POSTERIOR CONTRACTION FOR EMPIRICAL BAYESIAN APPROACH TO INVERSE PROBLEMS UNDER NON-DIAGONAL ASSUMPTION

JUNXIONG JIA, JIGEN PENG, AND JINGHUAI GAO

Abstract. We investigate an empirical Bayesian nonparametric approach to a family of linear inverse problems with Gaussian prior and Gaussian noise. We consider a class of Gaussian prior probability measures with covariance operator indexed by a hyperparameter that quantifies regularity. By introducing two auxiliary problems, we construct an empirical Bayes method and prove that this method can automatically select the hyperparameter. In addition, we show that this adaptive Bayes procedure provides optimal contraction rates up to a slowly varying term and an arbitrarily small constant, without knowledge about the regularity index. Our method needs not the prior covariance, noise covariance and forward operator have a common basis in their singular value decomposition, enlarging the application range compared with the existing results.

1. Introduction

Inverse problems for partial differential equations arise naturally from medical imaging, seismic exploration and so on. There are two difficulties for such types of problems: one is non-uniqueness and another one is instability. Bayesian approach formulates inverse problems as statistical inference problems, which enables us to overcome both of these difficulties. Because of that, Bayes' inverse method has attracted a lot of interest in recent years; see for instance [4, 5, 7, 9, 15, 16, 21, 26]. However, compared with the classical regularization techniques, the development of a theory of Bayesian posterior consistency is still in its infancy.

For clarity, let us provide some basic settings. Let $H$ be a separable Hilbert space, with norm $\| \cdot \|$ and inner product $(\cdot, \cdot)$, and let $\mathcal{A} : D(\mathcal{A}) \subset H \to H$ be a self-adjoint and positive-definite linear operator with bounded inverse. Then, we usually assume the following problem

$$y = \mathcal{A}^{-1} u + \frac{1}{\sqrt{n}} \xi,$$

(1.1)

where $\frac{1}{\sqrt{n}} \xi$ is noise and $y$ is a noisy observation of $\mathcal{A}^{-1} u$. The inverse problem can be thought to find $u$ from $y$. Following the framework shown in [8], we denote $\mu^0$ as the prior measure and $\mathbb{P}_{\text{noise}}$ as the noise measure and assume that

- Prior: $u \sim \mu^0$,
- Noise: $\xi \sim \mathbb{P}_{\text{noise}} = \mathcal{N}(0, C_1),$

2010 Mathematics Subject Classification. 74G75, 86A22.

Key words and phrases. Posterior consistency, Empirical Bayesian approach, Linear inverse problem, Non-simultaneous diagonal.
where \( \mathcal{N}(0, C_1) \) is a Gaussian measure with zero mean and covariance operator \( C_1 \). Let us denote \( \mu_n^\alpha \) to be the posterior measure which is the solution of the inverse problem. Usually, algorithms, such as Markov chain Monte Carlo (MCMC), can be employed to probe the posterior probability measure \( \mu \). If we assume the range of \( A^{-1} \) is included in the domain of \( C^{-1/2} \), then (1.1) can be rewritten as follow

\[
d = T u + \frac{1}{\sqrt{n}} \eta,
\]

(1.2)

where \( d = C_1^{-1/2} y, T = C_1^{-1/2} A^{-1} \) and \( \eta = C_1^{-1/2} \xi \). Obviously, we have \( \eta \sim \mathcal{N}(0, I) \) is a white Gaussian noise. We denote \( \mu_n^\alpha \) to be the posterior measure and still denote \( \mu_0 \) to be the prior measure. Here we will be particularly interested in the small noise limit where \( n \to \infty \). Under the frequentist setting, the data \( d = d^\dagger \) are generated by

\[
d^\dagger = T u^\dagger + \frac{1}{\sqrt{n}} \eta, \quad \eta \sim \mathcal{N}(0, I),
\]

(1.3)

where \( u^\dagger \) is a fixed element of \( H \). Our aim is to show that the posterior probability measure \( \mu_n^\alpha \) contracts to a Dirac measure centered on the fixed true solution \( u^\dagger \). In the following, we denote \( L(\eta) \) to be the law of a random variable \( \eta \) and denote \( E_0 \) to be the expectation corresponding to the distribution of the data \( d^\dagger \). A formal description of the concept of posterior consistency can be stated as follow.

**Definition 1.** A sequence of Bayesian inverse problems \( (\mu_0, T, L(\frac{1}{\sqrt{n}} \eta)) \) is posterior consistent for \( u^\dagger \) with rate \( \epsilon_n \downarrow 0 \) if for (1.3), there exists positive sequence \( M_n \) such that

\[
E_0 \left\{ \mu_n^\alpha \left\{ u : \| u - u^\dagger \| \geq M_n \epsilon_n \right\} \right\} \to 0, \quad \forall M_n \to \infty.
\]

(1.4)

Until now, there are a number of studies for posterior consistency and inconsistency for Bayes’ inverse method. When the forward operator \( A^{-1} \) is a nonlinear operator, Vollmer [29] provides a general framework and studies an inverse problem for elliptic equation in detail. Recently, Knapik and Salomond [18] provide a general framework for investigating the posterior consistency of linear inverse problems, which allows non-Gaussian priors. However, due to the difficulties brought by non-linearity and non-Gaussian, the existing results are mainly focus on the situation that the forward operator \( A^{-1} \) is linear, the prior measure is Gaussian \( \mu_0 = \mathcal{N}(0, C) \) and the noise measure is also Gaussian \( P_{\text{noise}} = \mathcal{N}(0, C_1) \). When \( A^{-1}, C \) and \( C_1 \) are all simultaneous diagonalizable, Knapik et al [20] provide a roadmap for what is to be expected regarding posterior consistency. This work reveals an important fact that the optimal convergence rates can be obtained if and only if the regularity of the prior and the truth are matched. Therefore, Knapik et al [19] propose an empirical Bayes procedure which provides an estimate of the regularity of the prior through the data. By choosing the regularity index adaptively, the optimal convergence rates are obtained up to a slowly varying factor. Later, Szabó et al [27] introduce the “polished tail” condition and investigate the frequentist coverage of Bayesian credible sets by choosing the regularity of the prior through an empirical Bayes method. Recently, by employing abstract tools from regularization theory, Agapiou and Mathé [2] study the posterior consistency by choosing a non-centered prior through empirical Bayes method. In 2018, Trabs [28] obtained optimal convergence rates up to logarithmic factors when the forward linear operator depends on an unknown parameter.
When $A^{-1}$, $C$ and $C_1$ are not simultaneous diagonalizable, optimal contraction rates can not be obtained by employing similar methods developed for the simultaneous diagonalizable case. Concerning this case, theories about partial differential equation (PDE) have been employed to obtain nearly optimal convergence rates in [1] by using precision operators. Using properties of the hypoelliptic operators, Kekkonen et al. [17] obtain nearly optimal convergence rates for Bayesian inversion with hypoelliptic operators. Recently, Mathé [22] studies the posterior consistency for non-commuting operators by employing the ideas of link conditions originating from classical regularization method in Hilbert scales. Besides these studies, to the best of our knowledge, there are little investigations of the posterior consistency for the non-simultaneous diagonalizable case. In order to achieve the optimal contraction rates, the regularity of the true solution should be known in the above mentioned investigations. Hence, how to generalize the empirical Bayes procedure developed for the simpler simultaneous diagonalizable case to the non-diagonal case is crucial for the applicability of the posterior consistent theory.

In [19, 27], the structures of the posterior probability measure for the regularity index are fully analyzed. For the non-diagonal case, the posterior probability measure for the regularity index is hard to define. The reason is that probability density functions can not be defined in infinite-dimensional space and some integrals can not be calculated easily to obtain the log-likelihood of the regularity index obtained in [19]. Hence, we can hardly define an empirical Bayes procedure for the non-diagonal case. To overcome these difficulties, we notice that regularity index is derived from data and the posterior contraction rates are only depending on the regularity index of the prior measure and the true function $u^\dagger$. So, if we introduce some artificial diagonal problem with the same regularity properties as for the non-diagonal problem, we may obtain optimal posterior contraction rates relying on the relations between the artificial diagonal problem and the non-diagonal problem.

For a positive constant $\alpha$, we consider Gaussian prior measure $\mu^0 = \mathcal{N}(0, C^\alpha)$. Firstly, we consider the following problem

$$d = m + \frac{1}{\sqrt{n}}\eta,$$

where $m := T u$. Then the prior measure for $m$ obviously is $\mathcal{N}(0, TC^\alpha T^*)$. Concerning this problem, the forward operator and the covariance operator of the noise are all the identity operator $I$. Hence, the posterior contraction rates of the problem (1.5) seems can be obtained by the ideas developed for the diagonal case. However, the operator $T$ appears in the prior probability measure, which makes difficult to derive estimations of the corresponding eigenvalues. By constructing an artificial prior probability measure with similar regularity properties as for the prior probability measure, we can construct maximum likelihood-based estimate for the regularity index and prove the optimal contraction rates for (1.5) up to a slowly varying term and an arbitrarily small constant. At last, we transform the results for problem (1.5) to the original problem (1.2) which can be achieved by employing the method developed in [18].

The outline of this paper is as follows. In Section 2 we provide a brief introduction about Hilbert scales and give some essential assumptions about the covariance operators of the prior and noise probability measure. In Section 3 we exhibit two auxiliary problems. Relying on detailed analysis about the two auxiliary problems,
we prove the posterior consistency of our original problem. In Section 4 two examples are given, which illustrate the usefulness of our results. In Section 5 we provide a brief summary and outlook. At last, the proofs of some auxiliary lemmas and properties are collected in Section 6.

2. Basic settings and Assumptions

In this section we present basic settings and show some important assumptions made in this paper. For the reader’s convenience, let us provide an explanation for some frequently used notations firstly.

Notations:
- The set of all bounded linear operators mapping from some Hilbert space $H$ to $H$ is denoted by $\mathcal{B}(H)$, and the corresponding operator norm is denoted by $\| \cdot \|_{\mathcal{B}(H)}$.
- The range of some operator $C$ is denoted by $\mathcal{R}(C)$, and the domain of the operator $C$ is denoted by $\mathcal{D}(C)$.
- For a linear operator $T$, its dual operator is denoted by $T^*$.
- The notation $E_0$ stands for the expectation corresponding to the distribution of the data $d^\dagger$ generated from the truth.
- For two sequences $a_n$ and $b_n$ of numbers, we denote by $a_n \preceq b_n$ ($a_n \succeq b_n$) that there are $M \in \mathbb{R}$ such that $a_n \leq Mb_n$ ($a_n \geq Mb_n$) for $n$ large enough. If $a_n \preceq b_n \preceq a_n$, we write $a_n \asymp b_n$.

2.1. Preliminaries. Firstly, we will present a brief introduction about Hilbert scales [10] which provides powerful tools for measuring the smoothness of the noise, the forward operator and the samples of the prior. Let $C : H \rightarrow H$ be a self-adjoint, positive-definite, trace class, linear operator with eigensystem $(\lambda_i^2, \phi_i)_{i=1}^{\infty}$.

Considering $H = \mathcal{R}(C) \oplus \mathcal{R}(C)^\perp = \overline{\mathcal{R}(C)}$, we know that $C^{-1}$ is a densely defined, unbounded, symmetric and positive-definite operator in $H$. Denote $\| \cdot \|$ to be the norm defined on the Hilbert space $H$. We define the Hilbert scales $(\mathcal{H}^t)_{t \in \mathbb{R}}$, with $\mathcal{H}^t := \overline{\bigcap_{n=0}^{\infty} \mathcal{D}(C^{-n})}$, $(u, v)_{\mathcal{H}^t} := (C^{-t/2}u, C^{-t/2}v)$, $\| u \|_{\mathcal{H}^t} := \| C^{-t/2}u \|$. In addition, the norms defined above possess the following properties.

Lemma 2.1. (Proposition 8.19 in [10]) Let $(\mathcal{H}^t)_{t \in \mathbb{R}}$ be the Hilbert scale induced by the operator $C$ defined above. Then the following assertions hold:

(1) Let $-\infty < s < t < \infty$. Then the space $\mathcal{H}^t$ is densely and continuously embedded in $\mathcal{H}^s$.
(2) If $t \geq 0$, then $\mathcal{H}^t = \mathcal{D}(C^{-t/2})$ and $\mathcal{H}^{-t}$ is the dual space of $\mathcal{H}^t$.
(3) Let $-\infty < q < r < s < \infty$ then the following interpolation inequality holds

$$\| u \|_{\mathcal{H}^r} \leq \| u \|_{\mathcal{H}^q} \| u \|_{\mathcal{H}^s},$$

where $u \in \mathcal{H}^s$.

Next, let us provide some necessary notations about infinite sequences. For a sequence $m_s = \{m_i\}_{i=1}^{\infty}$, we denote the $\ell^2$-norm by $\| m_s \|_0$, that is, $\| m_s \|_0^2 = \sum_{i=1}^{\infty} m_i^2$. In addition, the norms defined above possess the following properties.
\[ \sum_{i=1}^{\infty} m_i^2. \]  

The hyperrectangle and Sobolev space of sequence of order \( \beta > 0 \) and radius \( R > 0 \) are the sets

\[ \Theta^\beta (R) = \left\{ m_s \in \ell^2 : \sup_{i \geq 1} \frac{1}{i^{1+2\beta}} m_i^2 \leq R \right\}, \]

\[ S^\beta (R) = \left\{ m_s \in \ell^2 : \sum_{i=1}^{\infty} \frac{1}{i^{2\beta}} m_i^2 \leq R \right\}. \]

For a sequence \( m_s = \{m_i\}_{i=1}^{\infty} \), the hyperrectangle norm and Sobolev norm can be defined by

\[ \|m_s\|_{h,\beta}^2 = \sup_{i \geq 1} \frac{1}{i^{1+2\beta}} m_i^2, \quad \|m_s\|_{\beta}^2 = \sum_{i=1}^{\infty} \frac{1}{i^{2\beta}} m_i^2. \]

**Definition 2.2.** A sequence \( m_s = \{m_i\}_{i=1}^{\infty} \in \Theta^\beta (R) \) is self-similar if, for some fixed positive constants \( \epsilon, N_0 \) and \( \rho \geq 2 \),

\[ \sum_{i=N}^{\rho N} m_i^2 \geq \epsilon MN^{-2\beta}, \quad \forall N \geq N_0. \]

The class of self-similar elements of \( \Theta^\beta (R) \) are denoted by \( \Theta_{\text{ss}}^\beta (R) \). The parameters \( N_0 \) and \( \rho \) are fixed and omitted from the notation.

This definition is employed by Giné and Nickl [13] and Bull [3] to remove a “small” set of undesirable truths from the model, which allows to generate candidate confidence sets for the true parameter that are routinely used in practice. In [27], the authors propose the polished tail condition which includes the set of self-similar sequences. The set of self-similar sequences has been shown natural once one has adopted the Bayesian setup with variable regularity prior probability measures.

### 2.2. Assumptions

In this section, we give the main assumptions employed in our work. We assume that the prior probability measure and the probability measure of the noise \( \xi \) have the following form

\[ \mu^0 = \mathcal{N}(0, C^\alpha), \quad P_{\text{noise}} = \mathcal{N}(0, C_1); \]

where \( C : H \to H \) is a self-adjoint, positive-definite, trace class, linear operator and \( C_1 : H \to H \) is assumed to be a self-adjoint, positive-definite linear operator. The operator \( A : D(A) \to H \) is assumed to be a self-adjoint and positive-definite, linear operator with bounded inverse, \( A^{-1} : H \to H \).

Let us firstly present the following assumptions which describe the relations between Hilbert space and the space of sequences.

**Assumptions 1:** The covariance operator \( \mathcal{C} \) has eigenpairs \( \{\lambda_i^2, \phi_i\}_{i=1}^{\infty} \) on Hilbert space \( H \). For the singular values \( \{\lambda_i\}_{i=1}^{\infty} \), there exists a positive constant \( p > 0 \) such that

\[ \lambda_i^2 \asymp i^{-\frac{d}{2p}}, \quad \forall i = 1, 2, \ldots. \]

The formula (2.5) means that there exist a series of constants \( \{c_i\}_{i=1}^{\infty} \) and a positive constant \( C > 0 \) such that \( \lambda_i^2 \leq C \frac{1}{c_i} i^{-2p/d} \). Let \( \alpha_0 := \frac{d}{2p} \), for all
\[ \alpha > \alpha_0, \text{ we have} \]
\[ (2.6) \quad \sum_{i=1}^{\infty} i^{-\frac{2\alpha}{m^2}} < \infty. \]

Inequality (2.6) reflects that the operator \( C^{\alpha} \) is a trace class operator for \( \alpha > \alpha_0 \).

Secondly, we give the assumptions which are mainly concerned with the inter-
relations between the three operators \( C, C_1, \) and \( A^{-1} \). Similar to the assumptions
presented in [1], these assumptions reflect the ideas that
\[ (2.7) \quad C_1 \simeq C^\beta, \quad A^{-1} \simeq C^\ell, \]
for some \( \beta \geq 0, \ell \geq 0 \), where \( \simeq \) is used loosely to indicate two operators which
induce equivalent norms. Specifically speaking, we make the following assumptions.

**Assumptions 2:** Let us denote \( T := C_1^{-1/2} A^{-1} \) and \( M(\alpha) := TC^{\alpha} T^* \). Suppose
there exist \( \beta \geq 0, \ell \geq 0, \epsilon > 0 \) and for all \( \alpha > \alpha_0 \) such that
(1) \( \Delta \geq 1 \), where \( \Delta := 2\ell - \beta + 1 \);
(2) \( \| Tu \|_{H^r} \simeq \| C^{\frac{1}{2}(\Delta-1)} u \|_{H^r}, \quad \forall u \in H^{r-(\Delta-1)}, \quad r \geq 0; \)
(3) \( \| C^{\frac{1}{2}} T^* u \| \lesssim \| u \|_{H^{-\alpha-(\Delta-1)}}, \quad \forall u \in H^{-\alpha-(\Delta-1)}; \)
(4) \( \| M(\alpha) u \|_{H^{(\alpha+\Delta-1)}} \lesssim \| u \|_{H^\epsilon}, \quad \forall u \in H; \)
(5) \( \| C^{\frac{1}{2}(\alpha+\Delta-1)} M(\alpha) \frac{1}{2} C^{-\frac{1}{2}} \|_{B(H)} < \infty; \)
(6) \( \| M(\alpha)^2 C^{-x(\alpha+\Delta-1)} \|_{B(H)} < \infty, \quad \forall x \in [-\frac{1}{2}, 1]; \)
(7) \( \| M(\alpha) C^{-(\alpha+\Delta-1)} \|_{B(H)} < \infty. \)

At last, we assume the truth \( u^\dagger \) belongs to \( H^\gamma \) with \( \gamma \geq 1 \). Denote \( u^\dagger_i := (u^\dagger, \phi_i) \)
for \( i = 1, 2, \ldots, \) and \( u^\dagger_s = \{ u^\dagger_i \}_{i=1}^{\infty} \). Since
\[ (2.8) \quad \| u^\dagger \|_{H^\gamma}^2 = \sum_{i=1}^{\infty} \lambda_i^{-2\gamma} (u^\dagger_i)^2 \geq \sum_{i=1}^{\infty} i^{2\gamma} (u^\dagger_i)^2, \]
we know that \( u^\dagger_s \in S^{2\gamma}(R) \) for some \( R > 0 \). In addition, we easily find that
\( u^\dagger_s \in \Theta^{2\gamma}(R') \) for some \( R' > 0 \). Concerning the truth, we make the following
assumption which is crucial for our estimations.

**Assumption 3:** We assume that \( u^\dagger \in H^\gamma \) for some \( \gamma \geq 1 \), and the sequence \( m^\dagger_s \)
induced by \( m^\dagger = T u^\dagger \) belongs to \( \Theta^{2\gamma}(R) \) for some \( R > 0 \) with \( \beta = \frac{\ell}{2} (\gamma + \Delta - 1). \)

3. Posterior contraction

In this section, we will present our main results and show the proof details. Due
to the proofs are quite technical, it is worth to give a sketch of our main ideas
firstly.

As stated in the introduction, we will introduce two auxiliary problems which
are crucial for our analysis. Hence, it is necessary to give a clear summarization of
the three problems employed in our work.

**Original Problem:** Under the assumption \( R(A^{-1}) \subset D(C_1^{-1/2}) \), we summarize
the essential elements of the original inverse problem as follow
- **Forward operator:** \( T = C_1^{-1/2} A^{-1}, \)
• Data: \( d = T u + \frac{1}{\sqrt{n}} \eta \).
• Prior probability measure: \( \mu^0 = \mathcal{N}(0, C^\alpha) \) with \( \alpha > \alpha_0 \).
• Noise probability measure: \( P_{\text{noise}} = \mathcal{N}(0, I) \).

Since the operator \( T \) and \( C^\alpha \) cannot be diagonalized simultaneously, we introduce \( m = Tu \) to transform the forward operator to be the identity operator and propose the following transformed problem.

**Transformed Problem:** For the transformed inverse problem, the necessary elements can be summarized as follow

• Forward operator: \( I \),
• Data: \( d = m + \frac{1}{\sqrt{n}} \eta \),
• Prior probability measure: \( \mu^m_0 = \mathcal{N}(0, TC^\alpha T^*) \) with \( \alpha > \alpha_0 \),
• Noise probability measure: \( P_{\text{noise}} = \mathcal{N}(0, I) \).

The covariance operator \( TC^\alpha T^* \) of the transformed problem cannot be diagonalized under the eigenbasis presented in Assumptions 1. So we can hardly obtain useful estimations of the corresponding eigenvalues, which inspired us to introduce the following artificial diagonal problem.

**Artificial Diagonal Problem:** For the artificial diagonal problem, the essential elements can be summarized as follow

• Forward operator: \( I \),
• Data: \( d = m + \frac{1}{\sqrt{n}} \eta \),
• Prior probability measure: \( \mu^m_0 = \mathcal{N}(0, C^{\Delta-1+\alpha}) \) with \( \alpha > \alpha_0 \),
• Noise probability measure: \( P_{\text{noise}} = \mathcal{N}(0, I) \).

We will recast the artificial diagonal problem into the framework introduced in [19, 27] and prove a corresponding posterior consistency result. Then relations of the artificial diagonal problem and the transformed problem will be explored to provide a posterior contraction estimation for the transformed problem. At last, the general approach developed in [18] will be employed to transfer the posterior contraction estimation to our original problem.

### 3.1. Posterior contraction for the artificial diagonal problem.

Denote \( m_i := (m, \phi_i) \). Relying on Assumptions 1, we easily know that

\[
(3.1) \quad m_i \sim \mathcal{N}(0, \lambda_i^2(\Delta^{-1+\alpha})) = \mathcal{N}(0, c_i^2(\Delta^{-1+\alpha}) \frac{1}{\sqrt{n}} \eta_i).
\]

Denote \( \tilde{m}_i = c_i^{\Delta-1+\alpha} m_i \), \( d_i := (d, \phi_i) \) and \( \eta_i := (\eta, \phi_i) \), equality (1.5) can be rewritten as follow

\[
(3.2) \quad d_i = c_i^{\Delta-1+\alpha} \tilde{m}_i + \frac{1}{\sqrt{n}} \eta_i, \quad \text{for } i = 1, 2, \ldots
\]

with \( \tilde{m}_i \sim \mathcal{N}(0, i^{-\frac{2\alpha}{p}}(\Delta^{-1+\alpha})) \) and \( \eta_i \sim \mathcal{N}(0, 1) \). For convenience, let us define

\[
(3.3) \quad \tilde{\alpha} := \frac{p}{d}(\Delta - 1 + \alpha) - \frac{1}{2}.
\]

Since \( \Delta \geq 1 \) and \( \alpha > \alpha_0 \), we find that \( \tilde{\alpha} \geq 0 \). From the formula (3.3), we easily know that \( i^{-2\tilde{\alpha}} = i^{-\frac{2p}{d}(\Delta - 1 + \alpha)} \). Following the formula (2.2) employed in [19], we introduce log-likelihood for \( \tilde{\alpha} \) (relative to an infinite product of \( \mathcal{N}(0, 1/n) \)-distribution)
as follow
\begin{equation}
\ell_n(\tilde{\alpha}) = -\frac{1}{2} \sum_{i=1}^{\infty} \left( \log \left( 1 + \frac{n}{i^{1+2\tilde{\alpha}}} \right) - \frac{n^2}{i^{1+2\tilde{\alpha}} + 1} \right).
\end{equation}

For the artificial diagonal problem, we denote \( m^\dagger \) to be the truth which is equal to \( T u^\dagger \), and denote \( m_i^\dagger = \{(m_i^\dagger, \phi_i)\}_{i=1}^{\infty} \) and \( \tilde{m}_i^\dagger = (\tilde{m}_i^\dagger, \phi_i) \) with \( i = 1, 2, \ldots \). For convenience, let us define
\begin{equation}
\tilde{m}_i^\dagger := \sum_{i=1}^{\infty} \tilde{m}_i^\dagger \phi_i.
\end{equation}
Similarly, for problem (3.2), we denote \( m_i^\dagger = \{m_i^\dagger\}_{i=1}^{\infty} \) to be the truth. Considering \( m^\dagger = T u^\dagger \) and
\begin{equation}
\|m^\dagger\|_{H^{\gamma+\Delta-1}} = \|Tu^\dagger\|_{H^{\gamma+\Delta-1}} \asymp \|G^{(\Delta-1)}u^\dagger\|_{H^{\gamma+\Delta-1}} = \|u^\dagger\|_{H^\gamma} < \infty,
\end{equation}
we find that \( m^\dagger \in H^{\gamma+\Delta-1} \). Relying on Assumptions 1, we have
\begin{align*}
\|m^\dagger\|_{H^{\gamma+\Delta-1}}^2 &= \sum_{i=1}^{\infty} \lambda_i^{-2(\gamma+\Delta-1)} (m_i^\dagger)^2 \\
&= \sum_{i=1}^{\infty} c_i^{-2(\gamma+\Delta-1)} \frac{2^{2p}(\gamma+\Delta-1)}{2^{2p}} (m_i^\dagger)^2.
\end{align*}
The above equality indicates that
\begin{equation}
\|m_i^\dagger\|_{\tilde{\beta}} \lesssim \|m^\dagger\|_{H^{\gamma+\Delta-1}} \lesssim \|m_i^\dagger\|_{\tilde{\beta}},
\end{equation}
where \( \tilde{\beta} = \tilde{\beta}(\gamma + \Delta - 1) \). Through similar deductions, we can also find that
\begin{equation}
\|\tilde{m}_i^\dagger\|_{\tilde{\beta}} \lesssim \|\tilde{m}^\dagger\|_{H^{\gamma+\Delta-1}} \lesssim \|\tilde{m}_i^\dagger\|_{\tilde{\beta}}.
\end{equation}

With these preparations, we can define
\begin{equation}
h_n(\tilde{\alpha}) = \frac{1 + 2\tilde{\alpha}}{n^{1/2(1+2\tilde{\alpha})} \log n} \sum_{i=1}^{\infty} \frac{n^2 i^{1+2\tilde{\alpha}} (\log i) (\tilde{m}_i^\dagger)^2}{(i^{1+2\tilde{\alpha}} + 1)^2},
\end{equation}
which is similar to the formula (5.1) defined in [27]. In addition, we define
\begin{align*}
\tilde{\alpha}_n &= \inf \{ \tilde{\alpha} > 0 : h_n(\tilde{\alpha}) > \ell \} \land \sqrt{\log n}, \\
\overline{\alpha}_n &= \inf \{ \tilde{\alpha} > 0 : h_n(\tilde{\alpha}) > L (\log n)^2 \},
\end{align*}
where \( 0 < \ell < L < \infty \) and the infimum of the empty set is considered \( \infty \). Before going further, let us present some estimates about \( \tilde{\alpha}_n \) and \( \overline{\alpha}_n \).

Lemma 3.1. For any constant \( R > 0 \), there exist \( C_0 \) and \( C_1 \) such that
\begin{align*}
\inf_{\tilde{m}_i^\dagger \in S^\beta(R)} \tilde{\alpha}_n &\geq \tilde{\beta} - C_0/\log n, \\
\sup_{\tilde{m}_i^\dagger \in \Theta_\alpha^\beta(R) \cap S^\beta(R)} \overline{\alpha}_n &\leq \tilde{\beta} + C_1 (\log \log n)/\log n,
\end{align*}
for \( n \) large enough.
The proof details are shown in the Appendix. Then we define
\[
\hat{\alpha}_n := \arg\max_{\tilde{\alpha} \in [0, \log n]} \ell_n(\tilde{\alpha}) - \frac{C_1 \log \log n}{\log n},
\]
where \(C_1\) is a positive constant appeared in Lemma 3.1. Through some small modifications of the proof of Theorem 1 presented in [10], we have
\[
\inf_{\tilde{m}_n \in S^\beta(R) \cap \Theta^\beta_\alpha(R)} \mathbb{P}_0 \left( \tilde{\alpha}_n - C_1 \frac{\log \log n}{\log n} \leq \hat{\alpha}_n \leq \tilde{\alpha}_n - C_1 \frac{\log \log n}{\log n} \right) \to 1.
\]
(3.14)

Obviously, when \(\hat{\alpha}_n\) restricted to the following interval
\[
I_n := \left[ \tilde{\alpha}_n - C_1 \frac{\log \log n}{\log n}, \tilde{\alpha}_n - C_1 \frac{\log \log n}{\log n} \right],
\]
we have
\[
\hat{\alpha}_n \leq \tilde{\beta} \quad \text{and} \quad \hat{\alpha}_n \geq \tilde{\beta} - \frac{C_2 \log \log n}{\log n},
\]
where \(C_2 = C_0 + C_1\). The notation \(C_2\) will be used in all of the sections below.

Considering formula (3.3), we define
\[
\hat{\alpha}_n := d \left( \frac{\hat{\alpha}_n + 1}{2} \right) + 1 - \Delta,
\]
which is the estimation for the regularity index of the prior \(N(0, C_\alpha)\) of \(u\). For \(\hat{\alpha}_n \in I_n\), we easily deduce that
\[
\hat{\alpha}_n \leq \gamma + \frac{d}{2p}, \quad \text{and} \quad \hat{\alpha}_n \geq \gamma + \frac{d}{2p} - \frac{dC_2 \log \log n}{p \log n}.
\]
(3.16)

Now, we provide the following theorem which gives the convergence rates estimation for our artificial diagonal problem.

**Theorem 3.2.** Let \(\hat{m}_{dn}(\tilde{\alpha})\) be the posterior mean estimator for our artificial diagonal problem when the regularity index is \(\tilde{\alpha}\), then we have
\[
\sup_{m_n \in S^\beta(R) \cap \Theta^\beta_\alpha(R)} \mathbb{E}_0 \left\{ \sup_{\tilde{\alpha} \in I_n} \left\| \hat{m}_{dn}(\tilde{\alpha}) - m_n \right\|^2 \right\} = O(\epsilon^2_n),
\]
where \(\epsilon_n = L_n n^{-\beta/(1+2\beta)}\) and \(L_n = (\log n)^{C_2 \alpha} (\log \log n)^{1/2}\).

Since the proof is not the main ingredient of our work, it has been postponed to the Appendix.

### 3.2. Posterior contraction for the transformed problem

Denote \(\mathcal{M}(\alpha) = TC^\alpha T^*\) for \(\alpha > \alpha_0\). Now let us come back to the following transformed problem
\[
d = m + \frac{1}{\sqrt{n}} \eta,
\]
(3.20)

where
\[
\eta \sim N(0, I) \quad \text{and} \quad m \sim \mu^{m_0} = N(0, \mathcal{M}(\hat{\alpha}_n))
\]
with \(m = Tu\). Before diving into the discussions on posterior contraction, let us provide some important estimates. Consider the equation
\[
(n \mathcal{M}(\alpha) + I)u = r,
\]
(3.22)
Define the bilinear form $B : H \times H \to \mathbb{R}$,
\[ B(u, v) := (n^{1/2}C^{-1/2}T^*u, n^{1/2}C^{-1/2}T^*v) + (u, v), \quad \forall u, v \in H. \]

**Definition 3.3.** Let $r \in H$. An element $u \in H$ is called a weak solution of (3.22), if
\[ B(u, v) = (r, v), \quad \forall v \in H. \]

**Lemma 3.4.** Under the Assumptions 2, for any $r \in H$, there exists a unique weak solution $u \in H$ of (3.22). In addition, if $r \in H^t$ with $t \leq 2(\alpha + \Delta - 1)$, the weak solution $u \in H^t$.

Since the proof is simple, we postpone the proof to the Appendix. Relying on the definition of weak solution and interpolation inequalities, we can prove the following lemma.

**Lemma 3.5.** For any constant $s \in (\alpha_0, \alpha)$, under the Assumptions 2, the following norm bound holds:
\[ \|C^{\frac{1}{2}(\alpha+\Delta-1) - \frac{s}{2}} (\alpha, \alpha) M_{\alpha} + I)^{-1} C^{\frac{1}{2}(\alpha+\Delta-1) - \frac{s}{2}} \|_{B(H)} \lesssim n^{1+\frac{s}{2(\alpha+\Delta-1)}}. \]

For fluency of the description, we defer the proof to the Appendix. Denote $\mu_{\alpha_n}^d$ to be the posterior probability measure for our original problem and $\mu_{\alpha_n}^{d, \alpha}$ to be the posterior probability measure for the transformed problem. The two posterior probability measures are all Gaussian due to our assumptions. Using the relations between $m$ and $u$, we know that
\[ \mu_{\alpha_n}^{d, \alpha} \{ m : \|m - m^\dagger\| \geq M_n \varepsilon_n \} = \mu_{\alpha_n}^d \{ u : \|Tu - Tu^\dagger\| \geq M_n \varepsilon_n \}, \]
where $M_n$ and $\varepsilon_n$ are positive sequences satisfy $M_n \to \infty$ and $\varepsilon_n \downarrow 0$, respectively. Let us denote $E_{\alpha_n}$ and $\hat{m}_n (\alpha)$ to be the expectation and the conditional mean estimator with respect to the posterior probability measure $\mu_{\alpha_n}^d$, respectively. By Markov's inequality, we find that
\[ \sup_{m^\dagger \in S^\beta (R) \cap \Theta_{\alpha_n}^d (R)} E_0 \sup_{\alpha \in I_n} \sum_{\varepsilon_n} \sup_{E_{\alpha_n}} \left( \frac{1}{M_n \varepsilon_n^2} \right) \sup_{m \in S^\beta (R) \cap \Theta_{\alpha_n}^d (R)} \left( \|Tu - Tu^\dagger\|^2 \right) + o(1). \]

Denote $E_{\alpha,n}$ to be the expectation with respect to a probability measure $\mu_{\alpha,n}^d$ with zero mean and the same covariance operator as the posterior probability measure $\mu_{\alpha}^d$ of the original problem. Let $w_n := m - \hat{m}_n (\alpha)$, then we know that $w_n \sim \mu_{\alpha,n}^d$. Obviously, $w_n$ and $\hat{m}_n (\alpha)$ are independent. Since
\[ E_{\alpha,n} \|m - m^\dagger\|^2 = E_{\alpha,n} \|w_n + \hat{m}_n (\alpha) - m^\dagger\|^2, \]
we can insert the posterior mean estimator of our artificial diagonal problem into the right hand side of (3.22). Relying on Assumptions 2 and some calculations, we can prove the following theorem.

**Theorem 3.6.** For every positive constant $R > 0$ and an arbitrarily small positive constant $\tilde{c} > 0$, we have
\[ \sup_{m^\dagger \in S^\beta (R) \cap \Theta_{\alpha_n}^d (R)} E_0 \mu_{\alpha_n}^d \{ u : \|Tu - Tu^\dagger\| \geq M_n \varepsilon_n \} \to 0, \quad \text{as } n \to \infty, \]
where $\varepsilon_n = L_n n^{-\beta/(1+2\beta+\tilde{c})}$ and $L_n = (\log n)^{C_2+2} (\log \log n)^{1/2}$. 

Proof. Denote $F_{dn}(\alpha) := C^{\alpha + \Delta - 1} + n^{-1}I$, then the conditional mean estimator for the artificial diagonal problem has the following form

\begin{equation}
\hat{\mu}_{dn}(\alpha) = m^\dagger - \frac{1}{n} F_{dn}(\alpha)^{-1} m^\dagger + \frac{1}{\sqrt{n}} C^{\alpha + \Delta - 1} F_{dn}(\alpha)^{-1} \eta.
\end{equation}

For the non-diagonal problem with $m \sim \mathcal{N}(0, \mathcal{M}(\alpha))$, the conditional mean estimator can be written as follow

\begin{equation}
\hat{\mu}_{n}(\alpha) = m^\dagger - \frac{1}{n} F_n(\alpha)^{-1} m^\dagger + \frac{1}{\sqrt{n}} \mathcal{M}(\alpha) F_n(\alpha)^{-1} \eta,
\end{equation}

where $F_n(\alpha) := \mathcal{M}(\alpha) + n^{-1}I$. Following (3.28) and recalling that $w_n$ and $\hat{\mu}_{n}(\alpha)$ are independent, we have

\begin{equation}
E_0 \sup_{\tilde{\alpha} \in I_n} \mathbb{E}_{\alpha,n} \| m - m^\dagger \|^2 \leq E_0 \sup_{\tilde{\alpha} \in I_n} \mathbb{E}_{c,\alpha} \| w_n \|^2 + 2E_0 \sup_{\tilde{\alpha} \in I_n} \| \hat{\mu}_{n}(\alpha) - \hat{\mu}_{dn}(\alpha) \|^2 + 2E_0 \sup_{\tilde{\alpha} \in I_n} \| \hat{\mu}_{dn}(\alpha) - m^\dagger \|^2 = I + II + III.
\end{equation}

For the term III, Theorem 3.2 gives an appropriate estimation. For the term I, let us denote the covariance operator as $\mathcal{C}_n^\alpha$ which can be written as follow

\begin{equation}
\mathcal{C}_n^\alpha = (n\mathcal{M}(\alpha) + I)^{-1} \mathcal{M}(\alpha).
\end{equation}

Considering $\Delta \geq 1$, both of $\mathcal{T}$ and $\mathcal{T}^*$ are bounded linear operators. Because of $\mathcal{C}^\alpha$ is a trace class operator, we know that the operator $\mathcal{M}(\alpha)$ is a compact operator and can be diagonalized, which implies that $(n\mathcal{M}(\alpha) + I)^{-1} \mathcal{M}(\alpha) = \mathcal{M}(\alpha)^{1/2}(n\mathcal{M}(\alpha) + I)^{-1} \mathcal{M}(\alpha)^{1/2}$. Then for a white noise $\zeta$, we have the following estimate

\begin{equation}
I = \sup_{\tilde{\alpha} \in I_n} \text{Tr}(\mathcal{C}_n^\alpha) = \sup_{\tilde{\alpha} \in I_n} \text{Tr}(\mathcal{M}(\alpha)^{1/2}(n\mathcal{M}(\alpha) + I)^{-1} \mathcal{M}(\alpha)^{1/2})
\end{equation}

\begin{equation}
= \sup_{\tilde{\alpha} \in I_n} \mathbb{E} \left\| \mathcal{M}(\alpha)^{1/2}(n\mathcal{M}(\alpha) + I)^{-1} \mathcal{M}(\alpha)^{1/2} \zeta \right\|^2
\end{equation}

\begin{equation}
= \sup_{\tilde{\alpha} \in I_n} \mathbb{E} \left( \mathcal{C}^{s/2} \mathcal{C}^{-s/2} \mathcal{M}(\alpha)^{1/2}(n\mathcal{M}(\alpha) + I)^{-1} \mathcal{M}(\alpha)^{1/2} \mathcal{C}^{-s/2} \mathcal{C}^{s/2} \zeta \right)
\end{equation}

\begin{equation}
\leq \sup_{\tilde{\alpha} \in I_n} \left\| \mathcal{C}^{-s/2} \mathcal{M}(\alpha)^{1/2}(n\mathcal{M}(\alpha) + I)^{-1} \mathcal{M}(\alpha)^{1/2} \mathcal{C}^{-s/2} \right\|_{\mathcal{B}(H)} \mathbb{E} \left\| \mathcal{C}^{s/2} \zeta \right\|^2.
\end{equation}

Choosing $s > \alpha_0$ in the above estimate, we know that $\mathbb{E} \| \mathcal{C}^{s/2} \zeta \|^2 < \infty$ by Lemma 3.3 in [1]. Through the statement (5) of Assumptions 2, we have

\begin{equation}
\left\| \mathcal{C}^{-\frac{\alpha}{2}} \mathcal{M}(\alpha)^{\frac{\alpha}{2}} \mathcal{M}(\alpha)^{-\frac{\alpha}{2}} \right\|_{\mathcal{B}(H)} < \infty.
\end{equation}

Combining estimates (3.32), (3.33) and Lemma 3.5, we find that

\begin{equation}
I \lesssim \sup_{\tilde{\alpha} \in I_n} n^{-1 + s + \lambda - 1},
\end{equation}

where $\tilde{\alpha}$ are related to $\alpha$ by (3.3). Taking $n$ large enough, we easily get

\begin{equation}
1 + \frac{\tilde{\epsilon}2\beta}{1 + \tilde{\epsilon}2\beta} = \frac{1(1 + \tilde{\epsilon}C_2) \log \log n}{1 + 2\beta + \tilde{\epsilon} \log n} > 1.
\end{equation}
From the above estimate (3.35) and the lower bound estimate (3.18) of $\alpha$, we know that

$$-1 + \frac{\alpha_0}{\alpha + \Delta - 1} < -\frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}}.$$  

Hence, we can choose appropriate $s$ in (3.33) to obtain

(3.36) \[ I \lesssim n^{-\frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}}} \]

Now let us focus on the term II. Obviously, we have

$$\hat{m}_n(\alpha) - \hat{m}_d_n(\alpha) = \frac{1}{n} \left( F_{dn}(\alpha)^{-1} - F_n(\alpha)^{-1} \right) m^I + \frac{1}{\sqrt{n}} \left( M(\alpha) F_n(\alpha)^{-1} - C^{\alpha + \Delta - 1} F_{dn}(\alpha)^{-1} \right) \eta$$

$$= II_1 + II_2.$$

Concerning the term $II_1$, we find that

$$\left\| II_1 \right\| = \frac{1}{n} \left\| (M(\alpha) + n^{-1} I)^{-1} \left( C^{\alpha + \Delta - 1} - M(\alpha) \right) (C^{\alpha + \Delta - 1} + n^{-1} I)^{-1} m^I \right\|$$

$$\lesssim \frac{1}{n} \left\| (M(\alpha) + n^{-1} I)^{-1} C^{\alpha + \Delta - 1} (C^{\alpha + \Delta - 1} + n^{-1} I)^{-1} m^I \right\| + \frac{1}{n} \left\| (M(\alpha) + n^{-1} I)^{-1} M(\alpha) (C^{\alpha + \Delta - 1} + n^{-1} I)^{-1} m^I \right\|$$

$$= II_{11} + II_{12}.$$

The term $II_{11}$ can be estimated as follow

$$II_{11} \lesssim n^{-\frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}}} \left\| M(\alpha)^{-1} C^{\alpha + \Delta - 1} (C^{\alpha + \Delta - 1} + n^{-1} I)^{-1} m^I \right\|$$

$$\lesssim n^{-\frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}}} \left\| M(\alpha)^{-1} C^{\alpha + \Delta - 1} \right\|_{\mathcal{B}(H)} \left\| m^I \right\|_{H^{(\alpha + \Delta - 1) \frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}}}}$$

where the statement (6) of Assumptions 2 has been used for the last inequality.

Since $\alpha < \gamma + \frac{d}{2\beta}$, we have $(\alpha + \Delta - 1) \frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}} \leq \gamma + \Delta - 1$, which implies

(3.37) \[ \left\| m^I \right\|_{H^{(\alpha + \Delta - 1) \frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}}}} < \infty. \]

Hence, we obtain

(3.38) \[ II_{11} \lesssim n^{-\frac{\beta}{1 + 2\beta + \hat{\epsilon}}}. \]

Similarly, the term $II_{12}$ can be estimated as follow

$$II_{12} \lesssim n^{-\frac{\beta}{1 + 2\beta + \hat{\epsilon}}} \left\| M(\alpha)^{-1} C^{\alpha + \Delta - 1} (C^{\alpha + \Delta - 1} + n^{-1} I)^{-1} m^I \right\|$$

$$\lesssim n^{-\frac{\beta}{1 + 2\beta + \hat{\epsilon}}} \left\| M(\alpha)^{-1} \right\|_{\mathcal{B}(H)} \left\| m^I \right\|_{H^{(\alpha + \Delta - 1) \frac{2\bar{\beta}}{1 + 2\beta + \hat{\epsilon}}}}.$$  

Relying on the statement (6) of Assumptions 2 and (3.37), we obtain

(3.39) \[ II_{12} \lesssim n^{-\frac{\beta}{1 + 2\beta + \hat{\epsilon}}}. \]
Combining estimates (3.38) and (3.39), we obtain
\[(3.40)\]
\[\|\Pi_1\|^2 \lesssim n^{-\frac{2\hat{\beta}}{1+2\hat{\beta}+\epsilon}}.\]

Through some simple calculations, the term \(\Pi_2\) can be reformulated as follow
\[\Pi_2 = \frac{1}{n^{3/2}}(\mathcal{M}(\alpha) + n^{-1}I)^{-1}(\mathcal{M}(\alpha) - C^{\alpha+\Delta-1})(C^{\alpha+\Delta-1} + n^{-1}I)^{-1} \eta.\]

Then, we have
\[\mathbb{E}_0 \|\Pi_2\|^2 \leq n^{-3} \mathbb{E}_0 \left( \|\mathcal{M}(\alpha) + n^{-1}I\|_{\mathcal{H}}^{-1} \|C^{\alpha+\Delta-1} + n^{-1}I\|^{-1}\|\eta\|^2 + n^{-3} \mathbb{E}_0 \|\mathcal{M}(\alpha)\|_{\mathcal{H}}^{-1} \|C^{\alpha+\Delta-1} + n^{-1}I\|^{-1}\|\eta\|^2\right) \leq n^{-1} \mathbb{E}_0 \left( \|C^{\alpha+\Delta-1} + n^{-1}I\|^{-1}\|\eta\|^2 + n^{-1} \mathbb{E}_0 \|\mathcal{M}(\alpha)\|_{\mathcal{H}}^{-1} \|C^{\alpha+\Delta-1} + n^{-1}I\|^{-1}\|\eta\|^2\right) = \Pi_{21} + \Pi_{22}.

For the term \(\Pi_{21}\), we have the following estimates
\[(3.41)\]
\[\Pi_{21} \lesssim n^{-\frac{2\hat{\beta}}{1+2\hat{\beta}+\epsilon}} \mathbb{E}_0 \|C^{\frac{1+\hat{\beta}}{1+2\hat{\beta}+\epsilon}}\|^{-1}\|\eta\|^2.

Because \(\frac{1+\hat{\beta}}{1+2\hat{\beta}+\epsilon}(\alpha + \Delta - 1) > \alpha_0\) for \(n\) large enough, from (3.31) we obtain
\[(3.42)\]
\[\Pi_{21} \lesssim n^{-\frac{2\hat{\beta}}{1+2\hat{\beta}+\epsilon}},\]

where we used Lemma 3.3 in [1]. For the term \(\Pi_{22}\), we have
\[(3.43)\]
\[\Pi_{22} \lesssim n^{-1} \mathbb{E}_0 \left( \|\mathcal{M}(\alpha)C^{\alpha+\Delta-1}\|^{-1}\|\eta\|^2 + \mathbb{E}_0 \|\mathcal{M}(\alpha)\|_{\mathcal{H}}^{-1} \|C^{\alpha+\Delta-1} + n^{-1}I\|^{-1}\|\eta\|^2\right) \lesssim \Pi_{21} \lesssim n^{-\frac{2\hat{\beta}}{1+2\hat{\beta}+\epsilon}},\]

where the statement (7) of Assumptions 2 has been employed. Combining estimates (3.42) and (3.43), we obtain
\[(3.44)\]
\[\mathbb{E}_0 \|\Pi_2\|^2 \lesssim n^{-\frac{2\hat{\beta}}{1+2\hat{\beta}+\epsilon}}.

Combining estimations about I, II and III and recalling (3.25), we finish the proof. \(\square\)

3.3. Posterior estimation for the original problem. The final step is to obtain the posterior contraction estimation by employing Theorem 2.1 proved in [18] which provides a general framework for obtaining posterior consistency of linear inverse problems. Denote \(\mu_\ell = \mathcal{N}(A^{-1}u, C_1)\) and \(\mu'_\ell = \mathcal{N}(A^{-1}u, C_1)\). For \(u^1 \in \mathcal{H}^i\) and \(u \in H\), we know that \(A^{-1}u, A^{-1}u \in \mathcal{H}^{2i}\). The Cameron-Martin space for \(\mu_\ell\) and \(\mu'_\ell\) are all \(\mathcal{H}^i\). By Assumptions 2, we easily obtain \(A^{-1}u, A^{-1}u^1 \in \mathcal{H}^i\) which implies \(\mu_\ell\) and \(\mu'_\ell\) are equivalent. Inspired by the proof of Theorem 3.1 in [18], for given sequences of positive numbers \(k_n \to \infty\) and \(\rho_n \to 0\) and a constant \(c \geq 0\), we introduce
\[(3.45)\]
\[S_n = \left\{ u \in H : \sum_{i \geq k_n} u_i^2 \leq c \rho_n^2 \right\},\]
and
\[ B_n(A^{-1}u, \epsilon) = \begin{cases} u \in H : - \int \log \frac{dp}{d\mu} d\mu \leq n\epsilon^2, \\ \int \left| \log \frac{dp}{d\mu} - \int \log \frac{dp}{d\mu} d\mu \right|^2 d\mu \leq n\epsilon^2 \end{cases} \]

(3.46)

Through some simple calculations presented in the Appendix, we know that
\[ B_n(A^{-1}u, \epsilon) = \begin{cases} u \in H : \|T(u - u')\|^2 \leq \epsilon^2 \end{cases} \]

(3.47)

The proof of Lemma 1 in [12] can be adapted a little to achieve the following lemma.

**Lemma 3.7.** Let \( \delta_n \rightarrow 0 \) and let \( S_n \) be a sequence of sets \( S_n \subset H \). If \( \mu_n^0 = \mathcal{N}(0, C^0_n) \) is the prior probability measure on \( u \) satisfying
\[ \frac{\mu_n^0(S^c_n)}{\mu_n^0(B_n(A^{-1}u, \delta_n))} \lesssim \exp(-2n\delta_n^2), \]

then
\[ \mathbb{E}_0 \mu_{\alpha_n}^d(S^c_n) \rightarrow 0 \]

with \( \mu_{\alpha_n}^d \) represents the posterior probability measure of the original problem.

Relying on the above Lemma [3.7], we finally obtain our main result as follows.

**Theorem 3.8.** For every positive constant \( R > 0 \), \( C_3 := \frac{dC_2}{2p(1 + \frac{4}{\gamma} (\Delta - 1))} \) and an arbitrarily small positive constant \( \epsilon > 0 \), we have
\[ \sup_{m^j \in S^j(R) \cap \Theta^j_{c(R)}} \mathbb{E}_0 \mu_{\alpha_n}^d \{ u : \|u - u'\| \geq M_n \epsilon_n \} \rightarrow 0, \quad \text{as } n \rightarrow 0, \]

(3.48)

where
\[ \epsilon_n = (\log n)^{C_2 + C_3 + 4} (\log \log n)^{1/2} n^{-\frac{\gamma}{2p(1 + \frac{4}{\gamma} (\Delta - 1)) - 1}}. \]

**Proof.** **Step 1.** Let us firstly assume
\[ \mathbb{E}_0 \mu_{\alpha_n}^d(S^c_n) \rightarrow 0, \]

(3.49)

holds true. Let \( k_n \), \( \rho_n \) and \( \epsilon \) in the definition of \( S_n \) be fixed. For any \( u \in S_n \), we have
\[ \|u\|^2 = \sum_{j=1}^{\infty} u_j^2 = \sum_{j \leq k_n} u_j^2 + \sum_{j > k_n} u_j^2 \]
\[ \leq \sum_{j \leq k_n} u_j^2 + c\rho_n^2 = \sum_{j \leq k_n} \lambda_j^{2(\Delta - 1)} \lambda_j^{-2(\Delta - 1)} u_j^2 + c\rho_n^2 \]
\[ \leq \lambda_{k_n}^{-2(\Delta - 1)} \sum_{j \leq k_n} \lambda_j^{2(\Delta - 1)} u_j^2 + c\rho_n^2 \]
\[ \leq \lambda_{k_n}^{-2(\Delta - 1)} \|u\|_{L^{\gamma}_{0, (\Delta - 1)}}^2 + c\rho_n^2. \]

Denote \( u_j := \{u_j\}_{j=1}^{\infty} \) with \( u_j := (u_j^1, \phi_j) \) for \( j = 1, 2, \ldots \). Denote \( u_n^1 \) be function related to the projection of \( u_n^1 \) on the first \( k_n \) coordinates, that is, \( u_n^1 = \sum_{j=1}^{k_n} u_j^1 \phi_j. \)
Then, we have
\[
\|u_n^t - u^t\|^2 = \sum_{j=1}^{\infty} (u_{n,j}^t - u_j^t)^2 = \sum_{j > k_n} (u_j^t)^2 \\
= \sum_{j > k_n} \lambda_j^{-2\gamma} (u_j^t)^2 \lambda_j^{2\gamma} \leq \lambda_{k_n}^{2\gamma} \|u^t\|^2_{\mathcal{H}^{\gamma}},
\]
(3.51)
and
\[
\|u_n^t - u^t\|^2_{\mathcal{H}^{(\Delta - 1)}} = \sum_{j=1}^{\infty} \lambda_j^{2(\Delta - 1)} (u_{n,j}^t - u_j^t)^2 \\
= \sum_{j > k_n} \lambda_j^{2(\Delta - 1)} (u_j^t)^2 \lambda_j^{-2\gamma} \lambda_j^{2\gamma} \\
\leq \lambda_{k_n}^{2(\Delta - 1 + \gamma)} \|u^t\|^2_{\mathcal{H}^{\gamma}}.
\]
(3.52)
Using estimates (3.51) and (3.52), we have
\[
\|u - u^t\| \leq \|u - u_n^t\| + \|u_n^t - u^t\| \\
\leq \lambda_{k_n}^{(\Delta - 1)} \|u - u_n^t\|_{\mathcal{H}^{(\Delta - 1)}} + \sqrt{c} \rho_n + \lambda_{k_n}^{\gamma} \|u^t\|_{\mathcal{H}^{\gamma}} \\
\leq \lambda_{k_n}^{(\Delta - 1)} \left[ \|u - u_n^t\|_{\mathcal{H}^{(\Delta - 1)}} + \lambda_{k_n}^{\gamma} \|u^t\|_{\mathcal{H}^{\gamma}} \right] + \sqrt{c} \rho_n + \lambda_{k_n}^{\gamma} \|u^t\|_{\mathcal{H}^{\gamma}} \\
\leq \lambda_{k_n}^{(\Delta - 1)} \|u - u_n^t\|_{\mathcal{H}^{(\Delta - 1)}} + \sqrt{c} \rho_n + 2 \lambda_{k_n}^{\gamma} \|u^t\|_{\mathcal{H}^{\gamma}} \\
\leq \lambda_{k_n}^{(\Delta - 1)} \|u - u_n^t\|_{\mathcal{H}^{(\Delta - 1)}} + \rho_n + \lambda_{k_n}^{\gamma} \\
\leq \lambda_{k_n}^{(\Delta - 1)} \|T u - T u_n^t\| + \rho_n + \lambda_{k_n}^{\gamma}.
\]
(3.53)
For a large positive constant $M > 0$ (will be specified later), we denote $\bar{\epsilon} = \epsilon / M$. Let
\[
k_n = n^{\frac{1}{1+\frac{2d}{\gamma+(\Delta-1)+\frac{d}{\gamma}}}}, \quad \rho_n = L_n n^{\frac{1}{1+\frac{2d}{\gamma+(\Delta-1)+\frac{d}{\gamma}}}},
\]
and $\|T u - T u_n^t\| \leq M_n \bar{\epsilon}_n$, with $\bar{\epsilon}_n$ defined as in Theorem 3.6, we find that
\[
\|u - u_n^t\| \leq M_n L_n k_n^{\frac{1}{1+\frac{2d}{\gamma+(\Delta-1)+\frac{d}{\gamma}}}} n^{\frac{1}{1+\frac{2d}{\gamma+(\Delta-1)+\frac{d}{\gamma}}}} + L_n n^{\frac{1}{1+\frac{2d}{\gamma+(\Delta-1)+\frac{d}{\gamma}}}} + k_n^{-\frac{2d}{\gamma}} \\
\leq M_n L_n n^{\frac{1}{1+\frac{2d}{\gamma+(\Delta-1)+\frac{d}{\gamma}}}}.
\]
(3.55)
Applying estimate (3.18) to the last line of (3.55), we finally obtain
\[
\|u - u_n^t\| \lesssim M_n L_n (\log n)^{C_3} n^{-\frac{1}{1+2d/\gamma+(\Delta-1)}}.
\]
(3.56)
Taking $M > 0$ large enough such that
\[
\frac{d^2 + 2pd(\gamma + \Delta - 1) + 2\epsilon}{2p\gamma} + 1 \leq M,
\]
then we have
\[
\frac{1}{n^{1+\frac{2d}{\gamma+(\Delta-1)+\frac{d}{\gamma}}}} \leq n^{-\frac{2d}{2p(\gamma+(\Delta-1)+\frac{d}{\gamma})}}.
\]
Recalling Theorem 3.6 and Theorem 2.1 in [18], the proof is completed.
Step 2. In this step, we aim to prove (3.19). In the following, we use the values of \( k_n \) and \( \rho_n \) given in (3.54). Taking \( n \) large enough, following the proof of Lemma 5.2 in [18], we can obtain
\[
\mu_n^0(S_n^c) \leq \exp \left( -\frac{c}{8} \mu_n^2 k_n^\frac{2}{3} \hat{n} \right) \\
= \exp \left( -\frac{c}{8} L_n^2 n \alpha_0 \left( \frac{1}{2} \right)^{\frac{1}{2}\gamma_0} \right).
\]

Choose \( \hat{c}_n \) as in Theorem 3.6 and notice that for a constant \( C > 0 \),
\[
\mu_n^0(B_n(A^{-1}u^1, \hat{c}_n)) \geq \mu_n^0(u : \|Tu - T u^1\|^2 \leq \hat{c}_n^2)
\]
\[
\geq \mu_n^0(u : \|u - u^1\|^2 \leq C^{-1} \hat{c}_n^2)
\]
\[
\geq \mu_n^0 \left( u : \sum_{i=1}^{N} \lambda_i^{2(\Delta - 1)} (u_i - u_i^1)^2 \leq \frac{\hat{c}_n^2}{2C} \right) \mu_n^0 \left( u : \sum_{i=N+1}^{\infty} \lambda_i^{2(\Delta - 1)} (u_i - u_i^1)^2 \leq \frac{\hat{c}_n^2}{2C} \right).
\]

Then, following the proof of Lemma 5.1 in [18], we have
\[
\mu_n^0(B_n(A^{-1}u^1, \hat{c}_n)) \leq C_4 \exp \left( -C_3 \hat{c}_n \right)
\]
\[
= C_4 \exp \left( -C_3 L_n \frac{1}{n^{\frac{2}{\gamma_0}} \left( \frac{1}{2} \right)^{\frac{1}{2}\gamma_0}} \right).
\]

Combining (3.58) and (3.59), we arrive at
\[
\frac{\mu_n^0(S_n^c)}{\mu_n^0(B_n(A^{-1}u^1, \hat{c}_n))} \leq \exp \left( -\left( \frac{c}{8} - C_3 L_n \frac{1}{n^{\frac{2}{\gamma_0}} \left( \frac{1}{2} \right)^{\frac{1}{2}\gamma_0}} \right) n \hat{c}_n^2 \right).
\]

Choosing \( \epsilon \) in the definition of \( S_n \) large enough, we end the proof by using Lemma 3.7.

In the last part of this section, we would like to provide some discussions about the condition \( m_1^v \in \Theta_{ss}^\hat{\beta}(R) \). Concerning the self-similar sequence, we can prove the following theorem which may be compared to Proposition 3.5 of [27].

Theorem 3.9. For every \( 0 < \alpha < \hat{\beta} \), the prior probability measure \( \Pi_n \) is defined as follow
\[
\Pi_n := \bigotimes_{i=1}^{\infty} N(0, i^{-1-2\alpha}).
\]

Let \( \alpha \leq \frac{\hat{\beta}}{2} \) and the parameters appeared in the definition of self-similar sequence satisfy \( \rho - 1 \geq \epsilon R^{1+2\hat{\beta}} \), then we have \( \Pi_n(\cup_{N_i} \Theta_{ss}^\hat{\beta}(R)) = 1 \).

The proof details are postponed to the Appendix. This theorem obviously imply that the set of functions with self-similar coefficients
\[
m = \sum_{i=1}^{\infty} m_i \phi_i, \quad \{m_i\}_{i=1}^{\infty} \in \Theta_{ss}^\hat{\beta}(R)
\]
is large in the sense of Bayesian setup and almost every realization from the prior \( \mathcal{N}(0, C^{\alpha+\Delta-1}) (\alpha \leq \hat{\beta}) \) is a function with self-similar coefficients. Under the following two conditions
- \( C^{\frac{\alpha+\Delta-1}{2}}(H) = (TC^\alpha T^*)^{\frac{1}{2}}(H) \),
• The operator $C^{-\frac{1}{2}d}TC^\alpha T^*C^{-\frac{1}{2}d} = I$ is a Hilbert-Schmidt operator on the closure of the space $C^{-\frac{1}{2}d}(H)$, we know that $\mathcal{N}(0, C^{\alpha+\Delta-1})$ and $\mathcal{N}(0, TC^\alpha T^*)$ are equivalent probability measures \[24\]. That is to say, the self-similar set is also large when considering the original non-diagonal problem. The above two conditions can be verified for many specific problems, e.g., when these operators are some pseudo-differential operators, it can be verified through boundedness of these operators on certain Sobolev spaces \[25\].

4. Examples

In this section, we provide some nontrivial examples. For simplicity, we consider a simple closed manifold that is $d$-dimensional torus denoted as $\mathbb{T}^d$. Introduce the Hilbert space $H$ defined as follow

$$H := L^2(\mathbb{T}^d) = \left\{ u : \mathbb{T}^d \to \mathbb{R} \mid \int_{\mathbb{T}^d} |u(x)|^2 dx < \infty, \int_{\mathbb{T}^d} u(x) dx = 0 \right\}$$

of real valued periodic function $d \leq 3$ with inner-product and norm denoted by $(\cdot, \cdot)$ and $\| \cdot \|$ respectively. Let $A_0 := -\Delta$ be the negative Laplacian equipped with periodic boundary condition on $[0, 1)^d$, and restricted to functions with integrate to zero over $[0, 1)^d$.

It is well-known that this operator is positive self-adjoint and has eigensystem $\{\rho_j^2, \phi_j\}_{j=1}^\infty$. The eigenfunctions $\{\phi_j\}_{j=1}^\infty$ constitute the Fourier basis and form a complete orthonormal basis of $H$ and the eigenvalues $\rho_j^2$ behave asymptotically like $j^{2/d}$.

Before going further, let us provide two basic definitions concerned with the pseudo-differential operators. For further details, we refer to \[17\] \[25\].

**Definition 4.1.** Let $m \in \mathbb{R}$. Then $S^m(\mathbb{R}^d, \mathbb{R}^d)$ is the vector-space of all smooth functions $a \in C^\infty(\mathbb{R}^d, \mathbb{R}^d)$ such that

$$|\partial_\xi^\alpha \partial_\eta^\beta a(x, \xi)| \leq C_{\alpha,\beta,K}(1 + |\xi|)^{|m-|\beta||}, \quad \xi \in \mathbb{R}^d, \quad x \in K,$$

holds for all multi-indices $\alpha$ and $\beta$ and any compact set $K \subset \mathbb{R}^d$.

**Definition 4.2.** Let $Y : U \to \mathbb{R}^d$ be local coordinates of the manifold $N$. A bounded linear operator $A : \mathcal{D}(N) \to \mathcal{D}'(N)$ is called a pseudo-differential operator if for any local coordinates $Y : U \to \mathbb{R}^d$, $U \subset N$, there is a symbol $a \in S^m(\mathbb{R}^d, \mathbb{R}^d)$ such that for $u \in C_0^\infty(U)$ we have

$$Au(y_1) = \int_N k_A(y_1, y_2)u(y_2)dV_y(y_2),$$

where $k_A|_{N \times N\setminus\text{diag}(N)} \in C^\infty(N \times N \setminus \text{diag}(N))$ and $\text{diag}(N) = \{(y, y) \in N \times N \mid y \in N\}$. Also when $Y : U \to V \subset \mathbb{R}^d$ are local $C^\infty$-smooth coordinates $k_A(y_1, y_2)$ is given on $U \times U$ by

$$k_A(Y^{-1}(x_1), Y^{-1}(x_2)) = \int_{\mathbb{R}^d} e^{i(x_1-x_2)\xi}a(x_1, \xi)d\xi,$$

where $x_1, x_2 \in V \subset \mathbb{R}^d$ and $a = a_V \in S^m(V, \mathbb{R}^d)$. In this case we will write

$$A \in \Psi^m(N),$$

and say that in local coordinates $Y : U \to V \subset \mathbb{R}^d$ the operator $A$ has the symbol $a(x, \xi) \in S^m(V \times \mathbb{R}^d)$.

Following the proof of Lemma 1 in \[17\], we can obtain the following lemma.
Let \( B \in \Psi^{-2}(\mathbb{T}^d) \) be an injective elliptic pseudodifferential operator. Then we have the following estimates
\[
\|Bu\|_{\dot{H}^{r+2}\left(\mathbb{T}^d\right)} \lesssim \|u\|_{H^{r}\left(\mathbb{T}^d\right)} \lesssim \|Bu\|_{\dot{H}^{r}\left(\mathbb{T}^d\right)}, \quad \forall \ r \in \mathbb{R}.
\]

4.1. Example 1–non-diagonal forward operator. Let \( M_q : H \to H \) be the multiplication operator by a nonnegative function \( q \in C^\infty(\mathbb{T}^d) \). We define the forward operator \( A^{-1} := (A_0 + M_q)^{-1} \), assume the observational noise is white, so that \( C_1 = I \), and we set the operator \( C = A_0^{-2} \).

Denote the eigenvalues of the operator \( C \) to be \( \{\lambda_j^2\}_{j=1}^{\infty} \). Obviously, we have \( \lambda_j^2 = \rho_j^{-4} \). Since \( \sum_{j=1}^{\infty} \lambda_j^2 \lesssim \sum_{j=1}^{\infty} j^{-1} \leq \infty \) for \( d \leq 3 \), the operator \( C \) is trace class and the constant \( p \) appeared in Assumptions 1 is equal to 2.

Our aim is to show \( C_1 \simeq C^d \) and \( A^{-1} \simeq C^d \), where \( \beta = 0 \) and \( \ell = \frac{1}{2} \) in the sense specified in Assumptions 2. Obviously, we know \( \Delta = 2\ell - \beta + 1 = 2 \). Under these settings, we easily obtain the equivalent relations of Hilbert scales and Sobolev space as follow
\[
\mathcal{H}^d = \dot{H}^{2d}(\mathbb{T}^d).
\]

Now, we verify all of the conditions appeared in Assumptions 2.

(2): For this condition, we need to prove the following statement
\[
\|A^{-1}u\|_{\dot{H}^{2r}(\mathbb{T}^d)} \lesssim \|u\|_{H^{2r-2}(\mathbb{T}^d)}, \quad \forall \ r \geq 0.
\]

The forward operator \( A^{-1} \) is a pseudodifferential operator with symbol \( (|\xi|^2 + q(x))^{-1} \). That is to say, \( A^{-1} \in \Psi^{-2}(\mathbb{T}^d) \) and \( A^{-1} \) is an injective elliptic pseudodifferential operator. Through Lemma 4.3 we complete the verification.

(3): The operator \( \mathcal{C}^{\frac{d}{2}}T^* = \mathcal{C}^{\frac{d}{2}}A^{-1} \) is a pseudodifferential operator with symbol \( |\xi|^{-2\alpha}(|\xi|^2 + q(x))^{-1} \). Hence, we have \( \mathcal{C}^{\frac{d}{2}}T^* \in \Psi^{-2(1+\alpha)}(\mathbb{T}^d) \) which implies
\[
\|\mathcal{C}^{\frac{d}{2}}T^*u\|_{L^2(\mathbb{T}^d)} \lesssim \|u\|_{H^{-2(1+\alpha)}(\mathbb{T}^d)}.
\]

(4): To verify this condition, let us notice that the operator
\[
\mathcal{C}^{-(\alpha+\Delta-1)}M(\alpha) \in \Psi^0(\mathbb{T}^d),
\]
which indicates
\[
\|\mathcal{C}^{-(\alpha+\Delta-1)}M(\alpha)u\| \lesssim \|u\|_{L^2(\mathbb{T}^d)}, \quad \forall \ u \in L^2(\mathbb{T}^d).
\]

Conditions (5) to (7) can be verified similarly, here, we omit the details for concisely. We can now apply Theorem 8.38 to obtain the following convergence result.

**Theorem 4.4.** Let \( u^0 \in H^{2\gamma}(\mathbb{T}^d) \), \( \gamma \geq 1 \). Then, the convergence in \( H^{2\gamma}(\mathbb{T}^d) \) holds with
\[
\bar{e}_n = (\log n)^{C_2+C_3+4} (\log \log n)^{1/2} n^{\frac{2\gamma}{2\gamma+2}},
\]
for some large enough constants \( C_2 \) and \( C_3 \) and arbitrarily small constant \( \epsilon > 0 \).

Comparing with Theorem 8.2 proved in [1], we notice that this convergence rate is optimal up to a slowly varying factor and an arbitrarily small positive constant \( \epsilon > 0 \). Finally, let us verify the equivalence of probability measures \( \mathcal{N}(0, C^{\alpha+\Delta-1}) \) and \( \mathcal{N}(0, TC^{\alpha}T^*) \). Since the closure of \( \mathcal{C}^{2d/2}H \) is \( H \) under our assumptions, we
only need to prove that the operator \( J := C^{-\frac{\alpha}{2}} A^{-1} C^\alpha A^{-1} C^{-\frac{\alpha}{2}} - I \) is a Hilbert-Schmidt operator on \( L^2(T^d) \). Because the symbol of the operator \( J \) is

\[
\frac{2|\xi|^2 q(x) + q(x)^2}{(|\xi|^2 + q(x))^2},
\]

we easily derive that \( J \in \Psi^{-2}(T^d) \). Because of \(-2 < -\frac{d}{2}\) for \( d = 1, 2, 3 \), the operator \( J \) is obviously a Hilbert-Schmidt operator. Hence, concerning this example, the restriction of self-similar sequence in Theorem 3.8 is reasonable. For more explanations about self-similar sequence assumptions, we refer to [27].

4.2. Example 2–a fully non-diagonal example. As in Section 4.1, let the forward operator to be \( A^{-1} = (A_0 + M_q)^{-1} \). We assume that the observational noise is Gaussian with covariance operator \( C_1 = (A_0 + M_r)^{-2} \), where \( M_r \) is the multiplication operator by a nonnegative function \( r \in C^\infty(T^d) \). We also assume \( C = A_0^{-2} \).

Under this setting, we would like to verify Assumptions 2 with \( \ell = \frac{1}{2} \) and \( \beta = 1 \). We easily know that \( A^{-1} \in \Psi^{-2}(T^d) \), \( C_1 \in \Psi^{-4}(T^d) \) and \( C^{\alpha} \in \Psi^{-4\alpha}(T^d) \). Assumptions 2 and the equivalence of \( \mathcal{N}(0, C^{\alpha+\Delta-1}) \) and \( \mathcal{N}(0, TC^{\alpha}T^*) \) can be verified through similar analysis shown in Section 4.1. Then we apply Theorem 3.8 to obtain the following convergence result.

**Theorem 4.5.** Let \( u^1 \in H^{2\gamma}(T^d), \gamma \geq 1 \). Then, the convergence in (3.48) holds with

\[
\tilde{\epsilon}_n = (\log n)^{C_2 + C_3 + 4} (\log \log n)^{1/2} n^{\frac{2\gamma}{2\gamma + 4\gamma + \epsilon}},
\]

for some large enough constants \( C_2 \) and \( C_3 \) and arbitrarily small constant \( \epsilon > 0 \).

5. Conclusions

In this paper, we study an empirical Bayesian approach to a family of linear inverse problems. We assume that the covariance operator of the prior Gaussian measure depends on a hyperparameter that quantifies the regularity. We do not assume the prior covariance, noise covariance and forward operator have a common basis in their singular value decomposition, which enlarge the application range compared with the existing results. Under such weak assumptions, it is difficult to introduce maximum likelihood based estimation for the regularity index. In order to construct an empirical Bayesian approach, we propose two auxiliary problems: transformed problem and artificial diagonal problem. Since the regularity index only reflects the information about regularity, we provide a maximum likelihood estimation of the regularity index through the artificial diagonal problem. Then we deduce the posterior contraction estimates for the transformed problem by exploring the relations between the transformed problem and the artificial diagonal problem. Finally, the desired posterior contraction estimates are obtained through a general approach developed in a recent paper [18].

In order to give appropriate estimates for the transformed problem, we employ the self-similar sequence condition which is also used in [3, 14, 27]. For illustrating the appropriateness of this condition, we need to show the equivalence of the prior measures of the artificial diagonal problem and the transformed problem. The general conditions to ensure such equivalence are given and verified through boundedness of pseudo-differential operators for our two examples.
For the posterior consistency problem, there are numerous challenging problems need to be solved. Only inverse problems with linear forward operator have been considered in this paper, how to design empirical Bayesian methods for inverse problems with nonlinear forward operator is a difficult problem and it seems to depend on the specific structures of the certain problem.

6. Appendix

Here we gather the proofs of various results used in this paper. Including these proofs in the main text would break the flow of ideas related to posterior consistency.

Proof of Lemma 3.1

Proof. The proof of estimation (6.11) is similar to the proof of (i) of Lemma 1 presented in [19], so we omit the details. For the estimation (3.12), let us denote \( \tilde{\beta}_n := \tilde{\beta} + C_1 (\log \log n) / \log n \). Employing Lemma 6.1 proved in [27], we find that

\[
(6.1) \quad h_n(\tilde{\beta}) \geq \frac{\epsilon R}{(\rho^2 + 2\beta + 1)^2} n^{2(\tilde{\beta}_n - \tilde{\beta})/(1 + 2\tilde{\beta}_n)}
\]

holds for choosing \( n \) large enough such that \( C_1 (\log \log n) / \log n \leq 1 \). Because \( (\log \log n) / \log n \leq 1/4 \) for \( n \) large enough, we have

\[
(6.2) \quad n^{2(\tilde{\beta}_n - \tilde{\beta})/(1 + 2\tilde{\beta}_n)} \geq n^{\frac{1}{4}(\log \log n) - \frac{C_1}{1 + 2\tilde{\beta} + C_1 / 2}} = (\log n)^{2C_1/(1 + 2\tilde{\beta} + C_1 / 2)}.
\]

Hence, for \( C_1 \) large enough (depending on \( \tilde{\beta}, R, \epsilon \) and \( \rho \)), we obtain

\[
(6.3) \quad h_n(\tilde{\beta}) \geq L(\log n)^2
\]

for large \( n \), which implies that estimation (3.12) holds true. □

Proof of Theorem 3.2

Proof. Denote \( \hat{m}_n(\tilde{\alpha}) = \{\hat{m}_n(\tilde{\alpha})\}_{i=1}^{\infty} \) to be posterior mean estimator of the artificial diagonal problem (3.2). Since

\[
\|\hat{m}_n(\tilde{\alpha}) - m_\dagger\|^2 = \sum_{i=1}^{\infty} (\hat{m}_n,i(\tilde{\alpha}) - m_\dagger,i)^2 \approx \sum_{i=1}^{\infty} (\hat{m}_n,i(\tilde{\alpha}) - \tilde{m}_n,i)^2,
\]

we just need to prove the following estimation

\[
(6.3) \quad \sup_{m_\dagger \in S^\beta(R) \cap \Theta_n^\beta(R)} \mathbb{E}_0 \left\{ \sup_{\tilde{\alpha} \in I_n} \sum_{i=1}^{\infty} (\hat{m}_i(\tilde{\alpha}) - \tilde{m}_n,i)^2 \right\} = O(\epsilon_n^2).
\]

Inspired by the Subsection 6.1 in [19], we actually need to estimate the following terms

\[
(6.4) \quad \sum_{i=1}^{\infty} \frac{\epsilon^2 + 4\alpha (\hat{m}_n)^2}{(i^2 + 2\alpha + n)} + n \sum_{i=1}^{\infty} \frac{1}{(i^2 + 2\alpha + n)} = I + II.
\]

Employing Lemma 8 in [19] (with \( m = 0, \ell = 1, r = 1 + 2\alpha \) and \( s = 0 \)), we find that

\[
(6.5) \quad II \leq \sum_{i=1}^{\infty} \frac{1}{i^2 + 2\alpha + n} \leq n^{\frac{2\alpha}{1 + 2\alpha}} \leq (\log n)^{2C_2} n^{-\frac{2\beta}{1 + 2\alpha}},
\]

so that

\[
(6.6) \quad h_n(\tilde{\beta}) \geq \frac{\epsilon R}{(\rho^2 + 2\beta + 1)^2} n^{2(\tilde{\beta}_n - \tilde{\beta})/(1 + 2\tilde{\beta}_n)}
\]

holds for choosing \( n \) large enough such that \( C_1 (\log \log n) / \log n \leq 1 \). Because \( (\log \log n) / \log n \leq 1/4 \) for \( n \) large enough, we have

\[
(6.7) \quad n^{2(\tilde{\beta}_n - \tilde{\beta})/(1 + 2\tilde{\beta}_n)} \geq n^{\frac{1}{4}(\log \log n) - \frac{C_1}{1 + 2\tilde{\beta} + C_1 / 2}} = (\log n)^{2C_1/(1 + 2\tilde{\beta} + C_1 / 2)}.
\]

Hence, for \( C_1 \) large enough (depending on \( \tilde{\beta}, R, \epsilon \) and \( \rho \)), we obtain

\[
(6.8) \quad h_n(\tilde{\beta}) \geq L(\log n)^2
\]

for large \( n \), which implies that estimation (3.12) holds true. □
where we used estimates (3.16). For term I, it can be estimated by following the procedures used in the Subsection 6.1 of [19] with lower bound estimate of $\alpha$ similar to (6.5). Here, we omit the proof details.

Proof of Lemma 3.4

Proof. We use the Lax-Milgram theorem [11] in the Hilbert space $H$. The coercive of $B$ can be illustrated as follow

$$B(u, u) = n\|C^{\alpha/2}T^* u\|^2 + \|u\|^2 \geq \|u\|^2, \quad \forall u \in H.$$

Using the statement (3) of Assumptions 2, we know that $\|C^{\alpha/2}T^* u\| \lesssim \|u\|_{H^{\alpha-(\Delta-1)}}$. The continuity of the bilinear form $B$ can be derived as follow

$$|B(u, v)| \leq n\|C^{\alpha/2}T^* u\|\|C^{\alpha/2}T^* v\| + \|u\|\|v\| \lesssim \|u\|\|v\|, \quad \forall u, v \in H.$$

So the first statement has been proved. For the second statement about regularities, we notice that

$$u = r - n\mathcal{M}(\alpha)u.$$

Using the statement (4) of Assumptions 2, we easily deduce that $\mathcal{M}(\alpha)u \in H^{2(\alpha+\Delta-1)}$ for $u \in H$. Hence the right-hand side belongs to $H^t$, which implies $u \in H^t$ for $t \leq 2(\alpha + \Delta - 1)$. □

Proof of Lemma 3.5

Proof. Let $h \in H$. Then $C^{\frac{1}{2}(\alpha+\Delta-1)-\frac{s}{2}}h \in H$, since $\Delta \geq 1$ and $\alpha_0 < s < \alpha$. By Lemma 3.4 for $r = C^{\frac{1}{2}(\alpha+\Delta-1)-\frac{s}{2}}h$, there exists a unique weak solution of (3.22), $u' \in H$. Since for $v \in H^{(\alpha+\Delta-1)-s}$ we have $C^{-\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}}v \in H$, we conclude that for any $v \in H^{(\alpha+\Delta-1)-s}$

$$n\|C^{\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}} u\|^2 + n\|C^{\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}} v\|^2 + \left(C^{-\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}} u, C^{-\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}} v\right) + \left(C^{-\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}} h, C^{-\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}} v\right) = \left(C^{\frac{1}{2}(\alpha+\Delta-1)-\frac{s}{2}} h, C^{-\frac{1}{2}(\alpha+\Delta-1)+\frac{s}{2}} v\right),$$

where $u = C^{\frac{1}{2}(\alpha+\Delta-1)-\frac{s}{2}} u' \in H^{(\alpha+\Delta-1)-s}$. Choosing $v = u$, using the statement (3) of Assumptions 2, for a generic constant $c > 0$ and a small enough constant $\delta > 0$, we obtain

$$n\|u\|_{H^{\alpha-(\Delta-1)-s}}^2 + \|u\|_{H^{(\alpha+\Delta-1)-s}}^2 \leq c\|h\|\|u\|$$

$$\leq cn^{-\frac{1}{2}(1-\frac{\alpha+\Delta-1}{2})}\|h\|\left(n\frac{\|u\|_{H^{\alpha-(\Delta-1)-s}}}{\|u\|_{H^{(\alpha+\Delta-1)-s}}}\right)^{1-\frac{\alpha+\Delta-1}{2}}\|u\|_{H^{(\alpha+\Delta-1)-s}}$$

$$\leq c\|u\|_{H^{(\alpha+\Delta-1)-s}}^2 + c\|u\|_{H^{(\alpha+\Delta-1)-s}}^2 = \frac{c\|u\|_{H^{(\alpha+\Delta-1)-s}}^2 + c\delta}{2},$$

where interpolation inequality shown in Lemma 2.1 has been employed. Then we obviously have

$$\|u\|_{H^{(\alpha+\Delta-1)-s}} \lesssim n^{-\frac{1}{2}(1-\frac{\alpha+\Delta-1}{2})}\|h\|, \quad \|u\|_{H^{\alpha-(\Delta-1)-s}} \lesssim n^{-1+\frac{\alpha+\Delta-1}{2}}\|h\|.$$  

Finally, inserting the above estimates into the following interpolation inequality

$$\|u\| \leq \|u\|_{H^{\alpha-(\Delta-1)-s}}^1 + \|u\|_{H^{(\alpha+\Delta-1)-s}}^1,$$

we get

$$\|u\| \lesssim n^{-1+\frac{\alpha+\Delta-1}{2}}\|h\|.$$
Replacing \( u = C^{\frac{1}{2}}(\alpha + \Delta - 1)^{-\frac{1}{2}} (nM(\alpha) + I)^{-1} C^{\frac{1}{2}}(\alpha + \Delta - 1)^{-\frac{1}{2}} h \) gives the desired estimate (3.24).

□

Proof of (3.47)

Proof. Firstly, let us denote \( W_z \) as the white noise mapping for \( z \in H \). Recalling the Cameron-Martin formula, we can easily obtain

\[
\frac{d\mu'_t}{d\mu_t}(d) = \exp\left( -\frac{n}{2} \| T(u - u^\dagger) \|^2 + W_{\sqrt{n} T(u - u^\dagger)}(d) \right).
\]

Taking logarithm with respect to the above equality (6.6) and integrating the obtained equality, we obtain

\[
-\int \log \frac{d\mu'_t}{d\mu_t} d\mu_t = \int \frac{n}{2} \| T(u - u^\dagger) \|^2 - W_{\sqrt{n} T(u - u^\dagger)}(\cdot) d\mu_t
\]

\[
= \frac{n}{2} \| T(u - u^\dagger) \|^2,
\]

where we used the properties of the white noise mapping. To complete the proof, we should notice that

\[
\int | \log \frac{d\mu'_t}{d\mu_t} |^2 d\mu_t = \int (\log \frac{d\mu'_t}{d\mu_t})^2 d\mu_t - \left( \int \log \frac{d\mu'_t}{d\mu_t} d\mu_t \right)^2.
\]

For the first term on the right hand side of (6.8), we have

\[
\int (\log \frac{d\mu'_t}{d\mu_t})^2 d\mu_t = \frac{n^2}{4} \| T(u - u^\dagger) \|^4 + \int (W_{\sqrt{n} T(u - u^\dagger)}(\cdot))^2 d\mu_t
\]

\[
- \int n \| T(u - u^\dagger) \|^2 W_{\sqrt{n} T(u - u^\dagger)}(\cdot) d\mu_t
\]

\[
= \frac{n^2}{4} \| T(u - u^\dagger) \|^4 + n \| T(u - u^\dagger) \|^2.
\]

Noticing that

\[
\left( \int \log \frac{d\mu'_t}{d\mu_t} d\mu_t \right)^2 = \frac{n^2}{4} \| T(u - u^\dagger) \|^4,
\]

we find that

\[
\int \left| \log \frac{d\mu'_t}{d\mu_t} \right|^2 d\mu_t = n \| T(u - u^\dagger) \|^2.
\]

Combining equalities (6.7) and (6.10), we finally obtain

\[
B_n(A^{-1}u^\dagger, \epsilon) = \left\{ u \in H : \| T(u - u^\dagger) \|^2 \leq \epsilon^2 \right\}.
\]

□

Proof of Theorem 3.9
Proof. Let $u_1, u_2, \ldots$ be independent random variables with $u_i \sim \mathcal{N}(0, i^{-1-2\alpha})$, and let $\Omega_N$ be the event $\{\sum_{i=N}^{\rho N} u_i^2 < \epsilon RN^{-2\beta}\}$. Relying on the Borel-Cantelli lemma, it suffices to show that $\sum_{N \in \mathbb{N}} \Pi_\alpha(\Omega_N) < \infty$. Simple calculations yield

$$\sum_{i=N}^{\rho N} \mathbb{E}u_i^2 = \sum_{i=N}^{\rho N} i^{-1-2\alpha} \geq \frac{\rho - 1}{\rho^{1+2\alpha}} N^{-2\alpha}.$$ 

By Markov’s inequality, followed by the Marcinkiewicz-Zygmund and Hölder inequality, we obtain

$$\Pi_\alpha(\Omega_N) \lesssim N^{2\beta q} \mathbb{E} \left( \sum_{i=N}^{\rho N} (u_i^2 - \mathbb{E}u_i^2) \right)^q \lesssim N^{2\beta q} \left( \mathbb{E} \left( \sum_{i=1}^{\rho N} u_i^2 - \mathbb{E}u_i^2 \right)^2 \right)^{q/2} \lesssim N^{2\beta q} \sum_{i=N}^{\infty} \mathbb{E}(u_i^2 - \mathbb{E}u_i^2)^q r(q/2-1) \left( \sum_{i \geq N} i^{-r} \right)^{q/2-1}.$$ 

Because $\mathbb{E}(u_i^2 - \mathbb{E}u_i^2)^q \asymp i^{-2q\alpha-q}$, for $-2q\alpha-q+r(q/2-1) < -1$, the right-hand side of the last inequality is of the order $N^{-q/2}$. Choosing $q > 2$, the proof is completed. 

\[\Box\]

Acknowledgments

The authors would like to thank the anonymous referees for their comments and suggestions, which helped to improve the paper significantly. This work was partially supported by the National Natural Science Foundation of China under the grants no. 11871392, 11501439, 11771347, 11131006 and 41390450, the Postdoctoral Science Foundation Project of China under grant no. 2017T100733.

References

[1] S. Agapiou, S. Larsson, and A. M. Stuart. Posterior contraction rates for the Bayesian approach to linear ill-posed inverse problems. \textit{Stoch. Proc. Appl.}, 123(10):3828–3860, 2013.
[2] S. Agapiou and P. Mathé. \textit{Posterior contraction in Bayesian inverse problems under Gaussian priors}. New Trends in Parameter Identification for Mathematical Models. Springer, 2018.
[3] A. D. Bull. Honest adaptive confidence bands and self-similar functions. \textit{Electron. J. Stat.}, 6:1490–1516, 2012.
[4] M. Burger and F. Lucka. Maximum a posteriori estimates in linear inverse problems with log-concave priors are proper Bayes estimators. \textit{Inverse Probl.}, 30(11):114004, 2014.
[5] S. L. Cotter, M. Dashti, J. C. Robinson, and A. M. Stuart. Bayesian inverse problems for functions and applications to fluid mechanics. \textit{Inverse Probl.}, 25(11):115008, 2009.
[6] S. L. Cotter, G. O. Roberts, A. M. Stuart, and D. White. MCMC methods for functions: modifying old algorithms to make them faster. \textit{Stat. Sci.}, 28(3):424–446, 2013.
[7] M. Dashti, K. J. H. Law, A. M. Stuart, and J. Voss. MAP estimators and their consistency in Bayesian nonparametric inverse problems. \textit{Inverse Probl.}, 29(9):095017, 2013.
[8] M. Dashti and A. M. Stuart. The Bayesian approach to inverse problems. \textit{Handbook of Uncertainty Quantification}, pages 311–428, 2017.
[9] M. M. Dunlop and A. M. Stuart. MAP estimators for piecewise continuous inversion. \textit{Inverse Probl.}, 32(10):105003, 2016.
[10] H. W. Engl, M. Hanke, and A. Neubauer. \textit{Regularization of Inverse Problems}. Springer, Netherlands, 1996.
[11] Lawrence C Evans. \textit{Partial Differential Equations}. Providence, RI: American Mathematical Society, United States, second edition, 2010.
[12] S. Ghosal and A. Van Der Vaart. Convergence rates of posterior distributions for noniid observations. *Ann. Statist.*, 35(1):192–223, 2007.
[13] E. Giné and R. Nickl. Confidence bands in density estimation. *Ann. Statist.*, 38(2):1122–1170, 2010.
[14] M. Hoffmann and R. Nickl. On adaptive inference and confidence bands. *Ann. Statist.*, 39(5):2383–2409, 2011.
[15] J. Jia, J. Peng, and J. Gao. Bayesian approach to inverse problems for functions with a variable-index Besov prior. *Inverse Probl.*, 32(8):085006, 2016.
[16] J. Jia, S. Yue, J. Peng, and J. Gao. Infinite-dimensional Bayesian approach for inverse scattering problems of a fractional Helmholtz equation. *J. Funct. Anal.*, 275(9):2299–2332.
[17] H. Kekkonen, M. Lassas, and S. Siltanen. Posterior consistency and convergence rates for Bayesian inversion with hypoelliptic operators. *Inverse Probl.*, 32(8):085005, 2016.
[18] B. T. Knapik and J. B. Salomond. A general approach to posterior contraction in nonparametric inverse problems. *Bernoulli*, 24(3):2091–2121, 2018.
[19] B. T. Knapik, B. T. Szabó, A. W. van der Vaart, and J. H. van Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probab. Theory Rel.*, 164(3):771–813, 2016.
[20] B. T. Knapik, A. van Der Vaart, and J. H. van Zanten. Bayesian inference with Gaussian priors. *Ann. Statist.*, 39:2626–2657, 2011.
[21] J. Koponen, T. Huttunen, T. Tarvainen, and J. P. Kaipio. Bayesian approximation error approach in full-wave ultrasound tomography. *IEEE T. Ultrason. Ferr.*, 61(10):1627–1637, 2014.
[22] P. Mathé. Bayesian inverse problems with non-commuting operators. arXiv preprint arXiv:1801.09540, 2018.
[23] G. D. Prato. *An Introduction to Infinite-Dimensional Analysis*. Springer, Berlin, 2006.
[24] G. D. Prato and J. Zabczyk. *Stochastic Equations in Infinite Dimensions*. Cambridge University Press, United Kingdom, second edition, 2014.
[25] M. A. Shubin and S. I. Andersson. *Pseudodifferential Operators and Spectral Theory*. Springer, Berlin, 1987.
[26] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numer.*, 19:451–559, 2010.
[27] B. Szabó, A. W. van der Vaart, and J. H. van Zanten. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *Ann. Statist.*, 43(4):1391–1428, 2015.
[28] M. Trabs. Bayesian inverse problems with unknown operators. *Inverse Probl.*, 34:085001, 2018.
[29] Sebastian J. Vollmer. Posterior consistency for Bayesian inverse problems through stability and regression results. *Inverse Probl.*, 29(12):125011, 2013.