# Bayesian inference for Gaussian models: Inverse problems and evolution equations

Yan, D.

**Author**: Yan, D.
**Title**: Bayesian inference for Gaussian models: Inverse problems and evolution equations
**Issue Date**: 2020-03-03

# Chapter 10

# Bayesian Inference for Linear Evolution Equations

To prepare the main focus of this chapter, the inference for linear evolution equations, we first recall some key results from Chapter 9. In this chapter, we always assume the following condition, in which we choose a concrete but widely applicable example for $H$.

**Condition 10.1.** With $H = L^2(\mathfrak{D})$, Assumption 9.12 and Assumption 9.14 are satisfied.

A stochastic linear evolution equation is given by

$$\begin{cases} dX(t) + \mathcal{L}X(t)\,dt = f(t)\,dt + \frac{1}{\sqrt{n}}B\,dW^{\mathcal{Q}}(t) \\ X(0) = u \in H \end{cases}, \tag{10.1}$$

where $u$ is the initial condition and $f : [0,T] \to H$ is the drift. Under Condition 10.1, the *mild* solution of (10.1) exists and is given by

$$X(t) = S(t)u + \int_0^t S(t-s)f(s)\,ds + \frac{1}{\sqrt{n}}\int_0^t S(t-s)B\,dW^{\mathcal{Q}}(s). \tag{10.2}$$

Furthermore, since the functions $u$ and $f$ in (10.1) admit the following representations,

$$u(x) = \sum_{k\in\mathbb{N}^d} u_k\varphi_k(x), \quad \text{and} \quad f(x,t) = \sum_{k\in\mathbb{N}^d} f_k(t)\varphi_k(x),$$

where $\{\varphi_k\}_{k\in\mathbb{N}^d}$ is the eigenbasis of $\mathcal{L}$. The solution (10.2) admits a series representation

$$X^{(n)}(t) = \sum_{k\in\mathbb{N}^d} X_k^{(n)}(t)\varphi_k, \tag{10.3}$$

whose coefficients are real-valued processes

$$X_k^{(n)}(t) = e^{-t\ell_k}u_k + \int_0^t e^{-(t-s)\ell_k}f_k(s)\,ds + \frac{b_k\sqrt{q_k}}{\sqrt{n}}\int_0^t e^{-(t-s)\ell_k}\,dW_k(s), \tag{10.4}$$

where $W_k(t)$ are independent standard Wiener processes, and the other constants are from the aforementioned assumptions.

In this chapter, we investigate the Bayesian approach to the recovery of the parameters $u$ and $f$ in (10.1). In Section 10.1, we study the recovery of initial condition $u$ and in Section 10.2 we investigate the inference for drift $f$. In each section, we start with introducing a Gaussian prior that is tailored to the problem. Then, the contraction rates are proved using the general framework developed in Chapter 4. It is worthwhile to mention that our proofs do not rely on the conjugacy nor other Gaussian properties of the prior. The contraction rates for other priors can also be obtained using the same argument, namely verifying the conditions in Theorem 4.10.

## 10.1 Recovery of the Initial condition

In this section, suppose that all other parameters except the initial condition $u$ are known. With no loss of generality, we can assume that $f(t) = 0$, for all $t \in [0, T]$.

**Condition 10.2** (Observation of Final Value)**.** Fix $T > 0$. For $n \in \mathbb{N}$, we observe the solution $X^{(n)}(T)$ of (10.1) at time $T$, i.e.

$$X^{(n)}(T) = S(T)u + \frac{1}{\sqrt{n}} \int_0^T S(T-s)B\, dW^{\mathcal{Q}}(s). \tag{10.5}$$

### 10.1.1 Spatial Gaussian Priors

Since the operator $\mathcal{L}$ governs the spatial status of the evolution system, it is natural to consider a smoothness class that adapts to the structure of $\mathcal{L} = \Lambda^{(\nu)}$.

Centred Gaussian distributions on a separable Hilbert space correspond bijectively to covariance operators. By definition a random variable $F$ with values in $H$ is Gaussian if $\langle F, g\rangle$ is normally distributed, for every $g \in H$, and it has zero mean if these variables have zero means. The variances of these variables can then be written as

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

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

Since the spatial regularity is characterized by isotropic Sobolev spaces $H_{\boldsymbol{s}}$, we introduce the following Gaussian priors, which are fully adapted to the spatial smoothness.

Given a multi-index $\boldsymbol{\alpha} = (\alpha, \cdots, \alpha) \in \mathbb{R}_+^d$, a spatial Gaussian prior is the law of

$$u = \sum_{k\in\mathbb{N}^d} u_k \varphi_k, \quad \text{with} \quad u_k \overset{\text{independent}}{\sim} \mathcal{N}_{\mathbb{R}}\left(0, |k^{\boldsymbol{\alpha}}|^{-2}\right). \tag{10.6}$$

From Lemma 2.18, we conclude that $\alpha > \frac{d}{2}$ guarantees a 'proper' Gaussian prior on $H$.

### 10.1.2 Contraction Rate for Initial Condition

The observation (10.5) given in Condition 10.2 can be rewritten as a general linear problem,

$$X^{(n)} = \mathcal{A}u + \frac{\xi}{\sqrt{n}},$$

where

$$\mathcal{A} = S(T), \quad \xi = \int_0^T S(T-s)B\, dW^{\mathcal{Q}}(s),$$

and $\xi$ is a proper Gaussian random element in $H$ with the covariance operator

$$\Sigma = \int_0^T S(T-r)B\mathcal{Q}B^* S(T-r)\, dr = \sum_{k\in\mathbb{N}^d} \sigma_k^2 \langle \varphi_k, \cdot\rangle \varphi_k,$$

with

$$\sigma_k^2 := \frac{b_k^2 q_k}{2\ell_k}\left(1 - e^{-2\ell_k T}\right) \simeq \frac{b_k^2 q_k}{2\ell_k}. \tag{10.7}$$

Because of Assumption 9.14, we have

$$\|u\|^2_{\mathbb{H}_\xi} = \sum_{k\in\mathbb{N}^d} \frac{u_k^2}{\sigma_k^2} \simeq \sum_{k\in\mathbb{N}^d} |k|^{2\mu} u_k^2, \tag{10.8}$$

which implies $\mathbb{H}_\xi = H_{\boldsymbol{\mu}}$.

Notice that due to (9.7) and Assumption 9.12, the operator $\mathcal{A} = S(T)$ possesses a smoothing property and the recovery of $u$ from (10.5) is in fact an inverse problem. By applying a general contraction result (Theorem 4.10) for inverse problems, modified from Theorem 3.1 in [43], the contraction rate of the spatial Gaussian prior from Section 10.1.1 for the recovery of initial condition $u$ from (10.5) is obtained in the following theorem.

**Theorem 10.3** (Gaussian Prior for the Initial Condition Recovery).
*With $\boldsymbol{s} = (s, \cdots, s) \in \mathbb{R}^d, s > 0$, let $\{H_{\boldsymbol{s}}\}_{\boldsymbol{s}}$ be the isotropic smoothness class introduced in Section 2.3.1 with the orthonormal basis $\{\varphi_k\}_{k\in\mathbb{N}^d}$ from Assumption 9.12 and with $\lambda_{k_i} = k_i$, $i = 1, \cdots, d$. Consider the prior given in Section 10.1.1 with $\alpha > (\nu + d)/2$, and $X^{(n)}$ be given in (10.5). For any $u_0 \in H_{\boldsymbol{\beta}}$ with $\beta > 0$, the posterior distribution satisfies, for sufficiently large $M > 0$,*

$$\Pi_n\Big(u : \|u - u_0\|_{L^2} > M(\log n)^{-s} \mid X^{(n)}\Big) \to 0$$

*in $\mathbb{P}^{(n)}_{u_0}$, with*

$$s = \frac{1}{\nu}\left[\left(\alpha - \frac{\nu + d}{2}\right) \wedge \beta\right].$$

The contraction rate is of logarithmic order, because of the exponential smoothing property of the semigroup $S(T)$. Similar phenomenon has also been discovered in the recovery of the initial condition in white noise, c.f. [60] and Chapter 7. A noteworthy observation is that the rate obtained in Theorem 10.3 is identical to the rate for the white noise case, see e.g. Theorem 7.14. In other words, the 'smoother' noise in (10.5), which in contrary to the white noise realises as a proper Gaussian element, does not lead to a faster rate. This is because the extreme ill-posedness from 'inverting' $S(T)$ predominantly resolves the logarithmic rate, and any noise with RKHS $H_{\boldsymbol{\mu}}$ with $\mu \in \mathbb{R}^+$ will not improve the order of the rate.

## 10.2 Recovery of the Drift

In this section, suppose that all other parameters except the drift

$$f : [0, T] \to H$$

are known. With no loss of generality, we can assume that $u = 0$.

**Condition 10.4** (Indirect Observation of Drift Term)**.** Fix $T > 0$. For $n \in \mathbb{N}$, we observe continuously the solution $X^{(n)}(t)$ of (10.1) for $0 \leq t \leq T$, i.e.

$$X^{(n)}(t) = \int_0^t S(t-s)f(s)\, ds + \frac{1}{\sqrt{n}} \int_0^t S(t-s)B\, dW^{\mathcal{Q}}(s). \tag{10.9}$$

### 10.2.1 Spatial-Temporal Gaussian Priors

For the recovery of drift terms, the priors necessarily need to sit in the function space $L^2([0,T]; H)$, where $H = L^2(\mathfrak{D})$. Since $L^2([0,T]; H) \cong H \otimes L^2([0,T]; \mathbb{R}) \cong L^2(\mathfrak{D}_T)$, one may introduce a Gaussian prior on the space $L^2(\mathfrak{D}_T)$ following the same procedure in the previous paragraphs. However, as mentioned in Section 2.3.1, it may be of interest to distinguish the smoothness in each spatial and temporal directions.

We introduce zero Gaussian priors on $L^2([0,T]; H)$ using series expansion. In order to do that, we fix an orthonormal basis $\{\psi_k\}_k$ of $L^2([0,T])$.

From Section 2.3.1, recall that given the orthonormal basis $\{\psi_l\}_{l\in\mathbb{N}}$ of $L^2([0,T]; \mathbb{R})$,

$$\{\widetilde{\varphi}_{k,l}\}_{(i,j)\in\mathbb{N}^d\times\mathbb{N}} = \{\varphi_k \otimes \psi_l\}_{(i,j)\in\mathbb{N}^d\times\mathbb{N}}$$

is an orthonormal basis of $L^2([0,T]; H)$, of which any function $f(x,t)$ admits the representation

$$\sum_{(k,l)\in\mathbb{N}^d\times\mathbb{N}} f_{k,l}\varphi_k(x)\psi_l(t) \quad \text{with} \quad \|\{f_k\}\|_{\ell^2(\mathbb{N}^{d+1})} = \sum_{k\in\mathbb{N}^{d+1}} f_k^2 < \infty.$$

Given a multi-index $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_{d+1}) \in \mathbb{R}_+^{d+1}$ and $\boldsymbol{\beta}_* = (p, \cdots, p, 0) \in \mathbb{R}_+^{d+1}$, the spatial-temporal Gaussian prior is the law of

$$f(x,t) = \sum_{(k,l)\in\mathbb{N}^d\times\mathbb{N}} f_{k,l}\widetilde{\varphi}_{k,l}(x,t) = \sum_{(k,l)\in\mathbb{N}^d\times\mathbb{N}} f_{k,l}\varphi_k(x)\psi_l(t), \tag{10.10}$$

with

$$f_k \overset{\text{independent}}{\sim} \mathcal{N}_{\mathbb{R}}\left(0, |k^{\beta_*}|^{-2}|k^{\alpha}|^{-2}\right),$$

i.e. $f_k$ are independent zero mean Gaussian random variables with variances

$$|k^{\boldsymbol{\beta}_*}|^{-2}|k^{\boldsymbol{\alpha}}|^{-2}=\left(1+\sum_{i\leq d}k_i^{2p}\right)^{-2}\left(\sum_{i\leq d+1}k_i^{2\alpha_i}\right)^{-2}.$$

By Lemma 2.18, when $\mathcal{H}(\boldsymbol{\alpha}) > (d+1)/2$, the prior has sample paths that are $H_{\boldsymbol{\beta}_*}$-valued almost surely.

We conclude this section with the following remark. Using the basis constructed above, the function $f_k$ of (10.3) can also be expressed as

$$f_k(t) = \sum_{l\in\mathbb{N}} f_{k,l}\psi_l(t).$$

### 10.2.2 Contraction Rate for Drift Recovery

In this section we study the performance of Bayesian methods in the recovery of the drift term $f \in L^2([0,T];H)$. As shown in Condition 10.4, i.e. (10.9)

$$X^{(n)}(t) = \int_0^t S(t-s)f(s)\,ds + \frac{1}{\sqrt{n}}\int_0^t S(t-s)B\,dW^{\mathcal{Q}}(s),$$

the noise is a vector-valued Gaussian process, whose RKHS is determined by the operator-valued Kernel $S(t-s)$. This imposes a challenging technicality, as some analytical tools such as the operator version of Mercer theorem is required in order to obtain a workable structure of RKHS. However, a unique characteristic of the observation under discussion is that the same (operator-valued) integral kernel $k(t,s) = S(t-s)$ is applied to both the drift and the noise. This property offers us a workaround to avoid the aforementioned difficulty: whitening the process. To be specific, by a proper transform of the signal, we will show that the observation (10.9) along its spatial basis is statistically equivalent to a sequence version of the white noise model (Section 10.3.2.1), the latter of which can be further related to a Gaussian $(d+1)$-dimensional sequence model (Section 10.3.2.2). As a consequence, the problem is reduced to standard nonparametric estimation without inverse nature, which is a multi-dimensional problem because the underlying space-time domain is a compact set in $\mathbb{R}^{d+1}$.

Now we show the contraction rates of the Gaussian prior Section 10.2.1 in the recovery of a Drift term.

**Theorem 10.5** (Gaussian Prior for the Drift Recovery).
*Let $\beta_* = (p,\cdots,p,0) \in \mathbb{R}_+^{d+1}$. For any $f_0 \in H_\beta$ with $\beta > \beta_*$, where $H_\beta$ is an anisotropic smoothness class defined in Section 2.3.1 with the orthonormal basis $\{\varphi_k \otimes \psi_l\}_{(i,j)\in\mathbb{N}^d\times\mathbb{N}}$ such that $\{\varphi_k\}_{k\in\mathbb{N}^d}$ from Assumption 9.12, an orthonormal basis $\{\psi_l\}_{l\in\mathbb{N}}$ of $L^2([0,T])$, and $\lambda_{k_i} = k_i$, $i = 1,\cdots,d+1$. Let the prior be zero-mean spatial-temporal Gaussian proposed in Section 10.2.1 with $\mathcal{H}(\alpha) > (d+1)/2$, and $X^{(n)}$ be the observations in the form of (10.9). The posterior distribution satisfies, for sufficiently large $M > 0$,*

$$\Pi_n\Big(f : \|f-f_0\|_{H_{\beta_*}} > Mn^{-s} \mid X^{(n)}\Big) \overset{P_{f_0}^{(n)}}{\to} 0,$$

*where*

$$s = \left(\frac{\mathcal{H}(\alpha) - (d+1)/2}{2\mathcal{H}(\alpha)}\right) \wedge \left(\frac{1}{2 + 2\sup_{i\le d+1} \frac{(\alpha_i - \beta_i)\vee 0}{\beta_i}}\right).$$

*In particular, when* $\alpha_i = \frac{2\mathcal{H}(\beta)+d+1}{2\mathcal{H}(\beta)}\beta_i$ *for each* $1 \le i \le d+1$, *the two items in the expression of* $s$ *above are balanced and*

$$s = \frac{\mathcal{H}(\beta)}{2\mathcal{H}(\beta) + d + 1}.$$

**Remark 10.6.** The contraction rate is given in the norm of smoothness class $H_{\beta_*}$. With a proper choice of the basis functions, such as Fourier basis, the space $H_{\beta_*}$ can be connected to certain type of multidimensional Sobolev spaces.

## 10.3 Proofs

### 10.3.1 Proofs in Section 10.1

The theorem is a corollary to Theorem 4.10. The main tasks are to determine $\varepsilon_n$ satisfying the prior mass condition (4.12) of the direct problem, and next to identify $\eta_n$ from the prior mass condition (4.13) and the other conditions.

The first task is achieved in the following lemma.

**Lemma 10.7.** *For* $f_0 \in H_{\boldsymbol{s}}$, *the prior* $\Pi$ *from Section 10.1.1, as* $\varepsilon \downarrow 0$,

$$-\log \Pi(f : \|\mathcal{A}f - \mathcal{A}f_0\|_{\mathbb{H}_\xi} < \varepsilon) \lesssim \left(\log\frac{1}{\varepsilon}\right)^r, \tag{10.11}$$

*where*

$$r = \left(2\frac{\alpha-\beta}{\nu}\right) \vee \left(\frac{\nu+d}{\nu}\right).$$

**Remark 10.8.** Since $\nu \in \mathbb{N}$ from Assumption 9.12 and $d \in \mathbb{N}$, $r > 1$.

*Proof.* The probability in the left side is the decentred small ball probability $\Pi(g : \|g - g_0\|_{\mathbb{H}_\xi} < a\varepsilon)$ of the Gaussian random variable $G = \mathcal{A}F$ distributed according to the prior under the linear transform $\mathcal{A}$. Symbolically we denote the covariance operator (which is diagonal) of $F$ by $\Lambda_F$. Due to the property of Gaussian measure, the random element $G$ is also a centred Gaussian random element in $H$ with the covariance operator

$$\mathcal{A}\Lambda_F\mathcal{A}^* : h = \sum_{k\in\mathbb{N}^d} h_k\varphi_k \mapsto \sum_{k\in\mathbb{N}^d} e^{-2T|k^{\boldsymbol{\nu}}|}|k^{\boldsymbol{\alpha}}|^{-2}h_k\varphi_k. \tag{10.12}$$

The RKHS $\mathbb{H}_G$ of $G$ is given by

$$\left\{ g = \sum_{k\in\mathbb{N}} g_k\varphi_k \in H : \|g\| = \sum_{k\in\mathbb{N}^d} e^{2T|k^{\boldsymbol{\nu}}|}|k^{\boldsymbol{\alpha}}|^2 g_k^2 < \infty \right\}. \tag{10.13}$$

It is convenient to work with the following norm of $\mathbb{H}_G$,

$$\|g\|_{\mathbb{H}_G}^2 := \sum_{k\in\mathbb{N}^d} e^{2T|k|^\nu}|k|^{2\alpha}g_k^2,$$

which is equivalent to the norm in (10.13).

Recall (10.8), we have $\mathbb{H}_\xi = H_{\boldsymbol{\mu}}$ (as sets) and $\|h\|_{\mathbb{H}_\xi}\simeq \|h\|_{H_\mu}$. Due to the exponential smoothing property of $\mathcal{A}$, for any $f\in H$ and $\boldsymbol{s}\in\mathbb{R}_+^d$, we have $\mathcal{A}f\in H_{\boldsymbol{s}}$. Hence $\Pr(G\in\mathbb{H}_\xi)=1$. Hence the distribution of $G$ can be considered as a Gaussian measure on $\mathbb{H}_\xi$ with RKHS $\mathbb{H}_G$.

The left side of (10.11) is therefore up to constants equivalent to

$$\inf_{g\in\mathbb{H}_G:\|g-g_0\|_{\mathbb{H}_\xi}<\varepsilon}\|g\|_{\mathbb{H}_G}^2 - \log\Pi(\|g\|_{\mathbb{H}_\xi}<\varepsilon). \tag{10.14}$$

See [64, 65, 99], or Section 11.2, in particular, Proposition 11.19 in [35].

Let $P_j$ be the $H_0$-orthonormal projection to the $j$ basis $\{\varphi_k\}_{|k|_\infty\le j^{1/d}}$. Since $\mathcal{A}$ and $P_j$ commute, using Remark 9.13, we have

$$\begin{aligned}
&\|P_j\mathcal{A}f_0-\mathcal{A}f_0\|_{\mathbb{H}_\xi}^2\\
&=\sum_{|k|_\infty\ge j^{1/d}} e^{-2T\ell_k}|k|^{2\mu}f_{0,k}^2 \simeq_d \sum_{|k|_\infty\ge j^{1/d}} e^{-2T|k|^\nu}|k|^{2(\mu-\beta)}\left(|k|^{2\beta}f_{0,k}^2\right)\\
&\lesssim_d \exp\left(-2Tj^{\nu/d}\right)j^{-2(\beta-\mu)/d}\|f_0\|_\beta^2,
\end{aligned}\tag{10.15}$$

and hence $\|P_j\mathcal{A}f_0-\mathcal{A}f_0\|_{\mathbb{H}_\xi}$ is bounded above by $\varepsilon$ for $j\simeq_T(-\log\varepsilon)^{d/\nu}$. By substituting this value of $j$ into

$$\begin{aligned}
\|P_j\mathcal{A}f_0\|_{\mathbb{H}_G}^2 &= \sum_{|k|_\infty\le j^{1/d}}|k|^{2\alpha}f_{0,k}^2 \lesssim_d \sum_{|k|_\infty\le j^{1/d}}|k|^{2\alpha-2\beta}|k|^{2\beta}f_{0,k}^2\\
&\le j^{2\frac{(\alpha-\beta)\vee 0}{d}}\|f_0\|_\beta^2,
\end{aligned}$$

we conclude that the first term in (10.14) is bounded above by $(-\log\varepsilon)^{2\frac{\alpha-\beta}{\nu}\vee 0}$.

For the second term in (10.14), by Corollary 10.15, the metric entropy

$$\log N(\varepsilon,\{g\in\mathbb{H}_G:\|g\|_{\mathbb{H}_G}\le 1\},\|\cdot\|_{\mathbb{H}_\xi})\simeq(-\log\varepsilon)^{(\nu+d)/\nu}.$$

Hence, by [64] (see Lemma 6.2 in [100]),

$$-\log\Pi(\|\mathcal{A}f\|_{\mathbb{H}_\xi}<\varepsilon)\simeq\left(\log\frac{1}{\varepsilon}\right)^{(\nu+d)/\nu}.$$

Finally, the assertion of the lemma follows from combining the above results. □

It follows that (4.12) is satisfied for any $\varepsilon_n$ such that

$$e^{-(\log\frac{1}{\varepsilon_n})^r}\ge e^{-n\varepsilon_n^2},$$

where $r>1$ is given in the lemma above. Let $x=-\log\varepsilon_n$. The preceding display can be rewritten into

$$\frac{2}{r}xe^{\frac{2}{r}x}=x^*e^{x^*}\le\frac{2}{r}n^{1/r}.$$

The following lemma is useful for the proof.

**Lemma 10.9.** *Let $W(x)$ be the Lambert $W$ function, i.e. the inversion of the mapping $[e^{-1}, \infty) \ni x \mapsto xe^x$. We have, when $x \to \infty$,*

$$W(x) \sim \log x - \log\log x.$$

*Proof.* From the identity $x = W(x)e^{W(x)}$, $W(x)$ is an increasing function with respect to $x$ and $W(e) = 1$. In addition, we also have the following identities,

$$W(x) = \log\left(\frac{x}{W(x)}\right) \quad \text{and} \quad W(x) + \log W(x) = \log x.$$

From now on only consider the case $x > e$. The second relation in the last display implies that $W(x) < \log x$. Hence,

$$W(x) = \log\left(\frac{x}{W(x)}\right) > \log\left(\frac{x}{\log x}\right) = \log x - \log\log x.$$

On the other hand,

$$W(x) = \log\left(\frac{x}{\log\left(\frac{x}{W(x)}\right)}\right) < \log\left(\frac{x}{\log\left(\frac{x}{\log x}\right)}\right) = \log x - \log\log x + \log\log\log x.$$

Since $\log\log x \gg \log\log\log x$ as $x \to \infty$, the proof is complete. □

The asymptotic expansion of the Lambert $W$ function implies that $x$ such that

$$x = \frac{r}{2}W\left(\frac{2}{r}n^{1/r}\right) \sim \log\frac{\sqrt{n}}{(\log n)^{r/2}}$$

satisfies the last inequality above. Consequently, (4.12) is satisfied with

$$\varepsilon_n \simeq \frac{(\log n)^{r/2}}{\sqrt{n}}, \tag{10.16}$$

Now we construct the reconstruction operator $\mathcal{R}_n : H \to H$. For $g \in H$, we consider the following truncation regularizer,

$$\mathcal{R}_n g := \sum_{|k|_\infty < j_n^{1/d}} e^{T\ell_k} g_k \varphi_k, \tag{10.17}$$

where $j_n \to \infty$ as $n \to \infty$. Consequently, from Assumption 9.12, with some positive constant $c$, we have

$$\|\mathcal{R}_n\| = \sup_{|k|_\infty < j_n^{1/d}} \exp\left(|k^{\boldsymbol{\nu}}|T\right) = e^{cj_n^{\nu/d}T}, \tag{10.18}$$

and (4.6) is satisfied with $\rho_n = e^{cj_n^{\nu/d}T}$.

Besides, for $u_0 = \sum_{k\in\mathbb{N}^d} u_{0,k}\varphi_k \in H_{\boldsymbol{\beta}}$, by Lemma 2.20, we have

$$\|\mathcal{R}_n\mathcal{A}u_0 - u_0\|^2 = \sum_{|k|_\infty \geq j_n^{1/d}} u_{0,k}^2 \lesssim_d j_n^{-2\beta/d}\|u_0\|_{H_{\boldsymbol{\beta}}}^2. \tag{10.19}$$

The next step of the proof is to bound the prior probability in (4.13).

**Lemma 10.10.** *Let $\mathcal{R}_n$ be given as (10.17) and the corresponding $j_n = j$. There exist $a, b > 0$, such that for every $j \in \mathbb{N}$ and $t > 0$,*

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

*Proof.* Let $f^{(n)} = \mathcal{R}_n \mathcal{A} f$. Therefore, the probability on the left concerns the random variable $(\mathcal{R}_n \mathcal{A} - I)F$, if $F$ is a variable distributed according to the prior $\Pi$. Since $F$ is zero-mean normal with a covariance operator symbolically denoted by $\Lambda_F$, this variable is zero-mean Gaussian with covariance operator $(\mathcal{R}_n \mathcal{A} - I)\Lambda_F(\mathcal{R}_n \mathcal{A} - I)^*$. We shall compute the weak and strong second moments of the variable $(\mathcal{R}_n \mathcal{A} - I)F$, and next apply Borell's inequality for the norm of a Gaussian variable to obtain the exponential bound.

Since

$$(\mathcal{R}_n \mathcal{A} - I)F = - \sum_{|k|_\infty \geq j^{1/d}} F_k \varphi_k$$

with $F_k \overset{i.i.d.}{\sim} \mathscr{N}_{\mathbb{R}}(0, |k^{\boldsymbol{\alpha}}|^{-2})$, we have

$$\langle (\mathcal{R}_n \mathcal{A} - I)F, g \rangle = - \sum_{|k|_\infty \geq j^{1/d}} F_k g_k,$$

for arbitrary $g = \sum_{k \in \mathbb{N}^d} g_k \varphi_k$. Then, the weak second moment of $(\mathcal{R}_n \mathcal{A} - I)F$ is given by

$$\sup_{\|f\|_0 \leq 1} \mathbb{E}\langle (\mathcal{R}_n \mathcal{A} - I)F, f \rangle^2 \leq \sup_{\sum_k f_k^2 \leq 1} \sum_{|k|_\infty > j^{1/d}} |k^{\boldsymbol{\alpha}}|^{-2} f_k^2 \simeq j^{-2\alpha/d}.$$

The strong second moment of the Gaussian variable $(\mathcal{R}_n \mathcal{A} - I)F$ is

$$\mathbb{E}\|(\mathcal{R}_n \mathcal{A} - I)F\|^2 = \sum_{|k|_\infty \geq j^{1/d}} |k|^{-2\alpha}.$$

In the proof of Lemma 2.18, we have shown that for the hypercubes

$$C_n = \{k \in \mathbb{N}^d : k_i \lesssim n^{1/d}, \ i = 1, \cdots, d\},$$

the increment $C_n \backslash C_{n-1}$ covers index points of the order $n^{d-1}$. Hence, the strong second moment can be bounded by

$$\sum_{|k|_\infty \geq j^{1/d}} |k|^{-2\alpha} \leq_d \sum_{l \geq j^{1/d}} \sum_{k \in [C_l \backslash C_{l-1}]} \prod_{i \leq d} k_i^{-2\alpha/d} \simeq \sum_{l \geq j^{1/d}} l^{d-1-2\alpha} \leq j^{(d-2\alpha)/d},$$

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

Since the first moment of $\|(\mathcal{R}_n A - I)F\|_0$ 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]). □

Let $t^2 = 4n\varepsilon_n^2/(bj_n^{2\alpha/d})$. Then, $t \gtrsim j_n^{1/2-\alpha/d}$, due to the constraint (4.8), i.e. $j_n \lesssim n\varepsilon_n^2$. Substituting $t$ into Lemma 10.10,

$$\Pi(f : \|\mathcal{R}_n \mathcal{A} f - f\|_0^2 > 4n\varepsilon_n^2/(bj_n^{2\alpha/d})) \leq e^{-4n\varepsilon_n^2},$$

which together with (10.16) implies

$$\eta_n \gtrsim j_n^{-\alpha/d} (\log n)^{r/2}.$$

In addition, the constraints eqs. (4.9) and (4.10) impose

$$\begin{aligned} \eta_n &\gtrsim e^{Tcj_n^{\nu/d}} \frac{(\log n)^{r/2}}{\sqrt{n}}, \\ \eta_n &\gtrsim j_n^{-\beta/d}, \end{aligned}$$

where the right-hand side of the inequalities are given by (10.18) and (10.19).

We need to determine $j_n$ in order to solve for $\eta_n$. Since $d/\nu < r$, $(\log n)^{d/\nu} \ll n\varepsilon_n^2$. Hence, we can choose $j_n = \tilde{c}^{d/\nu} (\log n)^{d/\nu} \lesssim n\varepsilon_n^2$, with $\tilde{c}$ such that $\tilde{c}Tc < 1/2$. Substituting $j_n$ into the preceding constraints leads to

$$\begin{aligned} \eta_n &\gtrsim (\log n)^{-\frac{2\alpha-\nu-d}{2\nu}}, \\ \eta_n &\gtrsim n^{-(\frac{1}{2}-\tilde{c}Tc)} (\log n)^{r/2}, \\ \eta_n &\gtrsim (\log n)^{-\beta/\nu}. \end{aligned}$$

The second inequality above is negligible compared to the other two. The theorem follows from Theorem 4.10.

### 10.3.2 Proofs in Section 10.2

The major step of the proof can be summarized as follows.

(1) Section 10.3.2.1. Consider the sequence of scalar processes $\{X_k^{(n)}\}_{k \in \mathbb{N}^d}$ from (9.13). Using a sequence of transforms $\{\mathcal{T}_k\}_{\mathbb{N}^d}$, the sequence of processes $\{X_k^{(n)}\}_{k \in \mathbb{N}^d}$ is whitened in time.

(2) Section 10.3.2.2. The signal $f$ to recover is also isometrically transformed into $\widetilde{f}$. The transformed observation can be expressed with a Gaussian sequence,

$$\widetilde{X}_{k,l}^{(n)} = f_{k,l} + \frac{1}{\sqrt{n}} \widetilde{\xi}_{k,l} \in \mathbb{R}, \quad (k,l) \in \mathbb{N}^d \times \mathbb{N},$$

where the covariance structure of $\widetilde{\xi}_{k,l}$ is determined by the operators $B$ and $\mathcal{Q}$.

(3) Section 10.3.2.3. As the final preparation for the proof of Theorem 10.5, we establish a posterior contraction rate for the multi-dimensional white noise model.

(4) Section 10.3.2.4 Using the result obtained in Section 10.3.2.3, and an isometric property possessed by the prior and the noise $\widetilde{\xi}$, we finally conclude the proof of Theorem 10.5.

Recall that the eigenbasis $\{\varphi_k\}_{k\in\mathbb{N}^d}$ of $\mathcal{L}$ is an orthonormal basis in space and $\{\psi_k\}_{k\in\mathbb{N}}$ is an orthonormal basis in time, and denote their tensor product by $\{\widetilde{\varphi}_{k,l} = \varphi_k \otimes \psi_l\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}}$.

#### 10.3.2.1 Whitening Ornstein-Uhlenbeck processes

Due to (9.13), the observation (10.9) is equivalent to the following (functional) sequence model, for $k \in \mathbb{N}^d$, we observe

$$X_k^{(n)}(t) = \int_0^t e^{-\ell_k(t-s)} f_k(s)\, ds + \frac{b_k\sqrt{q_k}}{\sqrt{n}} \int_0^t e^{-\ell_k(t-s)}\, dW_k(s), \tag{10.20}$$

in the product space $(L^2([0,T];\mathbb{R}))^{\mathbb{N}^d}$ with the product measure $\bigotimes_{k\in\mathbb{N}^d}(\mu_{W_k})$, where $\mu_{W_k}$ are the probability measures induced by the processes $W_k$ given in (9.12), which are mutually independent Wiener processes.

Notice that the real-valued processes $X_k^{(n)}(t)$ is Ornstein-Uhlenbeck processes. We are going to convert them into the standard white noise model, and start with introducing a useful function together with its inverse. For $\lambda > 0$ and $t \in [0,T]$, define

$$\vartheta(t) = \frac{\log(2\lambda t + 1)}{2\lambda} \quad \text{and} \quad \vartheta^{-1}(t) = \frac{e^{\lambda t} - 1}{2\lambda}, \tag{10.21}$$

where $\vartheta^{-1}$ is well-defined since $\vartheta : [0,T) \to \mathbb{R}^+$ is bijective. With function $\vartheta$, we can define the following transform

$$(\mathcal{T}g)(t) := \sqrt{2\lambda t + 1}(g \circ \vartheta)(t), \quad t \in [0, \vartheta(T)], \tag{10.22}$$

for any continuous function $g$ on $[0,T]$.

Using the newly defined transform $\mathcal{T}$, the noise can be whitened as follows.

**Lemma 10.11.** *Let $W(t)$ be a Brownian motion, i.e. a standard real-valued Wiener process, and $\vartheta(t)$ be given in (10.21). If*

$$\xi(t) = \int_0^t e^{-\lambda(t-s)}\, dW(s),$$

*then $(\mathcal{T}\xi)(t) = \sqrt{2\lambda t + 1}(\xi \circ \vartheta)(t)$ is a Brownian motion.*

*Proof.* The process $M(t) = e^{\lambda t}\xi(t)$ is a continuous martingale whose quadratic variation is

$$[M]_t = \int_0^t e^{2\lambda s}\, ds = \frac{e^{2\lambda t} - 1}{2\lambda} = \vartheta^{-1}(t).$$

Consequently, $M \circ \vartheta(t)$ is a continuous martingale with

$$[M \circ \vartheta]_t = [M]_{\vartheta(t)} = t.$$

Therefore,

$$M \circ \vartheta(t) = e^{\lambda\vartheta(t)}\xi \circ \vartheta(t) = \sqrt{2\lambda t + 1}(\xi \circ \vartheta)(t)$$

is a Brownian motion. □

Similarly, the transform (10.22) can be applied to the deterministic integral.

**Lemma 10.12.** *Assume $f \in L^2[0,T]$. Let*

$$F(t) = \int_0^t e^{-\lambda(t-s)} f(s)\, ds$$

*and $\tilde{f}(u)$ be the transform of function $f$ such that*

$$\tilde{f}(u) = \frac{f \circ \vartheta(u)}{\sqrt{2\lambda u + 1}}. \tag{10.23}$$

*Then, the following statements hold.*

*(i)* $(\mathcal{T}F)(t) = \sqrt{2\lambda t + 1}(F \circ \vartheta)(t) = \int_0^t \tilde{f}(u)\, du.$

*(ii) If $\{\psi_k\}_k$ is an orthonormal basis for $L^2[0,T]$, then,*

$$\int_0^{\vartheta(T)} \tilde{\psi}_k \tilde{\psi}_l \, du = \delta_{kl} \quad \textit{and} \quad \int_0^{\vartheta(T)} \tilde{f}\tilde{\psi}_k \, du = \int_0^T f\psi_k \, du.$$

*Proof.* Since $f \in L^2$, $F$ is continuous. The first statement follows from

$$\begin{aligned}
(\mathcal{T}F)(t) =& \sqrt{2\lambda t + 1} \int_0^{\vartheta(t)} e^{-\lambda(\vartheta t - s)} f(s)\, ds \\
\overset{s=\vartheta(u)}{=}& \sqrt{2\lambda t + 1} e^{-\lambda\vartheta(t)} \int_0^t e^{\lambda\vartheta(u)} f \circ \vartheta(u) \vartheta'(u)\, du \\
=& \int_0^t f \circ \vartheta(u) \frac{1}{\lambda}(e^{\lambda\vartheta(u)})'\, du = \int_0^t \frac{f \circ \vartheta(u)}{\sqrt{2\lambda u + 1}}\, du,
\end{aligned}$$

where $(\cdot)'$ denotes the ordinary derivative.

The next statement is obtained by changing variables. The first equation follows from

$$\int_0^{\vartheta(T)} \tilde{\psi}_k \tilde{\psi}_l \, du = \int_0^{\vartheta(T)} \frac{\psi_k(\vartheta(u))\psi_l(\vartheta(u))}{2\lambda u + 1}\, du \overset{s=\vartheta(u)}{=} \int_0^T \psi_k(s)\psi_l(s)\, ds = \delta_{kl}.$$

The same argument also applies to the second one. □

Applying the transform $\mathcal{T}$ defined in (10.22) to

$$X(t) = \int_0^t e^{-\lambda(t-s)} f(s)\, ds + c\int_0^t e^{-\lambda(t-s)}\, dW(s), \quad t \in [0,T],$$

we obtain

$$\tilde{X}(t) := \mathcal{T}X(t) = \int_0^t \tilde{f}(s)\,ds + c\tilde{W}(t), \quad t \in [0, \vartheta(T)],$$

where $\tilde{f}$ is given in (10.23) and $\tilde{W}(t)$ is a Brownian motion.

Now given an orthonormal basis $\{\psi_l\}_{l\in\mathbb{N}}$ of $L^2[0,T]$, we can form, for $l \in \mathbb{N}$,

$$\int_0^{\vartheta(T)} \tilde{\psi}_l\, d\tilde{X}(s) = \int_0^T f(s)\psi_l(s)\,ds + c\int_0^{\vartheta(T)} \tilde{\psi}_l\, d\tilde{W} = f_l + cz_l, \tag{10.24}$$

where $z_l$ are i.i.d. standard Gaussian.

#### 10.3.2.2 Complete Sequence Model

Now consider the independent signals $X_k^{(n)}$ as given in (10.20). Define $\mathcal{T}_k$ as the transform (10.22) from the previous section, with $\lambda = \ell_k$ and $c_k = (\sqrt{q_k}b_k)/\sqrt{n}$. Then, we can transform the signals into

$$\tilde{X}_k^{(n)}(t) := \mathcal{T}_k X_k^{(n)}(t) = \int_0^t \tilde{f}_k(s)\,ds + c_k\tilde{W}_k(t), \tag{10.25}$$

where $\tilde{W}_k(t)$ are independent Brownian motions. Because of (10.24), we can form observations

$$\widetilde{X}_{k,l}^{(n)} = f_{k,l} + \frac{b_k\sqrt{q_k}}{\sqrt{n}} z_{k,l} \in \mathbb{R}, \quad (k,l) \in \mathbb{N}^d \times \mathbb{N},$$

where $z_{k,l} \overset{i.i.d.}{\sim} \mathscr{N}(0,1)$. In fact, these are the observations of the coordinates of the following multidimensional Gaussian sequence model,

$$\widetilde{X}^{(n)} = \widetilde{f} + \frac{1}{\sqrt{n}}\widetilde{\xi}, \tag{10.26}$$

where $\widetilde{f} = \{f_{k,l}\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}} \in \ell^2(\mathbb{N}^d \times \mathbb{N})$ are the coefficients of $f$ in the series representation with basis $\{\varphi_k \otimes \psi_l\}_{k\in\mathbb{N}^d, l\in\mathbb{N}}$ and $\widetilde{\xi} = \{\widetilde{\xi}_{k,l}\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}}$ is a random vector in $\mathbb{R}^{\mathbb{N}^{d+1}}$ whose entries are independent zero mean Gaussian random variables with variance $\left(\sum_{i\le d} k_i^p\right)^{-2}$. This variance is determined by the decay of $b_k\sqrt{q_k}$ under Assumption 9.14.

#### 10.3.2.3 Gaussian Posterior Contraction for Multi-dimensional White Noise Model

In this subsection, as the final preparation for proving Theorem 10.5, we prove the posterior contraction rate of the equivalent prior of Section 10.2.1 on sequence spaces, for a simple Gaussian sequence model. To be precise, we consider the following situation.

Let $\alpha \in \mathbb{R}_+^m$ be a multi-index. Consider a prior as the law of $F = \{F_k\}_{k\in\mathbb{N}^m}$, where $F_k$ are independent centred real Gaussian random variables with variance $|k^\alpha|^{-2}$. We impose $\mathcal{H}(\alpha) > m/2$ so that the prior are almost surely realised in the space $\ell^2(\mathbb{N}^m)$.

**Lemma 10.13.** *Consider the observation is given by the multi-dimensional white noise model,*

$$X_k^{(n)} = f_k + \frac{1}{\sqrt{n}} z_k, \quad \text{for } k \in \mathbb{N}^m, \tag{10.27}$$

*where $z_k$ are independent standard Gaussian random variables. If true parameter $f_0 = \{f_{0,k}\}_{k\in\mathbb{N}^m} \in h_\beta$, where $h_\beta$ is a Sobolev ellipsoid defined in Section 2.3.1 equipped with norm (2.17), then the Gaussian prior above behaves, as $\varepsilon \downarrow 0$,*

$$-\log \Pi(f : \|f - f_0\|_{\ell^2} < \varepsilon) \lesssim \varepsilon^{-r_1} \vee \varepsilon^{-r_2} \tag{10.28}$$

*with*

$$r_1 = 2 \sup_{i\le m} \frac{(\alpha_i - \beta_i) \vee 0}{\beta_i} \quad \text{and} \quad r_2 = \frac{m}{\mathcal{H}(\alpha) - m/2}.$$

*Proof.* Notice that the RKHS $\mathbb{H}_F$ of the prior $F$ is $h_\alpha \subset \ell^2$. The left side of (10.28) is up to constants equivalent to

$$\inf_{f\in h_\alpha:\|f-f_0\|_{\ell^2}<\varepsilon} \|f\|_{h_\alpha}^2 - \log \Pi(\|f\|_{\ell^2} < \varepsilon). \tag{10.29}$$

See [64, 65, 99], or Section 11.2, in particular, Proposition 11.19 in [35].

Let $P_N$ be the truncation of a sequence to $\{k < N : k_i < N_i,\ i \le m\}$ with $N = (N_1, \cdots, N_m)$. Applying Lemma 2.20, we obtain $\|P_I f_0 - f_0\|_{\ell^2} \le \varepsilon$, if $N_i \gtrsim \varepsilon^{-1/\beta_i}$ for all $i \le m$.

Taking $N_i \simeq \varepsilon^{-1/\beta_i}$, $i \le m$, an upper bound on the first term in (10.28) is obtained as,

$$\begin{aligned}
\|P_N f_0\|_{\mathbb{H}_F}^2 &= \sum_{k<N} |k^\alpha|^2 f_{0,k}^2 \le \sum_{k<N} \left[\sum_{i\le m} k_i^{2\beta_i} k_i^{2(\alpha_i-\beta_i)\vee 0}\right] f_{0,k}^2 \\
&\lesssim_d \sum_{k<N} |k^\beta|^2 \left[\sum_{i\le m} k_i^{2(\alpha_i-\beta_i)\vee 0}\right] f_{0,k}^2 \le \left[\sum_{i\le m} N_i^{2(\alpha_i-\beta_i)\vee 0}\right] \|f_0\|_{h_\beta}^2 \\
&\simeq \sum_{i\le m} \varepsilon^{-\frac{2(\alpha_i-\beta_i)\vee 0}{\beta_i}} \|f_0\|_{h_\beta}^2 \le d\, \varepsilon^{-\sup_{i\le m} \frac{2(\alpha_i-\beta_i)\vee 0}{\beta_i}} \|f_0\|_{h_\beta}^2
\end{aligned}$$

For the second term (small ball probability) in (10.28), by Corollary 2.24, the metric entropy $\log N(\varepsilon, \{f \in h_\beta : \|f\|_{h_\beta} \le 1\}, \|\cdot\|_{\ell^2})$ is of the order $\varepsilon^{-m/\mathcal{H}(\alpha)}$. Hence, under the condition $\mathcal{H}(\alpha) > d/2$, by [64] (see Lemma 6.2 in [100]),

$$-\log \Pi(\|f\|_{\ell^2} < \varepsilon) \simeq \varepsilon^{-\frac{m}{\mathcal{H}(\alpha)-m/2}}.$$

□

According to equations (1.2) nd (1.3) in [99], the minimal $\varepsilon_n$ satisfying

$$-\log \Pi(f : \|f - f_0\|_{\ell^2} < \varepsilon_n) \lesssim n\varepsilon_n^2$$

is the posterior contraction rate of the Gaussian prior considered in this section. By direct calculation, it can be shown that when for all $1 \leq i \leq m$,

$$\alpha_i = \left( \frac{2\mathcal{H}(\beta) + m}{2\mathcal{H}(\beta)} \right) \beta_i,$$

the rates $r_1$ and $r_2$ in (10.28) are balanced to $r_1 = r_2 = m/\mathcal{H}(\beta)$ and the posterior contraction rate reaches the minimax rate (see [53]), i.e.

$$\varepsilon_n \simeq n^{-\frac{\mathcal{H}(\beta)}{2\mathcal{H}(\beta)+m}}.$$

#### 10.3.2.4 Proof of Theorem 10.5

After the long preparation, the proof of Theorem 10.5 simply follows from assembling all the results obtained up to now.

Recall $\{\widetilde{\varphi}_{k,l}\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}} = \{\varphi_k \otimes \psi_l\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}}$ is a fixed basis of $L^2(\mathfrak{D}_T)$, satisfying the assumptions in Section 9.5. For a function $f$ in $L^2(\mathfrak{D}_T)$, its coefficients in the basis $\{\widetilde{\varphi}_{k,l}\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}}$ are denoted by $\widetilde{f} = \{f_{k,l}\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}}$. Recall that for the norms $\|\cdot\|_{L^2}, \|\cdot\|_{H_{\boldsymbol{\beta}}}, \|\cdot\|_{\ell^2}, \|\cdot\|_{h_{\boldsymbol{\beta}}}$ from Section 2.3, we have the isometries

$$\|f\|_{L^2} = \|\widetilde{f}\|_{\ell^2}, \quad \|f\|_{H_{\boldsymbol{\beta}}} = \|\widetilde{f}\|_{h_{\boldsymbol{\beta}}},$$

implying that it is sufficient to show the convergence of the coefficients in the sequence space.

Consider the change of variables $\widehat{f} = \{|k^{\beta_*}|f_{k,l}\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}}$. The model (10.26) can be rewritten into,

$$\widehat{X}^{(n)}_{k,l} = \widehat{f}_{k,l} + \frac{1}{\sqrt{n}} z_{k,l}, \quad \text{for } k \in k,l\}_{(k,l)\in\mathbb{N}^d\times\mathbb{N}},$$

where $z_{k,l}$ are independent standard Gaussian random variables. Notice that the prior of $f$ induces a prior of $\widehat{f}$ in $\ell^2(\mathbb{N}^d \times \mathbb{N})$. Therefore, using Lemma 10.13, we obtain the posterior contraction rate of the induced prior in $\|\cdot\|_{\ell^2(d+1)}$.

For the preceding change of variables, an isometry $\|\widehat{f}\|_{\ell^2} = \|\widetilde{f}\|_{h_{\beta_*}}$ holds, given $\widetilde{f}$ is in $h_{\beta_*}$. The isometry implies that rates of $\widehat{f}$ relative to $\|\cdot\|_{\ell^2}$ can be translated to the rates of $\widetilde{f}$ relative to $\|\cdot\|_{h_{\beta_*}}$. Consequently, the result can be translated to the rate in $\|\cdot\|_{H_{\beta_*}}$ and the proof is complete.

## 10.4 Entropy Number with Non-Polynomial Rates

In Section 2.4, we have shown the estimate of metric numbers of the embedding $\iota : H_{s+t} \to H_t$ with the scale. The same argument can be applied to $\iota : \widetilde{H} \to H_t$ where $\widetilde{H}$ is another Hilbert space contained in the smoothness class $\{H_t\}_t$, see the lemma below.

**Lemma 10.14.** *Given $\mathbb{H}_G$ from (10.13) and the isotropic Sobolev spaces $\{H_{\boldsymbol{s}}\}_{s\in\mathbb{R}}$ defined in Section 2.3.1. Then for the canonical embedding $\iota : \mathbb{H}_G \to H_{\boldsymbol{s}}$, when $j$ is large enough, the entropy number is of the order*

$$e^{-c_1 j^{\nu/(\nu+d)}} \leq e_j(\iota : \mathbb{H}_G \to H_s) \leq e^{-c_2 j^{\nu/(\nu+d)}}, \tag{10.30}$$

*where $c_1, c_2$ are universal positive constants.*

*Proof.* The proof follows the same argument from Section 2.4 (also see the Appendix B in [43]). The singular values of $\iota : \mathbb{H}_G \to H_s$ are of order $e^{-Tj^{\nu/d}} j^{-(\alpha-s)/d}$, for which the upper bound is obtained using the same argument in (10.15) and the lower bound is obtained by taking the unit vector $\varphi_k$ such that $k_i \simeq j^{1/d}$. Consequently, the approximation numbers $a_j(\iota : \mathbb{H}_G \to H_s)$ have the same order. In particular, $a_j(\iota : \mathbb{H}_G \to H_s) = O(e^{-Tj^{\nu/d}})$. By the second example in Section 3 of [103], we obtain the final statement of the lemma. □

Because $\varepsilon \mapsto H(\varepsilon, \iota)$ is the inverse mapping of $j \mapsto e_j(\iota)$, we obtain the corollary below.

**Corollary 10.15** (Metric entropy)**.** *The metric entropy of the unit ball of $\mathbb{H}_G$, given in (10.13), in isotropic Sobolev spaces $\{H_{\boldsymbol{s}}\}_{s\in\mathbb{R}}$ is given by*

$$H(\varepsilon, \iota) := \log N\left(\varepsilon, \{g \in \mathbb{H}_G : \|g\|_{\mathbb{H}_G} \leq 1\}, \|\cdot\|_{H_{\boldsymbol{s}}}\right) \sim \left(\log \frac{1}{\varepsilon}\right)^{\frac{\nu+d}{\nu}},$$

*as $\varepsilon \downarrow 0$.*