{(j,n)} + \frac{1}{\sqrt{n}}\mathcal{R}_j\xi^{(n)}. \tag{8.13}$$

The variable $\mathcal{R}_j\xi^{(n)} = \mathcal{R}_jQ_j\xi^{(n)}$ is a centered Gaussian random element in V_j with strong and weak second moments

$$\mathbb{E}\|\mathcal{R}_j\xi^{(n)}\|^2 \leq \|\mathcal{R}_j\|^2\mathbb{E}\|Q_j\xi^{(n)}\|^2 \lesssim \|\mathcal{R}_j\|^2 j \lesssim \frac{j}{\delta(j, \gamma)^2},$$

$$\sup_{\|f\| \leq 1} \mathbb{E}\langle\mathcal{R}_j\xi^{(n)}, f\rangle^2 = \sup_{\|f\| \leq 1} \mathbb{E}\langle\xi^{(n)}, \mathcal{R}_j^*f\rangle_G^2 \lesssim \sup_{\|f\| \leq 1} \|\mathcal{R}_j^*f\|_G^2 \leq \|\mathcal{R}_j^*\|^2 \lesssim \frac{1}{\delta(j, \gamma)^2}.$$

In both cases the inequality on $\|\mathcal{R}_j\|=\|\mathcal{R}_j^*\|$ at the far right side follows from (5.7), and we also use that, by Lemma 8.2, the covariance operator of $\xi^{(n)}$ is bounded above by a multiple of the projection onto $\widetilde{W}_n$, and hence the identity, so that the covariance operator of $Q_j\xi^{(n)}$ is bounded above by a multiple of Q_j.

The first inequality shows that the first moment $\mathbb{E}\|\mathcal{R}_j\xi^{(n)}\|$ of the variable $\|\mathcal{R}_j\xi^{(n)}\|$ is bounded above by $\sqrt{j}/\delta(j,\gamma)$. By Borell's inequality (e.g. Lemma 3.1 in [67] and subsequent discussion), applied to the Gaussian random variable $\mathcal{R}_j\xi^{(n)}$ in H_0, we see that, there exist positive constants a and b such that, for every $t>0$,

$$\Pr\Bigl(\|\mathcal{R}_j\xi^{(n)}\|> t+a\frac{\sqrt{j}}{\delta(j,\gamma)}\Bigr)\le e^{-bt^2\delta(j,\gamma)^2}.$$

For $t=2\sqrt{n}\eta_n/\sqrt{b}$ and η_n, ε_n and j_n satisfying (8.7), (8.8) and (8.9) this yields, for some $a_1>0$,

$$\Pr\Bigl(\|\mathcal{R}_{j_n}\xi^{(n)}\|> a_1\sqrt{n}\eta_n\Bigr)\le e^{-4n\varepsilon_n^2}. \tag{8.14}$$

We apply this to bound the two error probabilities of the tests τ_n.

Under f_0 the decomposition (8.13) is valid with $f=f_0$, and hence $\mathcal{R}_jY^{(n)}-f_0=n^{-1/2}\mathcal{R}_j\xi^{(n)}+f_0^{(j,n)}-f_0$. By the triangle inequality it follows that $\tau_n=1$ implies that $n^{-1/2}\|\mathcal{R}_{j_n}\xi^{(n)}\|\ge M_0\eta_n-\|f_0^{(j,n)}-f_0\|$. By the triangle inequality followed by (5.7) and (8.4), and (5.9),

$$\begin{aligned}\|f_0^{(j,n)}-f_0\|&\le\|\mathcal{R}_j\|\,\|\mathcal{I}_n\mathcal{A}f_0-\mathcal{A}f_0\|+\|\mathcal{R}_j\mathcal{A}f_0-f_0\|\\&\lesssim\frac{\delta_d(n,\beta+\gamma)}{\delta(j_n,\gamma)}\|\mathcal{A}f_0\|_{\beta+\gamma}+\delta(j_n,\beta)\|f_0\|_\beta\le M_1\eta_n,\end{aligned}$$

by assumption (8.9). Hence the probability of an error of the first kind satisfies

$$P_{f_0}^{(n)}\tau_n\le\Pr\Bigl(\frac{1}{\sqrt{n}}\|\mathcal{R}_{j_n}\xi^{(n)}\|\ge(M_0-M_1)\eta_n\Bigr),$$

For $M_0-M_1>a_1$, the right side is bounded by $e^{-4n\varepsilon_n^2}$, by (8.14).

Under f the decomposition (8.13) gives that $\mathcal{R}_jY^{(n)}-f_0=n^{-1/2}\mathcal{R}_j\xi^{(n)}+f^{(j,n)}-f_0$. By the triangle inequality $\tau_n=0$ implies that $n^{-1/2}\|\mathcal{R}_{j_n}\xi^{(n)}\|\ge\|f^{(j_n,n)}-f_0\|-M_0\eta_n$. For f such that $\|f-f_0\|>M\eta_n$ and $\|f-f^{(j_n,n)}\|\le\eta_n$, we have $\|f^{(j_n,n)}-f_0\|\ge(M-1)\eta_n$. Hence the probability of an error of the second kind satisfies

$$P_f^{(n)}(1-\tau_n)\le\Pr\Bigl(\frac{1}{\sqrt{n}}\|\mathcal{R}_{j_n}\xi^{(n)}\|\ge(M-1-M_0)\eta_n\Bigr),$$

For $M-1-M_0>a_1$, this is bounded by $e^{-4n\varepsilon_n^2}$, by (8.14).

We can first choose M_0 large enough so that $M_0-M_1>a_1$, and next M large enough so that $M-1-M_0>a_1$, to finish the proof. □

The theorem has a similar form as Theorem 6.1 obtained in Chapter 6 in the case of observation of a continuous signal (in white noise). Some interpretations of the theorem from the previous chapter are also applicable to the current one.

For completeness, we repeat them here. Inequality (8.10) is the usual *prior mass condition* for the 'direct problem' of estimating $\mathcal{A}f$ at the design points (see [33]). It determines the rate of contraction ε_n of the posterior distribution of $\mathcal{A}f$ to $\mathcal{A}f_0$ relative to the discrete seminorm $\|\cdot\|_n$. The rate of contraction η_n of the posterior distribution of f is slower due to the necessity of (implicitly) inverting the operator $\mathcal{A}$. The theorem shows that the rate η_n depends on the combination of the prior, through (8.11), and the inverse problem, through the various approximation rates. The factor $\delta_d(n, \beta+\gamma)/\delta(j_n, \gamma)$ in (8.9), which arises from having discrete observations only, will typically be negligible relative to $\delta(j_n, \beta)$. The Galerkin projection $f^{(j,n)}$ in (8.11) now incorporates the errors of both inversion and discretisation.

Same adaptation by mixture priors (c.f. Theorem 6.11) can also be achieved. The following theorem refines Theorem 8.6 by considering a mixture prior of the form

$$\Pi = \int \Pi_\tau \, dQ(\tau), \tag{8.15}$$

where Π_τ is a prior on H, for every given 'hyperparameter' τ running through some measurable space, and Q is a prior on this hyperparameter. The idea is to *adapt* the prior to multiple smoothness levels through the hyperparameter τ.

Theorem 8.7. *Consider the setup and assumptions of Theorem 8.6 with a prior of the form (8.15). Assume that (8.7), (8.8), (8.9) and (8.10) hold, but replace (8.11) by the pair of conditions, for numbers $\eta_{n,\tau}$ and $C > 0$ and every τ,*

$$\Pi_\tau(f : \|f - f_0\| < 2\eta_{n,\tau}) \le e^{-4n\varepsilon_n^2}, \quad \forall \tau \text{ with } \eta_{n,\tau} \ge C\eta_n, \tag{8.16}$$

$$\Pi_\tau(f : \|f^{(j_n,n)} - f\| > \eta_{n,\tau}) \le e^{-4n\varepsilon_n^2}. \tag{8.17}$$

Then the posterior distribution in the model (8.2) contracts at the rate η_n at f_0, i.e. for a sufficiently large constant M we have $\Pi_n(f : \|f - f_0\| > M\eta_n \mid Y_1, \dots, Y_n) \to 0$, in probability if $Y_1, \dots, Y_n$ follow (8.2) with $f = f_0$.

Proof. We take the parameter of the model as the pair (f, τ), which receives the joint prior given by $f \mid \tau \sim \Pi_\tau$ and $\tau \sim Q$. With abuse of notation, we denote this prior also by Π. The likelihood still depends on f only, but the joint prior gives rise to a posterior distribution on the pair (f, τ), which we also denote by $\Pi_n(\cdot \mid Y^n)$, by a similar abuse of notation.

By (8.15) and eqs. (8.16) and (8.17),

$$\Pi((f,\tau) : \eta_{n,\tau} \ge C\eta_n, \|f - f_0\| < 2\eta_{n,\tau}) \le e^{-4n\varepsilon_n^2},$$
$$\Pi((f,\tau) : \|f^{(j_n,n)} - f\| > \eta_{n,\tau}) \le e^{-4n\varepsilon_n^2}.$$

In view of (8.10) and Theorem 8.20 in [35], the posterior probabilities of the two sets in the left sides tend to zero. As in the proof of Theorem 8.6, we can apply a variation of Theorem 8.22 in [35] to see that it is now sufficient to show the existence of tests τ_n such that, for some $M \ge 2C$,

$$P_{f_0}^{(n)} \tau_n \to 0, \qquad \sup_{\substack{(f,\tau): \|f - f_0\| > M\eta_n \vee 2\eta_{n,\tau}, \\ \|f^{(j_n,n)} - f\| \le \eta_{n,\tau}}} P_f^{(n)}(1 - \tau_n) \le e^{-4n\varepsilon_n^2}.$$

(Note that $M\eta_n \vee 2\eta_{n,\tau} = M\eta_n$ if $\eta_{n,\tau} < C\eta_n$ and $M \geq 2C$.) We use the tests defined in (8.12), as in the proof of Theorem 8.6. The latter proof shows that the tests are consistent. The bound on the power can be adapted same as in the proof of Theorem 6.11, and hence the detail is omitted here.

□

In a typical application of the preceding theorem the priors Π_τ for τ such that $\eta_{n,\tau} \geq C\eta_n$ will be the priors on rough functions, with 'intrinsic' contraction rate $\eta_{n,\tau}$ slower than η_n. These 'bad' priors do not destroy the overall contraction rate, because they put little mass near the true function f_0, by condition (8.16). It is necessary to address these priors explicitly in the conditions, because they will typically fail the approximation condition (8.11), which must be relaxed to (8.17).

In Theorem 8.6 and Theorem 8.7, it is sufficient for the operator $\mathcal{A}$ to satisfy the following two properties: lifting property $\|\mathcal{A}f\| \simeq \|f\|_{-\gamma}$ with some $\gamma > 0$ and $\mathcal{A}f_0 \in G_{\beta+\gamma}$ if $f_0 \in H_\beta$. However, when analysing particular priors, the aforementioned conditions on $\mathcal{A}$ may not be strong enough for verifying the conditions (8.10) and (8.11), while they can be verified with the following condition, which is stronger but often satisfied in practice, e.g. Example 5.4. More examples can be found in [57].

Assumption 8.8 (Smoothing property of $\mathcal{A}$)**.** For some $\gamma > 0$ the operator $\mathcal{A}: H_{s-\gamma} \to G_s$ is injective and bounded for every $s \geq 0$, and, for every $f \in H_0 \cap H_{s-\gamma}$,

$$\|\mathcal{A}f\|_s \simeq \|f\|_{s-\gamma}. \tag{8.18}$$

8.3 Random Series Priors

In this section we study the performance of the random series prior from Section 6.3 to the inverse regression model (8.2). We briefly recall the set-up of the prior and the further details are referred to Section 6.3.

Suppose that $\{\phi_i\}_{i\in\mathbb{N}}$ is an orthonormal basis of $H = H_0$ that gives optimal approximation relative to the scale of smoothness classes $(H_s)_{s\in\mathbb{R}}$ in the sense that the linear spaces $V_j = \mathsf{Span}\{\phi_i\}_{i<j}$ satisfy Assumption 2.3. Consider a prior defined as the law of the random series

$$f = \sum_{i=1}^{M} f_i \phi_i, \tag{8.19}$$

where M is a random variable in $\mathbb{N}$ independent from the independent random variables $f_1, f_2, \ldots$ in $\mathbb{R}$.

Condition 8.9 (Random series prior)**.** (i) The probability density function p_M of M satisfies, for some positive constants b_1, b_2,

$$e^{-b_1 k} \lesssim p_M(k) \lesssim e^{-b_2 k}, \qquad \forall k \in \mathbb{N}.$$

(ii) The variable f_i has density $p(\cdot/\kappa_i)/\kappa_i$, for a given probability density p on $\mathbb{R}$ and a positive constant κ_i such that, for some $C > 0$ and $0 < v < w < \infty$,

$\beta_0 > 0$ and $\alpha > 0$,

$$e^{-C|x|^w} \lesssim p(x) \lesssim e^{-C|x|^v}, \tag{8.20}$$

$$i^{-\beta_0/d} \lesssim \kappa_i \lesssim i^{\alpha}. \tag{8.21}$$

The same contraction rate is obtained for the random series prior in the inverse regression model. The discussion on the result is referred to the discussion of Theorem 6.5.

Theorem 8.10 (Random Series Prior). *Let $(\phi_i)_{i\in\mathbb{N}}$ be an orthonormal basis of H_0 such that the spaces $V_j = \mathsf{Span}\{\phi_i\}_{i<j}$ satisfy Assumption 2.3 with $\delta(j,t) = j^{-t/d}$ relative to smoothness classes $(H_s)_{s\in\mathbb{R}}$ as in Definition 2.1 and sufficiently large t to be specified. Suppose that Assumption 8.1 holds with $\delta_d(n,s) = n^{-s/d}$, the operator $\mathcal{A}$ satisfies Assumption 8.8, and let $f_0 \in H_\beta$ for some $\beta \in (0,S]$ and $\beta+\gamma > s_d$. Then, for the random series prior defined in (8.19) and satisfying Condition 8.9 with $\beta_0 \le \beta$, and sufficiently large $M > 0$, for $\tau = (\beta+\gamma)(1+2\gamma/d)/(2\beta+2\gamma+d)$,*

$$\Pi_n\Big(f : \|f - f_0\|_0 > Mn^{-\beta/(2\beta+2\gamma+d)}(\log n)^{\tau} \mid Y^{(n)}\Big) \overset{P_{f_0}^{(n)}}{\to} 0.$$

8.4 Gaussian Priors

In this section we study the posterior contractions of Gaussian priors in the inverse regression model (8.2). Recall the definition of a Gaussian prior from Section 6.4. Centred Gaussian distributions on a separable Hilbert space correspond bijectively to covariance operators. By definition a random variable F with values in H_0 is Gaussian if $\langle F, g\rangle_0$ is normally distributed, for every $g \in H_0$, and it has zero mean if these variables have zero means. The variances of these variables can then be written as

$$\mathbb{E}\langle F, g\rangle_0^2 = \langle Cg, g\rangle_0,$$

for a linear operator $C : H_0 \to H_0$, called the *covariance operator*. A covariance operator C is necessarily self-adjoint, nonnegative, and of *trace class*, i.e., $\sum_{i\in\mathbb{N}}\langle C\phi_i, \phi_i\rangle < \infty$, for some (and then every) orthonormal basis $(\phi_i)_{i\in\mathbb{N}}$ of H_0; and every operator with these properties generates a Gaussian distribution.

In the setting of a Hilbert scale $(H_s)_{s\in\mathbb{R}}$ generated by the operator L it is natural to choose a Gaussian prior with covariance operator of the form $L^{-2\alpha}$, for some $\alpha > 0$. If L^{-1} has eigenvalues λ_j, then this operator is of trace class if $\sum_{j\in\mathbb{N}}\lambda_j^{-2\alpha} < \infty$. Thus α must be chosen big enough for the Gaussian prior to exist as a 'proper' prior on H_0. For instance, if $\lambda_j \simeq j^{-1/d}$, then every choice $\alpha > d/2$ yields a proper prior.

This leads to the following theorem on posterior contraction rates for Gaussian priors, the proof of which is given in Section 8.6.

Theorem 8.11 (Gaussian Prior). *Consider a Hilbert scale $(H_s)_{s\in\mathbb{R}}$ generated by an operator L as in the preceding such that $L^{-1} : H_0 \to H_0$ is compact with eigenvalues λ_j satisfying $\lambda_j \simeq j^{-1/d}$. Suppose that Assumption 8.1 holds with*

$\delta_d(n,s) = n^{-s/d}$, the operator $\mathcal{A} : H_0 \to G$ satisfies Assumption 8.8, $f_0 \in H_\beta$, for some $\beta > 0$ such that $\beta + \gamma > s_d$, and let the prior be zero-mean Gaussian with covariance operator $L^{-2\alpha}$, for some $\alpha > (d-\gamma) \vee d/2$. Then the posterior distribution satisfies, for sufficiently large $M > 0$,

$$\Pi_n\Big(f : \|f - f_0\|_0 > Mn^{-((\alpha-d/2)\wedge\beta)/(2\alpha+2\gamma)} \mid Y^{(n)}\Big) \overset{P_{f_0}^{(n)}}{\to} 0.$$

The results are comparable to Theorem 6.7, and hence we refer to Section 6.4 for the discussion.

8.5 Gaussian Mixtures

The posterior contraction rate resulting from a zero-mean Gaussian prior with covariance operator $L^{-2\alpha}$, as considered in Section 8.4, is equal to the minimax rate $n^{-\beta/(2\beta+2\gamma+d)}$ (see [19]) only when $\alpha - d/2 = \beta$, i.e., when the prior smoothness $\alpha - d/2$ matches the true smoothness β. As shown in Section 6.5, by mixing over Gaussian priors of varying smoothness the minimax rate can often be obtained simultaneously for a range of values β. In this section we consider the same mixture of the mean-zero Gaussian priors with covariance operators $\tau^2 L^{-2\alpha}$ over the 'hyperparameter' τ, from Section 6.5. Thus the prior Π is the distribution of τF, where F is a zero-mean Gaussian variable in H_0 with covariance operator $L^{-2\alpha}$, as in Section 8.4, and τ is an independent scale parameter satisfying Condition 6.10. We repeat the condition below for the reader's convenience.

Condition 8.12. The distribution Q of τ has support $[0,\infty)$ and satisfies

$$\begin{cases} -\log Q((t,2t)) \lesssim t^{-2}, & \text{as } t \downarrow 0, \\ -\log Q((t,2t)) \lesssim t^{d/(\alpha-d/2)}, & \text{as } t \to \infty. \end{cases}$$

Theorem 8.13 (Gaussian mixture prior)**.** *Consider a Hilbert scale $(H_s)_{s\in\mathbb{R}}$ generated by an operator L as in the preceding such that $L^{-1} : H_0 \to H_0$ is compact with eigenvalues λ_j satisfying $\lambda_j \simeq j^{-1/d}$. Suppose the operator $\mathcal{A} : H_s \to G_{s+\gamma}$ satisfies Assumption 8.8, assume that $f_0 \in H_\beta$, for some $(d/2-\gamma)\vee 0 < \beta \le \alpha$, and let the prior be a mixture of the zero-mean Gaussian distributions with covariance operators $\tau^2 L^{-2\alpha}$ over the parameters τ equipped with a prior satisfying Condition 8.12, for some $\alpha > (d-\gamma)\vee d/2$. Then the posterior distribution satisfies, for sufficiently large $M > 0$,*

$$\Pi_n\Big(f : \|f - f_0\|_0 > Mn^{-\beta/(2\beta+2\gamma+d)} \mid Y^{(n)}\Big) \overset{P_{f_0}^{(n)}}{\to} 0.$$

The proof is given in Section 8.6.

8.6 Proofs

8.6.1 Proof of Theorem 8.10

The theorem is a corollary to Theorem 8.6. We shall verify the conditions with

$$\varepsilon_n \simeq (\log n/n)^{(\beta+\gamma)/(2\beta+2\gamma+d)}, \quad j_n \simeq n^{d/(2\beta+2\gamma+d)}(\log n)^{(2\beta+2\gamma)/(2\beta+2\gamma+d)}.$$

Let P_j be the orthogonal projection of H on the linear span of the first $j-1$ basis elements ϕ_j, and define an additional sequence of integers by

$$i_n \simeq (n/\log n)^{d/(2\beta+2\gamma+d)}.$$

By the orthogonality of the basis (ϕ_i), the function ϕ_j is orthogonal to the space V_j spanned by $(\phi_i)_{i<j}$. Hence $P_j\phi_j = 0$, so that $\|\phi_j\|_{-s}\leq \delta(j,s)\|\phi_j\|_0 \lesssim j^{-s/d}$, for every j and $s \geq 0$, by (2.4). The same estimate is also true for $0 < -s < S$, directly by assumption (2.3). Therefore, by the triangle inequality, we have

$$\|f\|_s \lesssim \sum_j |f_j| j^{s/d}, \qquad \text{if } f = \sum_j f_j\phi_j.$$

Furthermore, since $f_0 \in H_\beta$ by assumption, the norm duality (2.1) gives that $|f_{0,i}| = |\langle f_0, \phi_i\rangle_0| \leq \|f_0\|_\beta \|\phi_i\|_{-\beta} \lesssim i^{-\beta/d}$.

First we verify the prior condition (8.10) of the direct problem. By Lemma 8.3, $\|\mathcal{A}f - \mathcal{A}f_0\|_n \simeq \|\mathcal{I}_n(\mathcal{A}f - \mathcal{A}f_0)\|$. By several applications of the triangle inequality, since $\beta+\gamma \in (s_d, S_d)$ and $\|\mathcal{A}f\|_{\beta+\gamma} \simeq \|f\|_\beta$,

$$\begin{aligned}\|\mathcal{I}_n(\mathcal{A}f - \mathcal{A}f_0)\| &\leq \|\mathcal{I}_n\mathcal{A}f - \mathcal{A}f\| + \|\mathcal{I}_n\mathcal{A}f_0 - \mathcal{A}f_0\| + \|\mathcal{A}f - \mathcal{A}f_0\| \\ &\lesssim \delta_d(n, \beta+\gamma)(\|f\|_\beta + \|f_0\|_\beta) + \|f - P_{i_n}f_0\|_{-\gamma} + \|f_0 - P_{i_n}f_0\|_{-\gamma} \\ &\lesssim \delta_d(n,\beta+\gamma)\|f - P_{i_n}f_0\|_\beta + \|f - P_{i_n}f_0\|_{-\gamma} \\ &\qquad + \delta_d(n,\beta+\gamma)(\|P_{i_n}f_0\|_\beta + \|f_0\|_\beta) + \delta(i_n,\gamma)\delta(i_n,\beta)\|f_0\|_\beta,\end{aligned}$$

by (2.4). The last term is of the order $\delta(i_n,\gamma)\delta(i_n,\beta) = i_n^{-(\gamma+\beta)/d} \simeq \varepsilon_n$, while the second last term is bounded above by $\delta_d(n,\beta+\gamma)(\|f_0\|/\delta(i_n,\beta) + \|f_0\|_\beta) \ll \varepsilon_n$, if $\beta+\gamma > d/2$.For $f = \sum_{i=1}^{i_n-1} f_i\phi_i \in V_{i_n}$ the sum of the first two terms is bounded above by

$$\delta_d(n,\beta+\gamma)\sum_{i=1}^{i_n-1}|f_i - f_{0,i}| i^{\beta/d} + \sum_{i=1}^{i_n-1}|f_i - f_{0,i}| i^{-\gamma/d} \lesssim \sum_{i=1}^{i_n-1}|f_i - f_{0,i}| i^{-\gamma/d}.$$

Same as shown in Section 6.6.1, the right side of this equation is bounded above by ε_n with prior probability at least $e^{-n\varepsilon_n^2/2}$. Since also $\Pi(M = i_n - 1) \geq e^{-b_1 i_n} \geq e^{-n\varepsilon_n^2/2}$, it follows that (8.10) is satisfied.

Next we verify (8.11). Since $\Pi(M \geq j_n) \leq e^{-b_2' j_n} \leq e^{-4n\varepsilon_n^2}$, by Condition 8.9, we may intersect the event in (8.11) with the event $M < j_n$. For $M < j$ the random series $f = \sum_{i=1}^M f_i\phi_i$ is contained in V_j and the Galerkin approximation

$\mathcal{R}_j\mathcal{A}f$ of f is exact. Since $f^{(j,n)} = \mathcal{R}_j\mathcal{I}_n\mathcal{A}f$, the triangle inequality followed by (8.4) give, for $s+\gamma \in (s_d, S_d)$,

$$\begin{aligned}\|f^{(j,n)} - f\| &\leq \|\mathcal{R}_j\mathcal{I}_n\mathcal{A}f - \mathcal{R}_j\mathcal{A}f\| + \|\mathcal{R}_j\mathcal{A}f - f\| \\ &\leq \frac{\delta_d(n, s+\gamma)}{\delta(j,\gamma)}\|f\|_s + \|\mathcal{R}_j\mathcal{A}f - f\| \\ &\leq \frac{\delta_d(n, s+\gamma)}{\delta(j,\gamma)\delta(j,s)}\|f\| + 0,\end{aligned}$$

if $f \in V_j$, by (2.3), for s such that $f \in H_s$. We conclude that it suffices to prove that

$$\Pi\Big(\sum_{i=1}^{j_n-1} f_i^2 > \eta_n^2 (j_n/n)^{2\gamma+2s}\Big) \leq e^{-4n\varepsilon_n^2}.$$

With the given choices of η_n and j_n, for some $a \in \mathbb{R}$,

$$\eta_n^2 (j_n/n)^{2\gamma+2s} = n^{(4(s+\gamma)(\beta+\gamma)-2d\beta)/(2\beta+2\gamma+d)}(\log n)^a.$$

For sufficiently large s this is an arbitary high power of n. We have that

$$\mathbb{E}\sum_{i=1}^{j_n-1} f_i^2 \simeq \sum_{i=1}^{j_n-1} \kappa_i^2 \lesssim j_n^{2\alpha+1}.$$

$$\mathbb{E}\Big|\sum_{i=1}^{j_n-1}(f_i^2 - \mathbb{E} f_i^2)\Big| \lesssim \sqrt{\sum_{i=1}^{j_n-1}\kappa_i^4} \lesssim j_n^{2\alpha+1/2}.$$

By the tail bound on the density p we further have that the $\psi_{v/2}$ Orlicz norm of f_i^2 is bounded above by κ_i^2. Therefore,

$$\Big\|\sum_{i=1}^{j_n-1}(f_i^2 - \mathbb{E} f_i^2)\Big\|_{\psi_{v/2}} \lesssim \begin{cases} j_n^{2\alpha+1/2} + (\log j_n)^{2/v}\max_{i<j_n}\kappa_i^2, & \text{if } v \leq 2, \\ j_n^{2\alpha+1/2} + (\sum_{i<j_n}\kappa_i^{2q})^{1/q}, & \text{if } 2 < v \leq 4, \end{cases}$$

where q is conjugate to $v/2$. In both cases the first term $j_n^{2\alpha+1/2}$ dominates. So provided

$$n^{(4(s+\gamma)(\beta+\gamma)-2d\beta)/(2\beta+2\gamma+d)}(\log n)^a \gtrsim j_n^{2\alpha+1}$$

we can first center $\sum_{i=1}^{j_n-1} f_i^2$ at mean zero, and next bound the tail of the centered variable with the help of the Orlicz norm. This will give a bound of the type

$$1/\psi_{v/2}\Big(\frac{n^{(4(s+\gamma)(\beta+\gamma)-2d\beta)/(2\beta+2\gamma+d)}(\log n)^a}{j_n^{2\alpha+1/2}}\Big).$$

Under the preceding display the quotient is a positive power n^t of n times a logarithmic factor and we obtain a bound of the form

$$e^{-(n^t)^{v/2}}.$$

Here t can be arbitrarily large by choosing s large. So condition (8.11) holds provided s can be chosen sufficiently large in the interpolation inequality.

8.6.2 Proof of Theorem 8.11

The theorem is a corollary to Theorem 8.6. The main tasks are to determine ε_n satisfying the prior mass condition (8.10) of the direct problem, and next to identify η_n from the prior mass condition (8.11) and the other conditions.

Similar to the proof of Theorem 8.10, the prior mass condition (8.10) is decomposed into different components, which will be studied separately. Recall that by Lemma 8.3, $\|\mathcal{A}f - \mathcal{A}f_0\|_n \simeq \|\mathcal{I}_n(\mathcal{A}f - \mathcal{A}f_0)\|$. By triangle inequality, we have

$$\|\mathcal{I}_n(\mathcal{A}f - \mathcal{A}f_0)\| \leq \|\mathcal{I}_n\mathcal{A}f - \mathcal{A}f\| + \|\mathcal{I}_n\mathcal{A}f_0 - \mathcal{A}f_0\| + \|\mathcal{A}f - \mathcal{A}f_0\|.$$

In the preceding display, the last term is same as the prior mass condition for white noise model and has been studied in Section 6.6.2. We recall the result in the following lemma.

Lemma 8.14 (Lemma 6.12). *Under the assumptions of Theorem 8.11, for $f_0 \in H_\beta$, as $\varepsilon \downarrow 0$,*

$$-\log \Pi(f : \|\mathcal{A}f - \mathcal{A}f_0\| < \varepsilon) \lesssim \begin{cases} \varepsilon^{-d/(\alpha+\gamma-d/2)}, & \text{if } d/2 < \alpha \leq \beta + d/2, \\ \varepsilon^{-(2\alpha-2\beta)/(\beta+\gamma)}, & \text{if } \alpha > \beta + d/2. \end{cases}$$

Hence, It follows that $\Pi(f : \|\mathcal{A}f - \mathcal{A}f_0\|_0 < \varepsilon_n/2) \geq e^{-n\varepsilon_n^2}$ with

$$\varepsilon_n \leq \begin{cases} 2^{d/(2\alpha+2\gamma)} n^{-(\alpha+\gamma-d/2)/(2\alpha+2\gamma)}, & \text{if } d/2 < \alpha \leq \beta + d/2, \\ 2^{(2\alpha-2\beta)/(2\alpha+2\gamma)} n^{-(\beta+\gamma)/(2\alpha+2\gamma)}, & \text{if } \alpha > \beta + d/2. \end{cases}$$

The conditions of ε_n given above can be further simplified into

$$\varepsilon_n \geq C 2^{d/(2\gamma+d)} n^{-(\beta\wedge(\alpha-d/2)+\gamma)/(2\alpha+2\gamma)}, \tag{8.22}$$

where $2^{d/(2\gamma+d)} < 2$ is independent of α and β and C will be determined below.

Since $\beta+\gamma > d/2 = s_d$, we have $\|\mathcal{I}_n\mathcal{A}f_0 - \mathcal{A}f_0\| \leq \delta_d(n, \beta+\gamma)\|f_0\|_\beta \simeq n^{-(\beta+\gamma)/d}$. Therefore, by selecting a sufficiently large constant C in the preceding display such that the right hand side of (8.22) is upper bounded by ε_n, we obtain that $\Pi(f : \|\mathcal{I}_n\mathcal{A}f_0 - \mathcal{A}f_0\| + \|\mathcal{A}f - \mathcal{A}f_0\|_0 < \varepsilon_n) \geq e^{-n\varepsilon_n^2}$.

Using a basic probability property, $P(A \cap B) = P(A) + P(B) - P(A \cup B) \geq P(B) - P(A^c)$,

$$\begin{aligned} &\Pi(f : \|\mathcal{I}_n\mathcal{A}f - \mathcal{A}f\| + \|\mathcal{I}_n\mathcal{A}f_0 - \mathcal{A}f_0\| + \|\mathcal{A}f - \mathcal{A}f_0\|_0 < \varepsilon_n) \\ \geq &\Pi(f : \|\mathcal{I}_n\mathcal{A}f_0 - \mathcal{A}f_0\| + \|\mathcal{A}f - \mathcal{A}f_0\|_0 < \varepsilon_n/2, \quad \|\mathcal{I}_n\mathcal{A}f - \mathcal{A}f\| < \varepsilon_n/2) \\ \geq &\Pi(f : \|\mathcal{I}_n\mathcal{A}f_0 - \mathcal{A}f_0\| + \|\mathcal{A}f - \mathcal{A}f_0\|_0 < \varepsilon_n/2) - \Pi(f : \|\mathcal{I}_n\mathcal{A}f - \mathcal{A}f\| \geq \varepsilon_n/2) \\ \geq &e^{-n\varepsilon_n^2} - \Pi(f : \|\mathcal{I}_n\mathcal{A}f - \mathcal{A}f\| \geq \varepsilon_n/2). \end{aligned}$$

Since the prior of f is a centred Gaussian distribution with covariance operator $L^{-2\alpha}$, f is in H_s almost surely for $s < \alpha - d/2$. Consequently, $\mathcal{A}f \in G_{s+\gamma}$ almost surely. Since α is chosen in the range $\alpha > (d-\gamma) \vee d/2$, which implies $\alpha + \gamma > d$, and consequently there exists an s satisfying $d/2 = s_d < s+\gamma < S_d$. Hence, (8.4) holds and we have

$$\|\mathcal{I}_n \mathcal{A} f - \mathcal{A} f\| \leq \delta_d(n, s+\gamma)\|f\|_s = \delta_d(n, s+\gamma)\|L^s f\|_0.$$

It leads to

$$\Pi(f : \|\mathcal{I}_n \mathcal{A} f - \mathcal{A} f\| \geq \varepsilon_n/2) \leq \Pi\left(f : \|L^s f\| \geq \frac{\varepsilon_n}{2\delta_d(n, s+\gamma)}\right).$$

Therefore, it suffices to show that, for the transformed centred Gaussian random variable $L^s F$ with covariance operator $L^{-2(\alpha-s)}$ (where F is centred Gaussian with covariance $L^{-2\alpha}$),

$$\Pi(\|F\| > r_n) < e^{-n\varepsilon_n^2},$$

where $r_n = \frac{\varepsilon_n}{2\delta_d(n,s+\gamma)}$ and ε_n is as given in (8.22).

For F, Since $\mathbb{E}\|F\|^2 = \sum_{i\in\mathbb{N}} i^{-\frac{\alpha-s}{d/2}} < \infty$, the first moment of $\mathbb{E}\|F\|$ is also bounded by Jensen's inequality. In particular, the upper bound of $\mathbb{E}\|F\|$ is independent of n. The weak second moment is given by

$$\sigma = \sup_{\|h\|_0 \leq 1} \mathbb{E}\langle F, h\rangle_0^2 = \sup_{\|h\|_0 \leq 1} \|h\|_{-(\alpha-s)}^2.$$

By the norm duality (2.1), the right side is equal to

$$\sup_{\|h\|_0\leq 1} \sup_{\|f\|_{\alpha-s}\leq 1} \langle f, h\rangle_0^2 \leq \sup_{\|f\|_{\alpha-s}\leq 1} \|f\|_0^2 \leq 1.$$

Then by Borell's inequality, when $r_n > \mathbb{E}\|F\|$,

$$\begin{aligned}\Pi(\|F\| > r_n) \leq& \Pi(\|F\| - \mathbb{E}\|F\| > r_n - \mathbb{E}\|F\|)\\ \leq& e^{-\frac{(r_n - \mathbb{E}\|f\|)^2}{2\sigma^2}} \leq e^{-r_n(r_n - \mathbb{E}\|F\|)} \ll e^{-n\varepsilon_n^2},\end{aligned}$$

where we use the fact that $n\varepsilon_n^2 \ll r_n \to \infty$ since $\delta_d(n, s) \simeq n^{-s/d}$ and $s+\gamma > d/2$.

Combining above results, we have shown that the (8.10)

$$\Pi(f \in H : \|\mathcal{A} f - \mathcal{A} f_0\|_n < \varepsilon_n) \geq \frac{1}{2} e^{-n\varepsilon_n^2}$$

is satisfied with ε_n given in (8.22).

The next step of the proof is to bound the prior probability in (8.11). The following lemma is a modification of Lemma 8.2 in [43].

Lemma 8.15. *Under the assumptions of Theorem 8.11, there exist $a, b > 0$, such that for every $j \in \mathbb{N}$ and $t > 0$,*

$$\Pi(f : \|f^{(j)} - f\|_0 > t + a j^{1/2-\alpha/d}) \leq e^{-bt^2 j^{2\alpha/d}}.$$

Proof. We have $f^{(j)} - f = (\mathcal{R}_j \mathcal{A}_n - I)f$, for $\mathcal{R}_j = \mathcal{A}^{-1}Q_j$ and $\mathcal{A}_n = \mathcal{I}_n \mathcal{A}$. Therefore, the probability on the left concerns the random variable $(\mathcal{R}_j \mathcal{A}_n - I)F$, if F is a variable distributed according to the prior Π. Since F is zero-mean normal with covariance operator $L^{-2\alpha}$, this variable is zero-mean Gaussian with covariance operator $(\mathcal{R}_j \mathcal{A}_n - I)L^{-2\alpha}(\mathcal{R}_j \mathcal{A}_n - I)^*$. We shall apply Borell's inequality to obtain the exponential bound, after computing the weak and strong second moments of the variable $(\mathcal{R}_j \mathcal{A}_n - I)F$.

Because $\langle (\mathcal{R}_j \mathcal{A}_n - I)F, g\rangle_0 = \langle F, (\mathcal{R}_j \mathcal{A}_n - I)^* g\rangle_0$ is zero-mean Gaussian with variance $\|L^{-\alpha}(\mathcal{R}_j \mathcal{A}_n - I)^* g\|_0^2 = \|(\mathcal{R}_j \mathcal{A}_n - I)^* g\|_{-\alpha}^2$, the weak second moment of $(\mathcal{R}_j \mathcal{A}_n - I)F$ is given by

$$\sup_{\|g\|_0 \leq 1} \mathbb{E}\langle (\mathcal{R}_j \mathcal{A}_n - I)F, g\rangle_0^2 = \sup_{\|g\|_0 \leq 1} \|(\mathcal{R}_j \mathcal{A}_n - I)^* g\|_{-\alpha}^2.$$

By the norm duality (2.1), the right side is equal to

$$\begin{aligned}\sup_{\|g\|_0 \leq 1} \sup_{\|f\|_\alpha \leq 1} \langle f, (\mathcal{R}_j \mathcal{A}_n - I)^* g\rangle_0^2 &\leq \sup_{\|f\|_\alpha \leq 1} \|(\mathcal{R}_j \mathcal{A}_n - I)f\|_0^2 \\ &\lesssim \left(\frac{\delta(n, \alpha+\gamma)}{\delta(j,\gamma)} + \delta(j,\alpha)\right)^2 \lesssim \delta(j,\alpha)^2,\end{aligned}$$

when $\delta(j,s) = j^{-s/d}$, $n \gg j$ and (5.9).

The strong second moment of the Gaussian variable $(\mathcal{R}_j \mathcal{A}_n - I)F$ is equal to the trace of its covariance operator. As

$$\mathrm{Trace}(S^* S) = \sum_i \|S\phi_i\|^2 = \sum_i \sum_j \langle S\phi_i, \phi_j\rangle^2 = \sum_i \|S^*\phi_i\|^2 = \|S\|_{HS}^2,$$

we have

$$\begin{aligned} &\mathbb{E}\|(\mathcal{R}_j \mathcal{I}_n \mathcal{A} - I)F\|^2 = \|(I - \mathcal{R}_{j_n} \mathcal{I}_n \mathcal{A})L^{-\alpha}\|_{HS}^2 \\ &\leq \|\mathcal{R}_{j_n}\|^2 \|\mathcal{I}_n - I : G_{\alpha+\gamma} \to G_0\|_{HS}^2 \|\mathcal{A}\|^2 \\ &+ \|(I - \mathcal{R}_{j_n}\mathcal{A}) : H_\alpha \to H_0\|_{HS}^2 \|L^{-\alpha} : H_0 \to H_\alpha\|^2 \\ &\lesssim_{(\mathcal{A},L)} \|\mathcal{R}_{j_n}\|^2 \|\mathcal{I}_n - I : G_{\alpha+\gamma} \to G_0\|_{HS}^2 + \|(I - \mathcal{R}_{j_n}\mathcal{A}) : H_\alpha \to H_0\|_{HS}^2, \end{aligned} \tag{8.23}$$

where the first inequality is because of an elementary application of triangle inequality and a norm estimation of the operators in Hilbert spaces (see Proposition A.20).

In the following argument we will use some results from approximation number (and more generally s-number) in Hilbert spaces. The necessary material is collected below, and for the detail we refer to Section 2.4. Recall that the lth *approximation number* of a bounded linear operator $T : X \to Y$ between normed spaces is defined as

$$a_l(T : G \to H) = \inf_{U:\mathrm{Rank}U < l} \|T - U\|_{X \to Y}$$

where the infimum is taken over all linear operators $U : X \to Y$ of rank (i.e., dimension of the range space) strictly less than j, and the norm on the right is the operator norm $\|T-U\|_{X\to Y} = \sup_{f:\|f\|_X\le 1}\|(T-U)f\|_Y$. Approximation number is in fact an example of a more general concept called *s-numbers*, which also include singular values. We will use the following fact: on Hilbert spaces there is only one s-number, i.e. singular value and approximation number are identical to each other.

Introduce a temporary notation $S := (I-\mathcal{R}_{j_n}\mathcal{A}) : H_\alpha \to H_0$, whose operator norm is $\|S : H_\alpha \to H_0\| \lesssim j_n^{-\alpha/d}$ by (5.9). Since $I : H_\alpha \to H_0$ is compact and $\mathcal{R}_{j_n}\mathcal{A}$ is of finite rank, S is compact as well. Hence it has singular values $s_l = a_l((I-\mathcal{R}_{j_n}\mathcal{A}) : H_\alpha \to H_0)$, and in particular, the first singular value is bounded by its operator norm, i.e. $s_1 \lesssim j_n^{-\alpha/d}$. Since $a_l(I : H_\alpha \to H_0) \simeq \delta(l,\alpha)$ (see the discussion following (2.6)), we have

$$a_l(S : H_\alpha \to H_0) \gtrsim \delta(j_n + l, \alpha),$$

which is because that the infimum on the right hand side is taken with all operators with rank less than $j_n + l$, while on the left hand side there is a fixed part $\mathcal{R}_{j_n}\mathcal{A}$ and the infimum is only taken over the operators with rank less than l. On the other hand, with the operator $U = R_{j_n+l}\mathcal{A} - \mathcal{R}_{j_n}\mathcal{A}$ of rank l, we have

$$a_l(S : H_\alpha \to H_0) \le \|I - \mathcal{R}_{j_n}\mathcal{A} - U\| \lesssim \delta(j_n + l, \alpha).$$

Combining the previous inequalities, we obtain

$$s_l = a_l(S : H_\alpha \to H_0) \simeq \delta(j_n + l, \alpha),$$

which leads to

$$\|(I-\mathcal{R}_{j_n}\mathcal{A}) : H_\alpha \to H_0\|_{HS}^2 = \sum_{l=1}^{\infty}(j_n + l)^{-2\alpha/d} = \sum_{l=j_n+1}^{\infty} l^{-2\alpha/d} \le j_n^{1-2\alpha/d},$$

where we used the estimate $\sum_{i>j_n} i^{-b} \le j_n^{1-b}/(b-1)$ for $b > 1$.

With the same argumentabove,

$$\|\mathcal{I}_n - I : G_{\alpha+\gamma} \to G_0\|_{HS}^2 \le n^{1-2(\alpha+\gamma)/d}.$$

Since $\alpha+\gamma > d/2$, $j_n \le n\varepsilon_n^2 < n$ and $\|\mathcal{R}_{j_n}\| \lesssim 1/\delta(j_n,\gamma) = j_n^{-\gamma/d}$, the first term in (8.23) is of order strictly smaller than the second term.

Since the first moment of $\|(\mathcal{R}_j\mathcal{A}_n - I)F\|_0$ is bounded by the root of its second moment, the lemma follows by Borell's inequality (see e.g. Lemma 3.1 and subsequent discussion in [67]). □

For $t^2 = n\varepsilon_n^2/(4bj_n^{2\alpha/d})$ the bound in the preceding lemma becomes $e^{-4n\varepsilon_n^2}$. Hence (8.11) is satisfied for

$$\eta_n \gtrsim \sqrt{n}\varepsilon_n/j_n^{\alpha/d} + j_n^{1/2-\alpha/d}.$$

Here we choose ε_n the minimal solution that satisfies the direct prior mass condition (8.10), given in (8.22). Next we solve for η_n under the constraints , (8.8) and

(8.9). The first of these constraints, $j_n \le n\varepsilon_n^2$, shows that the first term on the right side of the preceding display always dominates the second term. Therefore, we obtain the requirements $j_n \le n\varepsilon_n^2$ and

$$\eta_n \ge \sqrt{n}\, n^{-(\beta\wedge(\alpha-d/2)+\gamma)/(2\alpha+2\gamma)},$$
$$\eta_n \ge n^{-(\beta\wedge(\alpha-d/2)+\gamma)/(2\alpha+2\gamma)} j_n^{\gamma/d},$$
$$\eta_n \ge j_n^{-\beta/d}.$$

Depending on the relation between α and $\beta + d/2$, two situations need to be discussed separately.

(i) $\alpha < \beta + d/2$. We choose $j_n \simeq n^{d/(2\alpha+2\gamma)} = n\varepsilon_n^2$ and then see that the first two requirements in the preceding display both reduce to $\eta_n \ge n^{-(\alpha-d/2)/(2\alpha+2\gamma)}$, while the third becomes $\eta_n \ge n^{-\beta/(2\alpha+2\gamma)}$ and becomes inactive.

(ii) $\alpha > \beta + d/2$. We choose $j_n \simeq n^{d/(2\alpha+2\gamma)} \le n\varepsilon_n^2$, and then see that all three requirements reduce to $\eta_n \ge n^{-\beta/(2\alpha+2\gamma)}$.

Finally apply Theorem 8.6 to complete the proof.

8.6.3 Proof of Theorem 8.13

Let Π_τ denote the zero-mean Gaussian distribution on H with covariance operator $\tau^2 L^{-2\alpha}$ (where $\alpha > d/2$). The following lemmas are the counterparts of Lemmas 6.14 to 6.16.

Lemma 8.16. *Under the assumptions of Theorem 8.13, for $f_0 \in H_\beta$ and $(d/2-\gamma)\vee 0 < \beta \le \alpha$, as $\varepsilon \downarrow 0$,*

$$-\log \Pi_\tau(f : \|\mathcal{A}f - \mathcal{A}f_0\| < \varepsilon) \lesssim \frac{1}{\tau^2}\left(\frac{1}{\varepsilon}\right)^{(2\alpha-2\beta)/(\beta+\gamma)} + \left(\frac{\tau}{\varepsilon}\right)^{d/(\alpha+\gamma-d/2)}.$$

Lemma 8.17. *Under the assumptions of Theorem 8.13, for $f_0 \in H_\beta$ and $(d/2-\gamma)\vee 0 < \beta \le \alpha$, as $\varepsilon \downarrow 0$,*

$$-\log \Pi_\tau(f : \|f\|_0 < \varepsilon) \gtrsim \left(\frac{\tau}{\varepsilon}\right)^{d/(\alpha-d/2)}.$$

Lemma 8.18. *Under the assumptions of Theorem 8.13, there exist $a, b > 0$ such that, for every $j \in \mathbb{N}$ and $x, \tau > 0$,*

$$\Pi_\tau(f : \|f^{(j)} - f\|_0 > \tau x + \tau a j^{1/2-\alpha/d}) \le e^{-bx^2 j^{2\alpha/d}}$$

Proofs. The proofs are identical to the one in Section 6.6.3, and hence they are omitted. □

As preparation for the proof of Theorem 8.13, we first show that the minimax rate can be obtained by a Gaussian prior with the deterministic scaling, dependent on β, given by

$$\tau_n = n^{(\alpha-d/2-\beta)/(2\beta+2\gamma+d)}. \tag{8.24}$$

Theorem 8.19. *Assume the conditions on the Hilbert scale, the forward operator $\mathcal{A}$ and the true parameter f_0 in Theorem 8.11 hold. Suppose that the priors Π are zero-mean Gaussian with covariance operators $\tau_n^2 L^{-2\alpha}$ with τ_n as given in (8.24) and $\alpha > (d-\gamma) \vee d/2$. Then for $(d/2-\gamma) \vee 0 < \beta \le \alpha$, the posterior distribution satisfies, for sufficiently large $M > 0$,*

$$\Pi_n\Big(f : \|f - f_0\|_0 > Mn^{-\beta/(2\beta+2\gamma+d)} \mid Y^{(n)}\Big) \overset{P_{f_0}^{(n)}}{\to} 0.$$

Proof. The theorem is a corollary to Theorem 8.6. The proof follows the same lines as the proof of Theorem 8.11. By Lemma 8.16, inequality (8.10) is satisfied for

$$\varepsilon_n \gtrsim n^{-(\beta+\gamma)/(2\beta+2\gamma+d)}.$$

By Lemma 8.18, inequality (8.11) is satisfied for

$$\eta_n \gtrsim \tau_n(\sqrt{n}\varepsilon_n j_n^{-\alpha/d} + j_n^{1/2-\alpha/d}).$$

We choose $j_n \simeq n\varepsilon_n^2$, and the minimal solution $\varepsilon_n = n^{-(\beta+\gamma)/(2\beta+2\gamma+d)}$ to the second last display. It is then straightforward to verify that (8.8), (8.9) and (8.11) are satisfied for $\eta_n \simeq n^{-\beta/(2\beta+2\gamma+d)}$. □

Theorem 8.13 is a corollary of Theorem 8.7, with the choices

$$\eta_n \simeq n^{-\beta/(2\beta+2\gamma+d)}, \qquad \varepsilon_n \simeq n^{-(\beta+\gamma)/(2\beta+2\gamma+d)},$$
$$j_n \simeq n\varepsilon_n^2 = n^{d/(2\beta+2\gamma+d)}.$$

Conditions (8.7), (8.8), and (8.9) are satisfied for these choices. It remains to verify (8.10), and (8.17)–(8.16).

For ease of notation, for the moment, define η_n and ε_n as in the preceding display, with exact equality (i.e., with the constant set equal 1). Let τ_n be the 'optimal' scaling rate defined in (8.24).

Verification of (8.10). For $\tau \simeq \tau_n$ and $\varepsilon \simeq \varepsilon_n$ as given and $\beta \le \alpha$, both terms in the right side of Lemma 8.16 are of the order $n\varepsilon_n^2$. The lemma yields, for $\tau_n \le \tau \le 2\tau_n$ and some constant $a_1 > 0$,

$$-\log \Pi_\tau(f : \|\mathcal{A}f - \mathcal{A}f_0\| < \varepsilon_n) \le a_1 n\varepsilon_n^2.$$

This shows that

$$\begin{aligned}\Pi(f : \|\mathcal{A}f - \mathcal{A}f_0\| < \varepsilon_n) &= \int_0^\infty \Pi_\tau(f : \|\mathcal{A}f - \mathcal{A}f_0\| < \varepsilon_n)\, dQ(\tau) \\ &\ge e^{-a_1 n\varepsilon_n^2} Q(\tau_n, 2\tau_n).\end{aligned}$$

If $\alpha - d/2 < \beta$, then $\tau_n \to 0$, and Condition 8.12 on Q gives that

$$-\log Q(\tau_n, 2\tau_n) \lesssim \tau_n^{-2} = n^{(2\beta-2\alpha+d)/(2\beta+2\gamma+d)} \le n^{d/(2\beta+2\gamma+d)} = n\varepsilon_n^2,$$

if $\beta \leq \alpha$. If $0 < \beta < \alpha - d/2$, then $\tau_n \to \infty$, and Condition 8.12 on Q gives that

$$-\log Q(\tau_n, 2\tau_n) \lesssim \tau_n^{d/(\alpha-d/2)} = n^{(d(\alpha-d/2-\beta)/(\alpha-d/2)(2\beta+2\gamma+d))}$$
$$\leq n^{d/(2\beta+2\gamma+d)} = n\varepsilon_n^2.$$

Finally if $\alpha - d/2 = \beta$, then $\tau_n = 1$ and $Q(\tau_n, 2\tau_n) \gtrsim 1$. Thus in all three cases $Q(\tau_n, 2\tau_n)$ is bounded below by a power of $e^{-n\varepsilon_n^2}$. Combining this with the preceding, we see that $\Pi(f : \|\mathcal{A}f - \mathcal{A}f_0\|_n \leq \varepsilon_n) \geq e^{-a_2 n\varepsilon_n^2}$, for some positive constant a_2, which we can take bigger than 1. Then (8.10) is satisfied for ε_n equal to $\sqrt{a_2}$ times the current ε_n.

Verification of (8.16). Lemma 8.17 gives that

$$\Pi_\tau(f : \|f - f_0\|_0 < 2\eta_{n,\tau}) \leq \Pi_\tau(f : \|f\|_0 < 2\eta_{n,\tau}) \leq e^{-a_3(\tau/\eta_{n,\tau})^{d/(\alpha-d/2)}},$$

for some constant a_3. This is bounded above by $e^{-4a_2 n\varepsilon_n^2}$ if

$$\eta_{n,\tau} = 2a_4\tau\, n^{(d/2-\alpha)/(2\beta+2\gamma+d)} = 2a_4\tau\, \eta_n/\tau_n,$$

for a sufficiently small constant $a_4 > 0$.

Verification of (8.17). Choosing $x = a_4\eta_n/\tau_n = \eta_{n,\tau}/(2\tau)$ in Lemma 8.18, we see that the left side of (8.17) is bounded above by $e^{-4a_2 n\varepsilon_n^2}$ if j_n satisfies

$$a j_n^{1/2-\alpha/d} \leq a_4\eta_n/\tau_n, \qquad \text{and} \qquad b a_4^2(\eta_n/\tau_n)^2 j_n^{2\alpha/d} \geq 4a_2 n\varepsilon_n^2.$$

Both inequalities become equalities for j_n of the order $j_n \simeq n^{d/(2\beta+2\gamma+d)}$, as indicated at the beginning of the proof. Since $1/2-\alpha/d < 0$ and $2\alpha/d > 0$, the left side of the first inequality is decreasing in j_n and the left side of second inequality is increasing. Thus both inequalities are satisfied for $j_n = a_5 n^{d/(2\beta+2\gamma+d)}$ and a sufficienty large constant a_5.

Finally we choose ε_n and j_n in Theorem 8.7 equal to $\sqrt{a_2}$ and a_5 times the orders indicated at the beginning of the proof. Then (8.7) is satisfied, and (8.8) and (8.9) are satisfied if η_n is chosen of the indicated order times a sufficiently large constant.