

REGULARIZED POSTERIOR IN LINEAR ILL-POSED INVERSE PROBLEMS

Jean-Pierre Florens and Anna Simoni

Toulouse School of Economics

Abstract: We study the Bayesian solution of a signal-noise problem stated in infinite dimensional Hilbert spaces. The functional parameter of interest is characterized as the solution of a functional equation which is ill-posed because of compactness of the operator appearing in it. We show that the posterior distribution of the parameter of interest is inconsistent in the frequentist sense. This fact confirms the eventual frequentist inconsistency in Bayes nonparametric estimation pointed out, for instance, in Diaconis and Freedman (1986).

Our contribution is to propose a new method to deal with this problem: we regularize the posterior mean and variance by using a Tikhonov regularization scheme. The resulting distribution is called *regularized posterior distribution* and we prove it is consistent in a frequentist sense. Prior inconsistency issues are also discussed.

Key words and phrases: Bayesian estimation of density and regression, functional data, gaussian priors, inverse problems, posterior consistency, Tikhonov and Hilbert Scale regularization.

1. Introduction

Let \mathcal{X} and \mathcal{Y} be two infinite dimensional separable Hilbert spaces over \mathbb{R} and $x \in \mathcal{X}$ and $\hat{Y} \in \mathcal{Y}$ be two Hilbert-valued random functions. We want to obtain the Bayes estimate of the signal x from the observation \hat{Y} . The observed trajectory \hat{Y} is a linear noisy transformation of x through the statistical model

$$\hat{Y} = Kx + U, \quad U \in \mathcal{Y} \tag{1.1}$$

where U is a stochastic measurement error and $K : \mathcal{X} \rightarrow \mathcal{Y}$ is a known, compact, linear operator with infinite dimensional range. Its adjoint is denoted with K^* , then, by definition, K^* is such that $\langle K\varphi, \psi \rangle = \langle \varphi, K^*\psi \rangle$, $\forall \varphi \in \mathcal{X}$ and $\psi \in \mathcal{Y}$. The spaces \mathcal{X} and \mathcal{Y} are supposed to be Polish, with inner products and norms both denoted by $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$, respectively. We require Polish

spaces because this guarantees the existence of a regular version of the posterior distribution of x , even when the dimension of the problem is infinite. As an example of spaces, we could take \mathcal{X} and \mathcal{Y} to be both the L^2 space. An L^2 space, endowed with a gaussian measure defined on it, is a Polish space, see Hiroshi and Yoshiaki (1975).

The error term U is an Hilbert-valued gaussian random variable with zero mean and covariance operator Σ_n : $U \sim \mathcal{GP}(0, \Sigma_n)$, where n can be interpreted as the sample size. The Hilbert-valued random element x is supposed to induce a gaussian measure on \mathcal{X} : $x \sim \mathcal{GP}(x_0, \Omega_0)$, with $x_0 \in \mathcal{X}$ and $\Omega_0 : \mathcal{X} \rightarrow \mathcal{X}$.

A model of type (1.1) is encountered in many applications in different fields. For instance, in statistical field, applications are *e.g.* density and regression estimation or gaussian white noise model; in signal and image processing, example of applications are image deblurring or extrapolation of a band- or time-limited signal.

Recovering the solution of (1.1) is known in the literature as solving an *inverse problem* and such a problem is said to be *well-posed* when a unique solution exists and depends continuously on the data. When one of these requirements fails to hold we say the inverse problem to be *ill-posed*. While lack of uniqueness or existence can be easily dealt with, for instance by adopting a generalized inverse, the lack of continuity of the solution is more troublesome since it prevents convergence of the recovered solution towards the true one as the noise in the data \hat{Y} decreases. The lack of continuity of the solution is due to the inversion of a compact operator K on an infinite dimensional space \mathcal{Y} . The inverse K^{-1} is not always defined and not continuous on the whole \mathcal{Y} so that some regularization of this inverse is demanded.

Classical regularization techniques are well developed in the literature, see Kress (1999). In reverse, in this paper we focus on Bayesian methods for estimating the solution of the ill-posed inverse problem (1.1). From a Bayesian point of view, the solution to an inverse problem is the posterior distribution of the quantity of interest, therefore the ill-posedness linked to the inversion of K is overcome. This reformulation of an inverse problem as a parameter estimation is due to Franklin (1970).

The posterior distribution suffers of problems too. Indeed, the infinite dimension

of the problem prevents the posterior distribution from being consistent in the frequentist sense. If problem (1.1) was formulated in finite dimensional spaces then its Bayes solution would be the standard multivariate gaussian posterior distribution of x given \hat{Y} and it would be consistent. Otherwise stated, in finite dimensional inverse problems it is possible to remove the ill-posedness (*i.e.* the ill-conditioning of a square matrix, for instance) by incorporating the available prior information, see Kaipio and Somersalo (2004, Chap.3). This is no longer true when dimension is infinite since the covariance operator $Var(\hat{Y})$ is no longer continuously invertible, so covariance operators do not have the regularization properties that have in the finite dimensional case. This prevent the posterior mean from being continuous in \hat{Y} and then from being a consistent estimator of x . Hence, the posterior distribution is not consistent.

This problem has been solved in past literature by restricting the space of definition of $\hat{Y} - Kx_0$, see Mandelbaum (1984), Prenter and Vogel (1985) and Lehtinen, Paivarinta and Somersalo (1989). However this solution is not always appropriate since the observed data may not satisfy this restriction.

Our contribution consists in dealing with the lack of continuity of $Var(\hat{Y})^{-1}$ by applying a regularization scheme to this inverse. We consider two alternative regularization schemes. The first one is the classical Tikhonov regularization scheme: $(\alpha_n I + Var(\hat{Y}))^{-1}$, with I the identity operator on \mathcal{Y} and $\alpha_n > 0$ the regularization parameter; the second one is a Tikhonov regularization scheme in the Hilbert scale induced by the prior covariance operator Ω_0 : $(\alpha_n L^{2s} + Var(\hat{Y}))^{-1}$, with $L = \Omega_0^{-\frac{1}{2}}$ and $s \in \mathbb{R}$. The posterior distribution that results with each of these regularization schemes is slightly modified and we call it *regularized posterior distribution*. We analyze its asymptotic properties from a frequentist point of view and we prove posterior consistency, see Diaconis and Freedman (1986) or Section 4 for a definition of it. The rate of contraction of the moments of each regularized posterior distribution are computed and they result to be considerably improved with the second regularization scheme. Moreover, we compute the rate of contraction of the Tikhonov regularized posterior distribution.

The regularization that we introduce is justified as a regularized projection and cannot be interpreted as resulting from a prior specification. Another strategy to deal with the inconsistency of the posterior distribution would be to restrict our

model to models where K and Σ_n are linked and where the prior has a specific form which concentrates to a Dirac measure at a suitable speed. This analysis has been done in Florens and Simoni (2009) but we consider here the general case.

The paper is developed as follows. Section 2 characterizes the Bayes experiment associated to (1.1) and provides some example of application. In Section 3 we define the *regularized posterior distribution* for both the regularization schemes; its consistency is proved in Section 4. All the proofs are given in Appendix A and numerical simulations are provided in Appendix B.

This paper analyzes the very general case where both operators K and Σ_n are known. Extensions to the case where they are unknown require minor modifications in the proofs of consistency and are treated in Simoni (2009, Chap.1).

2. The Model

2.1 Sampling probability measure and examples

Quantities \hat{Y} , x and U in equation (1.1) are Hilbert-random variables. Let \mathcal{F} denote the σ -field of subsets of the sample space \mathcal{Y} , we endow the measurable space $(\mathcal{Y}, \mathcal{F})$ with the sampling distribution $\mathbb{P}(\hat{Y}|x)$ of \hat{Y} given x , denoted with P^x and characterized by Assumption 1 below.

Assumption 1. *Let P^x be a conditional probability measure on $(\mathcal{Y}, \mathcal{F})$ given x such that $\mathbb{E}(\|\hat{Y}\|^2) < \infty$, $\hat{Y} \in \mathcal{Y}$. P^x is a gaussian measure that defines a mean element $Kx \in \mathcal{Y}$ and a covariance operator $\Sigma_n : \mathcal{Y} \rightarrow \mathcal{Y}$.*

For a characterization of gaussian measures in Hilbert spaces we refer to Baker (1973). Assumption 1 implies that the covariance operator Σ_n is linear, bounded, nonnegative, selfadjoint and trace-class. A covariance operator needs to be trace-class in order the associated measure be able to generate trajectories belonging to an Hilbert space, therefore the covariance operator cannot be proportional to the identity operator. The fact that Σ_n is trace-class entails that $\Sigma_n^{\frac{1}{2}}$ is Hilbert-Schmidt (HS, hereafter). HS operators are compact and compactness of $\Sigma_n^{\frac{1}{2}}$ implies compactness of Σ_n .

The covariance operator Σ_n is supposed to be known and to decrease to 0 as the noise in the data \hat{Y} decreases. The measurement error is in several applications

inversely linked to the sample size n , then we can write $\Sigma_n \rightarrow 0$ as $n \rightarrow \infty$. This is true, for instance when the curve \hat{Y} is obtained through a mathematical transformation of an n -sample of finite dimensional observations (like the empirical cumulative distribution function, the empirical characteristic function or the Nadaraya-Watson estimator of a regression function) or when \hat{Y} is the sample mean of functional data (like many examples in signal and image processing). So, our methodology allows for very general observational schemes.

We stay as general as possible since we do not require that operator Σ_n be linked in some way to operator K in the sampling mechanism. If this was the case, then we could exploit the regularity in operator K and our problem would be greatly simplified. We have studied this situation in Florens and Simoni (2009).

The Bayes approach that we develop can be used for all the classical examples where theory of linear inverse problems applies. Statistics and econometrics offer several examples of applications, see Vapnik (1998) and Carrasco, Florens and Renault (2007); we develop some of them below.

EXAMPLE 1: *Inverse problems in image science.* Let suppose that we observe n curves \tilde{Y}_i independently generated from the model

$$\tilde{Y}_i = Kx + U_i, \quad U_i \sim \mathcal{GP}(0, \Sigma).$$

The empirical mean is then a sufficient statistics for doing inference on x . We compute $\hat{Y} = \frac{1}{n} \sum_i \tilde{Y}_i$ and we rewrite $\hat{Y} = Kx + U$, with $U = \frac{1}{n} \sum_i U_i$ and $\text{Var}(U) = \frac{1}{n} \Sigma$. This example is often encountered in image science. The covariance operator is usually known in image science, but there exist cases (like **EXAMPLE 2** below) where Σ is unknown. In this situation we can either estimate Σ in a frequentist way and develop the asymptotic theory in a similar way as developed in Simoni (2009, Chap.1) or extend the Inverse-Wishart prior distribution to covariance operators and develop a fully Bayesian estimation procedure (this point is in our research agenda).

EXAMPLE 2: *Density estimation.* Let $\mathcal{X} = \mathcal{Y} = L^2_\pi([0, 1])$ be the spaces of square integrable functions on $[0, 1]$, integrable with respect to the uniform distribution π . We want to recover the density $f \in \mathcal{X}$ associated to the distribution F of the random variable ξ from an *i.i.d.* sample (ξ_1, \dots, ξ_n) drawn from F . Let

$\hat{F}_n(\bar{\xi}) = \frac{1}{n} \sum_{i=1}^n 1\{\xi_i \leq \bar{\xi}\}$, then f can be obtained by solving the functional equation

$$\hat{F}_n(\bar{\xi}) = \int_0^1 f(u) 1\{u \leq \bar{\xi}\} du + U_n,$$

The sampling probability P^f is inferred from asymptotic properties of the empirical distribution function, so that it is asymptotically a gaussian measure with mean F and covariance operator $\Sigma_n = \frac{1}{n} \int_0^1 F(u \wedge v) - F(u)F(v) du$. Even if the error term is only asymptotically gaussian, the estimation method that we propose remains valid since the gaussianity is only used to construct the estimator and not to prove the result of consistency of the Bayes estimator or of the posterior distribution.

EXAMPLE 3: Regression estimation. Let (ξ, w) be a \mathbb{R}^{1+p} -valued random vector with cdf F and $L_F^2(w)$ be the space of square integrable functions of w , integrable with respect to F . We define the regression function of ξ given w as a function $m(w) \in L_F^2(w)$.

Let $g(w, t) : \mathbb{R}^p \times \mathbb{R}^p \rightarrow \mathbb{R}$ be a known function defining an HS integral operator with respect to (w, t) , where t belongs to \mathbb{R}^p provided with a suitable probability measure. Then $\mathbb{E}(g(w, t)\xi) = \mathbb{E}(g(w, t)m(w))$, where the expectation is taken with respect to F , and $m(w)$ is the solution of a linear inverse problem. Take for simplicity $F(\xi|w)$ unknown and $F(\cdot, w)$ known and suppose to dispose of a random sample (ξ_i, w_i) . Then $\hat{\mathbb{E}}(g(w, t)\xi) := \frac{1}{n} \sum_{i=1}^n g(w_i, t)\xi_i$ and the statistical inverse problem becomes $\hat{\mathbb{E}}(g(w, t)\xi) = \mathbb{E}(g(w, t)m(w)) + U_n(t)$. The empirical process $\sqrt{n}(\hat{\mathbb{E}}(g(w, t)\xi) - \mathbb{E}(g(w, t)\xi))$ weakly converges toward a zero mean gaussian process and this characterizes the sampling distribution.

Other examples of application are for instance *hazard rate function estimation with right-censored survival data*, *deconvolution*, *instrumental regression estimation*. A brief development of them can be found in the Appendix B of Chapter 1 of Simoni (2009).

2.2 Prior Specification and Identification

In the following we denote with $\mathcal{R}(\cdot)$ the range of an operator and with $\mathcal{D}(\cdot)$ its domain. Let μ denote the prior measure induced by x on the parameter space \mathcal{X} endowed with the σ -field \mathcal{E} . We specify a conjugate prior:

Assumption 2. Let μ be a gaussian measure on $(\mathcal{X}, \mathcal{E})$ that defines a mean element $x_0 \in \mathcal{X}$ and a covariance operator $\Omega_0 : \mathcal{X} \rightarrow \mathcal{X}$ that is trace-class.

Then, $\mathbb{E}(\|x\|^2) < \infty$, $\forall x \in \mathcal{X}$ and the covariance operator Ω_0 is compact. The covariance operator Ω_0 is assumed to be fixed and is not shrinking to 0. This would be the case, for instance, when Ω_0 is an inverse function of the sample size, as for the Zellner's g -prior, see Florens and Simoni (2009).

We introduce the *Reproducing Kernel Hilbert Space* ($\mathcal{R.K.H.S.}$ in the following) associated to the covariance operator Ω_0 and denoted with $\mathcal{H}(\Omega_0)$. Let $\{\lambda_j^{\Omega_0}, \varphi_j^{\Omega_0}\}_j$ be the eigensystem of Ω_0 . We define the space $\mathcal{H}(\Omega_0)$ embedded in \mathcal{X} as:

$$\mathcal{H}(\Omega_0) = \{\varphi : \varphi \in \mathcal{X} \quad \text{and} \quad \sum_{j=1}^{\infty} \frac{|\langle \varphi, \varphi_j^{\Omega_0} \rangle|^2}{\lambda_j^{\Omega_0}} < \infty\} \quad (2.1)$$

and, following Proposition 3.6 in Carrasco, Florens and Renault (2007), $\mathcal{H}(\Omega_0) = \mathcal{R}(\Omega_0^{\frac{1}{2}})$. The $\mathcal{R.K.H.S.}$ is a subset of \mathcal{X} that gives the geometry of the distribution of x . The support of a centered gaussian process, taking its values in an Hilbert space \mathcal{X} , is the closure in \mathcal{X} of the $\mathcal{R.K.H.S.}$ associated with the covariance operator of this process (denoted with $\overline{\mathcal{H}(\Omega_0)}$ in our case). Then, for the prior distribution, $x - x_0 \in \overline{\mathcal{H}(\Omega_0)}$ with μ -probability 1, but, with μ -probability 1, $x - x_0$ is not in $\mathcal{H}(\Omega_0)$, see van der Vaart and van Zanten (2008a).

We adopt a frequentist perspective for studying our procedure, then we admit the existence of a true value x_* , of the parameter x , having generated the data \hat{Y} and we assume that

Assumption 3. $(x_* - x_0) \in \mathcal{H}(\Omega_0)$, i.e. there exists $\delta_* \in \mathcal{X}$ such that $(x_* - x_0) = \Omega_0^{\frac{1}{2}} \delta_*$.

This assumption is only a regularity condition and it will be exploited for proving asymptotic results. For instance, when the kernel of Ω_0 is the variance of a standard Brownian motion in $\mathcal{C}[0, 1]$, the $\mathcal{R.K.H.S.}$ is the space of absolutely continuous functions f on $[0, 1]$ with at least one square integrable derivative and such that $f(0) = 0$, see Carrasco and Florens (2000) and van der Vaart and van Zanten (2008b).

The discussion just before implies that the prior distribution is not able to generate a trajectory x that satisfies Assumption 3 or, in other words, the true value x_*

having generated \hat{Y} cannot have been drawn from μ . Anyway, if Ω_0 is injective, even if μ puts zero probability on $\mathcal{H}(\Omega_0)$, this space is dense in \mathcal{X} and therefore μ can generate trajectories as close as possible to the true value x_* . We find a similar result for a Dirichlet process, in nonparametric probabilities estimation, in the sense that it puts zero probability mass on absolutely continuous probability measures but it is able to generate probability functions close to them. This kind of problem is known as *prior inconsistency* and it is due to the fact that, because of the infinite dimensionality of the parameter space, the support of the prior can cover only a very "small" part of it.

From a Bayesian point of view we say that a model is identified if the posterior distribution completely revises the prior distribution, for what we do not need to introduce strong assumptions, see Florens, Mouchart and Rolin (1990) Section 4.6 for an exhaustive explanation of this concept. Nevertheless, this paper focuses on the frequentist consistency of the posterior distribution and for that we need the following assumption for identification (see Section 4 below).

Assumption 4. *The operator $K\Omega_0^{\frac{1}{2}} : \mathcal{X} \rightarrow \mathcal{Y}$ is one-to-one on \mathcal{X} .*

This assumption guarantees continuity of the *regularized posterior mean* that we define below.

The classical hypothesis for identification of x in model (1.1) requires that K be one-to-one. This is a stronger condition since, if $\Omega_0^{\frac{1}{2}}$ is one-to-one, K one-to-one implies $K\Omega_0^{\frac{1}{2}}$ one-to-one, but the reverse is not true. Therefore, frequentist consistency in a Bayesian model requires a weaker identification condition than a classical model does.

2.3 Construction of the Bayesian Experiment

The relevant probability space associated to (1.1) is the real linear product space $\mathcal{X} \times \mathcal{Y}$ defined as the set $\mathcal{X} \times \mathcal{Y} := \{(x, y); x \in \mathcal{X}, y \in \mathcal{Y}\}$ with addition, scalar multiplication and scalar product defined in the usual way. The product σ -field associated to $\mathcal{X} \times \mathcal{Y}$ is denoted with $\mathcal{E} \otimes \mathcal{F}$ and the probability measure defined on $(\mathcal{X} \times \mathcal{Y}, \mathcal{E} \otimes \mathcal{F})$ is denoted with Π and constructed by recomposing μ and P^x .

The marginal distribution of \hat{Y} , obtained by integrating out x with respect to the prior distribution, is denoted with P and its covariance operator is $\Upsilon_{yy} =$

$(\Sigma_n + K\Omega_0 K^*)$. We denote with Υ the covariance operator associated to Π defined as $\Upsilon(\varphi, \psi) = (\Omega_0 \varphi + \Omega_0 K^* \psi, (\Sigma_n + K\Omega_0 K^*)\psi + K\Omega_0 \varphi)$, for all $(\varphi, \psi) \in \mathcal{X} \times \mathcal{Y}$.

Lemma 1. *The covariance operators Υ and Υ_{yy} are trace class. In particular, Υ_{yy} trace class is a necessary condition for Υ being trace class.*

Next, we state that the joint and predictive probabilities, Π and P , are gaussian.

Theorem 1.

- (i). *Under Assumptions 1 and 2, the joint measure Π on $(\mathcal{X} \times \mathcal{Y}, \mathcal{E} \otimes \mathcal{F})$ is gaussian with mean function $m_{xy} = (x_0, Kx_0) \in \mathcal{X} \times \mathcal{Y}$ and covariance operator Υ .*
- (ii). *Let P be a gaussian measure on $(\mathcal{Y}, \mathcal{F})$ with mean function $m_y = Kx_0$ in \mathcal{Y} and covariance operator Υ_{yy} . Then, P is the marginal distribution on $(\mathcal{Y}, \mathcal{F})$ associated to the joint gaussian measure Π defined in (i).*

The aim of this paper will be to determine the inverse decomposition of Π into the marginal P and the posterior distribution $\mu^{\hat{Y}}$, the conditional distribution of x given \hat{Y} . Existence of this inverse decomposition is ensured if a regular version of the posterior probability exists.

3. Bayes solution of the ill-posed inverse problem

Due to the infinite dimension of problem (1.1), application of Bayes theorem is not evident and in computing the posterior distribution three points require a particular attention: (i) the existence of a regular version of the conditional probability on \mathcal{E} given \mathcal{F} , (ii) the fact that it is a gaussian measure and (iii) continuity of the posterior mean and posterior consistency.

- (i). The conditional probability on \mathcal{E} given \mathcal{F} exists and it is unique since it is the projection on a closed convex subset of $L^2(\mathcal{X} \times \mathcal{Y})$, where $L^2(\mathcal{X} \times \mathcal{Y})$ is the Hilbert space of random variables defined on $\mathcal{X} \times \mathcal{Y}$ that are square integrable with respect to Π . A conditional probability is called *regular* if there exists a transition probability characterizing it. The existence of such a transition for $\mu^{\hat{Y}}$ is guaranteed by *Jirina Theorem* if the space $(\mathcal{X} \times \mathcal{Y})$ is Polish, see Neveu (1965).
- (ii). By slightly modifying the proof given in Section 2.2 of Mandelbaum (1984) it is easy to show that $\mu^{\hat{Y}}$ is gaussian since the associated characteristic function

takes the form

$$\mathbb{E}(e^{i\langle x, h \rangle} | \hat{Y}) = e^{i\langle A\hat{Y} + b, h \rangle - \frac{1}{2}\langle (\Omega_0 - AK\Omega_0)h, h \rangle}, \quad h \in \mathcal{X}.$$

Then, $x|\hat{Y}$ has mean: $A\hat{Y} + b$, and variance $V = \Omega_0 - AK\Omega_0$. The function $b := (I - AK)x_0$ is recovered from $\mathbb{E}(x) = \mathbb{E}(\mathbb{E}(x|\hat{Y}))$ and A is characterized by exploiting the definition of covariance operator, for which we have $\langle \text{Cov}(\hat{Y}, x)\varphi, \psi \rangle = \langle (\Sigma_n + K\Omega_0 K^*)A^*\varphi, \psi \rangle$, $\forall \varphi \in \mathcal{X}, \psi \in \mathcal{Y}$ and $\text{Cov}(\hat{Y}, x) = K\Omega_0$ is a component of operator Υ determined in Theorem 1. Hence, $A : \mathcal{Y} \rightarrow \mathcal{X}$ is solution of

$$A(\Sigma_n + K\Omega_0 K^*)\psi = \Omega_0 K^* \psi, \quad \psi \in \mathcal{Y} \quad (3.1)$$

and then $A = \Omega_0 K^* (\Sigma_n + K\Omega_0 K^*)^{-1}$.

(iii). This expression for A is not well-defined since $(\Sigma_n + K\Omega_0 K^*)$ is a compact operator with infinite range so that its inverse is not continuous on the whole \mathcal{Y} and the posterior mean is not continuous in \hat{Y} . Thus, the posterior mean and the posterior distribution are inconsistent in the frequentist sense (but consistency in the Bayes sense is still verified). Actually, Bayes approach to (1.1), by changing the nature of the problem, changes the nature of the ill-posedness. Here, we have to deal with the ill-posedness in the inverse problem (3.1) that characterizes the inconsistency of the posterior distribution. Diaconis and Freedman (1986) stress that posterior inconsistency is frequent in nonparametric Bayes experiments.

Past literature on Bayesian inverse problems, see Mandelbaum (1984) and Lehtinen, Paivarinta and Somersalo (1989), proposed, as strategy to solve this problem of non-continuity, to restraint the space of the observable \hat{Y} . It was implicitly assumed that \hat{Y} belongs to $\mathcal{R}(\Sigma_n + K\Omega_0 K^*)$ or to a subspace of it. We do not wish to make this kind of restriction since we admit any trajectory \hat{Y} in $\overline{\mathcal{R}(\Sigma_n + K\Omega_0 K^*)}$. Thus, a different strategy, based on Tikhonov regularization, will be proposed in the next paragraph.

3.1 Tikhonov Regularized Posterior distribution

We propose to solve the problem of unboundedness of A by applying a Tikhonov regularization scheme to equation (3.1). By abuse of notation, hereafter we use I for denoting the identity operator on both \mathcal{X} and \mathcal{Y} . We define

the Tikohnov regularized operator A_α as:

$$A_\alpha = \Omega_0 K^* (\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1} \quad (3.2)$$

where $\alpha_n > 0$ is a regularization parameter appropriately chosen such that $\alpha_n \rightarrow 0$ with n .

We could interpret A_α as the operator that we would obtain if we have considered the new Bayesian experiment $\hat{Y} = Kx + U + \eta$, with η a further error term with variance $\alpha_n I$. In this case the sampling distribution would characterize a covariance operator $(\alpha_n I + \Sigma_n)$ that is not trace-class so that the trajectories generated by this distribution would not be in the Hilbert space \mathcal{Y} . Even if this interpretation gives a model that is not well specified in \mathcal{Y} , it is interesting since it could suggest a Bayes method for selecting the regularization parameter through the specification of a prior distribution on α_n .

The regularized versions of b and V , with A replaced by A_α are

$$\begin{aligned} b_\alpha &= (I - A_\alpha K)x_0, \\ V_\alpha &= \Omega_0 - \Omega_0 K^* (\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1} K \Omega_0. \end{aligned} \quad (3.3)$$

These regularized objects characterize a new distribution that is gaussian with mean $(A_\alpha \hat{Y} + b_\alpha)$ and covariance operator V_α ; it is trivial to show that V_α is trace-class. This distribution is called *regularized posterior distribution* and is denoted with $\mu_\alpha^{\hat{Y}}$. It is a new object that we define to be the solution of the signal-noise problem and that will be shown to be consistent in Section 4. The idea of regularizing a distribution consists therefore in regularizing the moments characterizing it. We keep as point estimator of x the regularized posterior mean

$$\mathbb{E}_\alpha(x|\hat{Y}) = x_0 + \Omega_0 K^* (\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1} (\hat{Y} - Kx_0). \quad (3.4)$$

This estimator is justified since it minimizes the penalized mean squared error obtained by approximating x by a linear transformation of \hat{Y} . Otherwise stated, the bounded linear operator $A_\alpha : \mathcal{Y} \rightarrow \mathcal{X}$ is the unique solution to the problem:

$$A_\alpha = \operatorname{argmin}_{A \in \mathfrak{B}_2(\mathcal{Y}, \mathcal{X})} \mathbb{E} \|A\hat{Y} - x\|^2 + \alpha_n \|A\|_{HS}^2$$

where $\|A\|_{HS}^2 := \operatorname{tr} A^* A$ denotes the HS norm, $\mathfrak{B}_2(\mathcal{Y}, \mathcal{X})$ the set of all bounded operators on \mathcal{Y} to \mathcal{X} for which $\|A\|_{HS} < \infty$ and for simplicity we have set $x_0 = 0$.

The penalization is required because otherwise the solution to the minimization problem would be unbounded.

3.2 Tikhonov regularization in the Prior Variance Hilbert scale (PVHS)

We propose in this subsection an alternative regularization scheme for recovering A . It is a *Tikhonov regularization in the Hilbert scale induced by the inverse of the prior covariance operator*, see Engl, Hanke and Neubauer (2000) for general theory, and it is appealing when we know that x_* is highly regular, as under Assumption 5 (ii) below. Let $L = \Omega_0^{-\frac{1}{2}}$ be a densely defined unbounded self-adjoint strictly positive operator in the Hilbert space \mathcal{X} . More clearly, if $\mathcal{D}(L)$ denotes the domain of L , L is a closed operator in \mathcal{X} satisfying: $\mathcal{D}(L) = \mathcal{D}(L^*)$ is dense in \mathcal{X} , $\langle Lx, y \rangle = \langle x, Ly \rangle$ for all $x, y \in \mathcal{D}(L)$, and there exists $\gamma > 0$ such that $\langle Lx, x \rangle \geq \gamma \|x\|^2$ for all $x \in \mathcal{D}(L)$. The norm $\|\cdot\|_s$ is defined as $\|x\|_s := \|L^s x\|$. The Hilbert Scale \mathcal{X}_s induced by L is defined as the completion of the domain of L^s , $\mathcal{D}(L^s)$, with respect to the norm $\|\cdot\|_s$, see Krein and Petunin (1966); moreover $\mathcal{X}_s \subseteq \mathcal{X}_{s'}$ if $s' \leq s$, $\forall s \in \mathbb{R}$. Usually, when a regularization scheme in Hilbert Scale is adopted, the operator L , and consequently the Hilbert Scale, is created *ad hoc*. In our case the Hilbert Scale is not created ad-hoc but is suggested by the prior distribution and this represents a considerable difference and advantage with respect to standard methods. We make the following Assumption:

Assumption 5.

- (i) $\|K\Omega_0^{\frac{1}{2}}x\| \sim \|\Omega_0^{\frac{a}{2}}x\|$, $\forall x \in \mathcal{X}$;
- (ii) $(x_* - x_0) \in \mathcal{X}_{\beta+1}$, i.e. $\exists \rho_* \in \mathcal{X}$ such that $(x_* - x_0) = \Omega_0^{\frac{\beta+1}{2}} \rho_*$
- (iii) $0 < a \leq s \leq \beta + 1 \leq 2s + a$.

Assumption (i) is necessary in order the regularization in Hilbert Scale works. It means that the specification of the prior distribution is related to the sampling model, so the prior variance is linked to the sampling model (1.1) and, in particular, to operator K . This kind of prior specification is not new in Bayesian literature since it is similar to the Zellner's *g-prior*, see Zellner (1986) or Agliari, Parisetti (1988). Parameter a can be interpreted as the *degree of ill-posedness* in the Bayesian experiment. It is usually different than the degree of ill-posedness

in the classical problem $\hat{Y} = Kx$ since it is determined by the rate of decreasing of the spectrum of operator $K\Omega_0^{\frac{1}{2}}$ and not by that one of K . Therefore, the prior is specified not only by taking into account the sampling model but also the degree of ill-posedness of the problem.

Assumption (ii) is known as a *source condition*; it concerns the regularity of x_* and it allows to reach a faster speed of convergence of the regularized solution the larger is β . We have $\mathcal{X}_{\beta+1} \equiv \mathcal{R}(\Omega_0^{\frac{\beta+1}{2}}) \subset \mathcal{R}(\Omega_0^{\frac{1}{2}})$, then Assumption 5 (ii) implies Assumption 3 and $\delta_* \in \mathcal{R}(\Omega_0^{\frac{\beta}{2}})$. The meaning of such an assumption is that the prior distribution contains information about the regularity of the true value of x . In fact, parameter β is interpreted as the *regularity parameter*. These two remarks stress the fact that we are not taking whatever Hilbert Scale, but the Hilbert Scale linked to the prior. Either we first choose the Hilbert Scale and then we use the information contained in it to specify the prior distribution or we use the information contained in the prior distribution to specify the Hilbert Scale.

The restriction $\beta + 1 \geq s$ means that $(x_* - x_0)$ has to be at least an element of \mathcal{X}_s and it guarantees that the norm $\|L^s x\|$ exists $\forall x \in \mathcal{X}_{\beta+1}$. The upper bound $(2s + a)$ of β is the qualification of this regularization scheme: we can at most exploit a degree of regularity of $(x_* - x_0)$ equal to $(2s + a)$.

Under Assumption 5, the Tikhonov regularized solution in \mathcal{X}_s to equation (3.1) is:

$$A_s = \Omega_0 K^* (\alpha_n L^{2s} + \Sigma_n + K \Omega_0 K^*)^{-1}. \quad (3.5)$$

The regularized posterior distribution is thus defined similarly as in Section 3.1 with A_α substituted by A_s and is denoted with $\mu_s^{\hat{Y}}$. The regularized posterior mean and variance are

$$\begin{aligned} \mathbb{E}_s(x|\hat{Y}) &= A_s \hat{Y} + (I - A_s K)x_0 \\ V_s &= \Omega_0 - A_s K \Omega_0. \end{aligned} \quad (3.6)$$

This regularization method has the advantage that it permits to better exploit the regularity of the true function x_* . If x_* satisfies Assumption 5 (ii), then $\mathbb{E}_s(x|\hat{Y})$ satisfies the same regularity, see proof of Theorem 5, while $\mathbb{E}_a(x|\hat{Y})$ does not. Moreover, a classical Tikhonov regularization method allows to obtain a rate of convergence to zero of the regularization bias that is at most of order

2; on the contrary with a Tikhonov scheme in an Hilbert Scale the smoother the function x_* is, the faster the rate of convergence to zero of the regularization bias will be (up to $(2s + a)$). We will show in Section 4.1 that $\mathbb{E}_s(x|\hat{Y})$ reaches a faster speed of convergence toward x_* .

4. Asymptotic Analysis

In this section we study frequentist asymptotic properties of the regularized posterior mean and variance. Then, we analyze frequentist consistency of the regularized posterior distribution. We first consider the Tikhonov regularized posterior distribution $\mu_\alpha^{\hat{Y}}$ and then, in subsection 4.1, we analyze the moments of the PVHS-regularized posterior distribution $\mu_s^{\hat{Y}}$.

Let consider the regularized posterior mean as a point estimator for x_* , as suggested by a penalized squared loss function. We denote with Φ_β the β -regularity space associated to operator $K\Omega_0^{\frac{1}{2}}$, i.e. $\Phi_\beta \equiv \mathcal{R}(\Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$ for some $\beta > 0$. The convergence of $\|\mathbb{E}_\alpha(x|\hat{Y}) - x_*\|$ to 0 in P^{x_*} -probability when $n \rightarrow \infty$ and the rate of contraction are stated in the following theorem. We refer to Appendix A for its proof.

Theorem 2. *Under Assumptions 3 and 4, if $\alpha_n \rightarrow 0$, $\frac{1}{\alpha_n} \text{tr} \Sigma_n \rightarrow 0$ and $\frac{1}{\alpha_n^3} \|\Sigma_n\|^2 \sim \mathcal{O}_p(1)$, then $\|\mathbb{E}_\alpha(x|\hat{Y}) - x_*\| \rightarrow 0$ in P^{x_*} -probability. Moreover, if $\delta_* \in \Phi_\beta$, for some $\beta > 0$, the MISE is of order*

$$\mathbb{E} \|\mathbb{E}_\alpha(x|\hat{Y}) - x_*\|^2 = \mathcal{O}_p \left(\alpha_n^{\beta \wedge 2} + \frac{1}{\alpha_n^4} \|\Sigma_n\|^2 \alpha_n^{(\beta+1) \wedge 2} + \frac{1}{\alpha_n} \text{tr} \Sigma_n \right).$$

The larger β is, the smoother the function $\delta_* \in \Phi_\beta$ will be and faster the regularization bias will converge to zero. However, for a Tikhonov regularization scheme, β cannot be greater than 2 (this is the reason why we bound it by 2 in α_n^β). This is known as *saturation effect*, see section 4.2 in Engl, Hanke and Neubauer (2000); then with classical Tikhonov regularization scheme it is useless to have a function x_* with a degree of smoothness larger than 2. In the remaining of this section, even if we do not explicitly write $\beta \wedge 2$, it must be understood that if $\beta > 2$ it has to be set at 2.

Condition $\frac{1}{\alpha_n^3} \|\Sigma_n\|^2 \sim \mathcal{O}_p(1)$ is sufficient to guarantee that $\frac{1}{\alpha_n^4} \|\Sigma_n\|^2 \alpha_n^{(\beta+1) \wedge 2} \rightarrow 0$. If $\text{tr} \Sigma_n$ is of the same order as $\|\Sigma_n\|$, for instance $\text{tr} \Sigma_n \sim \|\Sigma_n\| \sim \mathcal{O}_p(\frac{1}{n})$, then an α_n satisfying $\alpha_n \rightarrow 0$ and $\alpha_n^{\frac{3}{2}} n \rightarrow \infty$ guarantees convergence to zero of the

second and third rates in the MISE. Classical conditions for convergence of the solution of stochastic ill-posed problems are $\alpha_n \rightarrow 0$ and $\alpha_n^2 n \rightarrow \infty$ (see Vapnik (1998)). Therefore, we require weaker conditions.

If $\text{tr}\Sigma_n \sim \|\Sigma_n\|$, the fastest global rate of convergence is obtained when $\alpha_n^\beta = \frac{1}{\alpha_n} \|\Sigma_n\|$. Then, the optimal regularization parameter α_n^* is proportional to

$$\alpha_n^* \propto \|\Sigma_n\|^{\frac{1}{\beta+1}}$$

and the optimal speed of convergence of the MISE is proportional to $\|\Sigma_n\|^{\frac{\beta}{\beta+1}}$. With the optimal value α_n^* , the condition $\frac{1}{\alpha_n^3} \|\Sigma_n\|^2 \sim \mathcal{O}_p(1)$ is satisfied for $\beta \geq \frac{1}{2}$. Assuming the trace and the norm of the covariance operator be of the same order is not really stringent. For instance, in almost all real examples they are both of order $\frac{1}{n}$.

As Corollary 1 below shows, a faster rate of convergence for $\mathbb{E}_\alpha(x|\hat{Y})$ can be obtained if we add more conditions on Σ_n and on its spectrum.

Corollary 1. *If $\Sigma_n = \frac{1}{n}\Sigma$, where Σ has the same eigenfunctions as $K\Omega_0 K^*$ and under the assumption that $\Sigma(K\Omega_0 K^*)^{-\gamma}$ is trace-class for $\gamma \in [0, \frac{3\beta-1}{2} \wedge 1]$, then the optimal rate of $\mathbb{E}\|\mathbb{E}_\alpha(x|\hat{Y}) - x_*\|^2$ is $n^{-\frac{\beta}{\beta+1-\gamma}}$ and the corresponding optimal α_n^* is proportional to $n^{-\frac{1}{\beta+1-\gamma}}$.*

We proceed to study the convergence of the regularized posterior variance operator when it is applied to an element $\varphi \in \mathcal{X}$. Furthermore, we compute the rate of convergence of a restriction of V_α to a subset of its domain.

Theorem 3. *Under Assumption 4, if $\alpha_n \rightarrow 0$ and $\frac{1}{\alpha_n^3} \|\Sigma_n\|^2 \sim \mathcal{O}_p(1)$ then $\forall \varphi \in \mathcal{X}$, $\|V_\alpha \varphi\| \rightarrow 0$. Moreover, if the posterior variance is applied to $\varphi \in \mathcal{X}$ such that $\Omega_0^{\frac{1}{2}} \varphi \in \mathcal{R}(\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$, for some $\beta > 0$, it is of order*

$$\|V_\alpha \varphi\|^2 = \mathcal{O}(\alpha_n^\beta + \frac{1}{\alpha_n^4} \|\Sigma_n\|^2 \alpha_n^{(\beta+1)\wedge 2}).$$

With the optimal α_n^* (optimal for $\mathbb{E}_\alpha(x|\hat{Y})$), under the conditions in the above theorem and for $\beta \geq \frac{1}{2}$, the rate of $\|V_\alpha \varphi\|^2$ is $\|\Sigma_n\|^{\frac{\beta}{\beta+1}}$.

We wish to compare the optimal rate of convergence of $\mathbb{E}_\alpha(x|\hat{Y})$, called *Bayes estimator* hereafter, with the rate of the classical Tikhonov solution of (1.1), i.e. $x_\alpha = (\alpha_n I + K^* K)^{-1} K^* \hat{Y}$, that is suggested by the classical literature on inverse

problems and that will be called *classical estimator*. We refer to Engl, Hanke and Neubauer (2000) and Carrasco, Florens and Renault (2007) for a review of the classical method. For simplicity, we set $x_0 = 0$. To make this comparison possible we have to consider the particular case: $\Omega_0 = c_1(K^*K)^b$, with $b > 0$ and c_1 a constant of proportionality. In this particular case we show that the fastest rate of convergence of $\mathbb{E}_\alpha(x|\hat{Y})$ is slower than the rate of convergence of x_α . The regularity condition required by the classical method is $x_* \in \mathcal{R}((K^*K)^{\frac{b}{2}})$, for some $b > 0$, and the optimal speed of convergence is $(\text{tr}\Sigma_n)^{\frac{b}{b+1}}$, with $b \leq 2$ or b set equal to 2 if $b \geq 2$. Therefore, if we choose β in order to have the same regularity condition, *i.e.* $\mathcal{R}((K^*K)^{\frac{(b+1)\beta}{2}}) = \mathcal{R}((K^*K)^{\frac{b}{2}})$ and then $\beta = \frac{b}{b+1}$, the fastest rate of convergence of the Bayes estimator is proportional to $(\text{tr}\Sigma_n)^{\frac{b}{2b+1}}$ that is slower than the rate of x_α . This result is due to the fact that the Bayes approach, by changing the nature of the problem, increases the degree of ill-posedness that is now linked to the rate of decrease of the spectrum of $K\Omega_0^{\frac{1}{2}}$ and not of K as in the classical problem. However, no comparison can be done outside this particular form taken by Ω_0 .

Finally, we analyze frequentist asymptotic properties of the whole regularized posterior distribution $\mu_\alpha^{\hat{Y}}$. Following Diaconis and Freedman (1986), we give the following definition of *posterior consistency* (also called *frequentist consistency*) for a general posterior distribution $\mu^{\hat{Y}}$ on $(\mathcal{X}, \mathcal{E})$.

Definition 1. *For a given $x \in \mathcal{X}$, the pair $(x, \mu^{\hat{Y}})$ is consistent if $\mu^{\hat{Y}}$ converges weakly to δ_x as $n \rightarrow \infty$ under P^x -probability or P^x -a.s., where δ_x is the Dirac measure on x . The posterior probability $\mu^{\hat{Y}}$ is consistent if $(x, \mu^{\hat{Y}})$ is consistent for all $x \in \mathcal{X}$.*

If $(x, \mu^{\hat{Y}})$ is consistent in the previous sense, the Bayes estimate for x is consistent too. The meaning of this definition is that, for any neighborhood \mathcal{U} of the true parameter x , the posterior probability of the complement of \mathcal{U} , \mathcal{U}^c , converges toward zero when $n \rightarrow \infty$: $\mu^{\hat{Y}}(\mathcal{U}^c) \rightarrow 0$ in P^x -probability, or P^x -a.s..

By exploiting the results in Theorems 2 and 3 it is easy to show that, for a sequence ε_n with $\varepsilon_n \rightarrow 0$, $\mu_\alpha^{\hat{Y}}\{x \in \mathcal{X}; \forall \varphi \in \mathcal{X}, | \langle x - x_*, \varphi \rangle | \geq \varepsilon_n\}$ converges to 0. However, it is not possible to obtain an uniform convergence and the rate of contraction depends on the direction φ .

A stronger result of posterior consistency is given in the next theorem and it

requires a further assumption.

Theorem 4. *Under the assumptions of Theorem 2 and if there exists a $\kappa > 0$ such that $\sum_j \frac{\langle \Omega_0 \varphi_j, \varphi_j \rangle}{\rho_j^{2\kappa}} < \infty$, where $(\rho_j^2, \varphi_j)_j$ is the eigensystem associated to $\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}}$, then, for a sequence ε_n with $\varepsilon_n \rightarrow 0$, $\mu_\alpha^{\hat{Y}}\{x \in \mathcal{X}; \|x - x_*\| \geq \varepsilon_n\} \rightarrow 0$ in P^{x_*} -probability. Moreover, if $\delta_* \in \Phi_\beta$, for some $\beta > 0$, it is of order*

$$\mu_\alpha^{\hat{Y}}\{x \in \mathcal{X}; \|x - x_*\| \geq \varepsilon_n\} \sim \frac{1}{\varepsilon_n^2} \mathcal{O}_p\left(\alpha_n^\beta + \frac{1}{\alpha_n^4} \|\Sigma_n\|^2 \alpha_n^{(\beta+1) \wedge 2} + \frac{1}{\alpha_n} \text{tr} \Sigma_n + \alpha_n^\kappa\right).$$

In order to determine the rate of contraction, $\frac{1}{\alpha_n} \text{tr} \Sigma_n$ must be equated to α_n^β if $\kappa \geq \beta$, and to α_n^κ otherwise. Then, the rate of contraction of the posterior distribution is $\varepsilon_n = \|\Sigma_n\|^{\frac{\beta \wedge \kappa}{2(\beta \wedge \kappa + 1)}}$.

4.1 Convergence of the PVHS-regularized posterior distribution

We analyze frequentist asymptotic properties of mean and variance of the PVHS-regularized posterior distribution $\mu_s^{\hat{Y}}$, under Assumption 5. The rate of contraction of $\mathbb{E}_s(x|\hat{Y})$ is faster than that one of $\mathbb{E}_\alpha(x|\hat{Y})$ and it is the same as the rate of the classical solution of (1.1) obtained through a classical Tikhonov regularization in an Hilbert scale. The attainable speed of convergence is given in the following theorem that is proved in Appendix A.

Theorem 5. *Let $\mathbb{E}_s(x|\hat{Y})$ and V_s be as in (3.6). Under Assumptions 4 and 5, $\|\mathbb{E}_s(x|\hat{Y}) - x_*\|$ and $\|V_s \varphi\|$ converge to 0 in P^{x_*} -probability. Moreover,*

$$\mathbb{E}\|\mathbb{E}_s(x|\hat{Y}) - x_*\|^2 \sim \mathcal{O}_p\left(\alpha_n^{\frac{\beta+1}{a+s}} + \alpha_n^{\frac{1-a}{a+s}} \text{tr} \Sigma_n + \frac{1}{\alpha_n^2} \|\Sigma_n\|^2 \alpha_n^{\frac{\beta-a}{a+s}} + \frac{1}{\alpha_n^2} \|\Sigma_n\|^2 \text{tr} \Sigma_n \alpha_n^{\frac{1-a}{a+s}}\right)$$

and, the restriction of V_s to the subset of $\varphi \in \mathcal{X}$ such that $\Omega_0^{\frac{1}{2}} \varphi \in \mathcal{R}(\Omega_0^{\frac{\beta}{2}})$, has the order

$$\|V_s \varphi\|^2 \sim \mathcal{O}\left(\alpha_n^{\frac{\beta+1}{a+s}} + \frac{1}{\alpha_n^2} \|\Sigma_n\|^2 \alpha_n^{\frac{\beta-a}{a+s}}\right).$$

The optimal α_n (optimal for $\mathbb{E}_s(x|\hat{Y})$) is $\alpha_n^* \propto (\text{tr} \Sigma_n)^{\frac{a+s}{a+\beta}}$ and the corresponding optimal rate of $\mathbb{E}\|\mathbb{E}_s(x|\hat{Y}) - x_*\|^2$ and $\|V_s \varphi\|^2$ is proportional to $(\text{tr} \Sigma_n)^{\frac{\beta+1}{a+\beta}}$. With α_n set equal to its optimal values α_n^* , the remaining rates goes to zero if $\beta > \frac{a+2s}{3}$. This constraint is binding with respect to the constraint in Assumption 5 (iii), i.e. $\frac{a+2s}{3} \geq s - 1$, if $a \geq s - 3$. It should be noticed that parameter s , that

characterizes the norm in the Hilbert scale, does not play any role in the speed of convergence.

An advantage of the Tikhonov regularization in Hilbert Scale is that we can obtain rates of convergence for other norms, namely $\|\cdot\|_r$ for $-a \leq r \leq \beta + 1 \leq a + 2s$. The speed of convergence of these norms gives the speed of convergence of the estimate of the r -th derivative of x .

The rate of convergence can be improved if more assumptions on Σ are satisfied:

Corollary 2. *Let $(\lambda_j, \varphi_j, \psi_j)$ be the singular system of $K\Omega_0^{\frac{s+1}{2}}$. If $\Sigma_n = \frac{1}{n}\Sigma$, with Σ having eigenfunctions ψ_j and under the assumption that $\Sigma(K\Omega_0^{s+1}K^*)^{-\gamma}$ is trace-class for $\gamma \in [0, \frac{3\beta-a-2s}{2(a+s)} \wedge 1]$, then the optimal speed of $\mathbb{E}\|\mathbb{E}_s(x|\hat{Y}) - x_*\|^2$ is of order $n^{-\frac{\beta+1}{\beta+a-\gamma(a+s)}}$ and the corresponding optimal α_n^* is proportional to $n^{-\frac{a+s}{\beta+a-\gamma(a+s)}}$.*

The proof of this Corollary is similar to that one of Corollary 1 and then it is omitted. It is also possible to have a result for $\mu_s^{\hat{Y}}$ similar to that one in Theorem 4, this result is immediate and then omitted.

Now, we fix $x_0 = 0$ and we want to compare the rate of convergence of $\|\mathbb{E}_s(x|\hat{Y}) - x_*\|^2$ to the rate of the classical Tikhonov regularized solution in \mathcal{X}_s of (1.1). Such a solution is $x_s = (\alpha_n L^{2s} + K^*K)^{-1}K^*\hat{Y}$ and $\|x_s - x_*\|^2 \sim \mathcal{O}_p((tr\Sigma_n)^{\frac{u}{\bar{a}+u}})$, under the assumptions $\|Kx\| \sim \|L^{-\bar{a}}x\|$ and $x \in \mathcal{X}_u$ for some $u \geq 0$, with \bar{a} the degree of ill-posedness, see Section 8.5 in Engl, Hanke and Neubauer (2000). Hence, it results that $\|K\Omega_0^{\frac{1}{2}}x\| \sim \|L^{-\bar{a}}\Omega_0^{\frac{1}{2}}x\|$ and, by substituting to L the operator $\Omega_0^{-\frac{1}{2}}$, this norm is equivalent to $\|\Omega_0^{\frac{\bar{a}+1}{2}}x\|$. Comparison of this assumption to Assumption 5 (i) implies that the degree of ill-posedness in the Bayesian problem is greater than the degree of ill-posedness in the classical problem: $a = \bar{a} + 1$, as previously stated. Despite of this, if we take the same regularity condition in the two problems, *i.e.* $x \in \mathcal{X}_u = \mathcal{X}_{\beta+1}$ and then $\beta + 1 = u$, the rate of convergence of $\mathbb{E}_s(x|\hat{Y})$ and of x_s are the same.

This confirms the improvement, in terms of speed of convergence, that we have by using a Tikhonov regularization in the PVHS instead of a classical Tikhonov regularization. Let consider for instance the particular case $\Omega_0 = (K^*K)$, then $a = 0$, and let impose the same regularity condition in \mathcal{X} and in the Hilbert scale \mathcal{X}_s . The regularity condition in Theorem 2 requires that $\delta_* \in \mathcal{R}(\Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}})^{\frac{c}{2}} \equiv$

$\mathcal{R}((K^*K)^c)$ for some $c > 0$, that implies $(x_* - x_0) \in \mathcal{R}((K^*K)^{c+\frac{1}{2}})$. The regularity condition for the PVHS-regularization is $(x_* - x_0) \in \mathcal{R}(\Omega_0^{\frac{\beta+1}{2}}) \equiv \mathcal{R}((K^*K)^{\frac{\beta+1}{2}})$; henceforth the conditions are equal if $2c = \beta$. Taking this value for β , the optimal rate of convergence of $\mathbb{E}_s(x|\hat{Y})$, under assumptions of Theorem 5 is $(tr\Sigma_n)^{\frac{2c+1}{2c+2}}$ that is faster than the rate of $\mathbb{E}_\alpha(x|\hat{Y})$ (that is proportional to $(tr\Sigma_n)^{\frac{c}{c+1}}$).

Even without restricting to this particular form for Ω_0 it is possible to show the improvement in term of speed of convergence obtained with an Hilbert scale. To this end, it is sufficient that Assumption 5 (i) holds since it implies the equivalence $\|(\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{\frac{c}{2}} v\| \sim \|\Omega_0^{\frac{ac}{2}} v\|$, for some $v \in \mathcal{X}$. Then, if we require equality between Assumption 5 (ii) and the assumption in Theorem 2, we have $\|\Omega_0^{\frac{\beta}{2}} v\| \sim \|\Omega_0^{\frac{ac}{2}} v\|$ and then $\beta = ac$ (or $\beta = (\bar{a} + 1)c$). The optimal Bayesian rate of convergence with an Hilbert scale is $(tr\Sigma_n)^{\frac{ac+1}{a+ac}}$ that is fastest than the Bayes rate of convergence with a classical Tikhonov: $(tr\Sigma_n)^{\frac{c}{c+1}}$, $\forall c > 0$.

Acknowledgment

We thank Jan Johannes, Renauld Lestringand, Sebastien Van Bellegem, Anna Vanhems and the participants to BNRW01 in Cambridge, SFdS meeting in Aussois, Rencontre de Statistiques Mathématique in Marseille, 2008-JSM in Denver and to the seminars in Toulouse (GREMAQ and LSP), CREST (Paris) for helpful comments.

Appendix A

Proof of Lemma 1

Notice that $tr(\Sigma_n + K\Omega_0 K^*) = tr\Sigma_n + tr(K\Omega_0 K^*)$. Since Σ_n is trace class, we only have to prove that $K\Omega_0 K^*$ is trace class. This is trivial to prove if K is compact. Let $(\tilde{\lambda}_j^2, \tilde{\varphi}_j)$ be the eigensystem associated to K^*K . Then, $tr(K\Omega_0 K^*) = tr(K^* K \Omega_0)$ and by using the definition of trace

$$tr(K^* K \Omega_0) = \sum_j \langle K^* K \Omega_0 \tilde{\varphi}_j, \tilde{\varphi}_j \rangle = \sum_j \tilde{\lambda}_j^2 \langle \Omega_0 \tilde{\varphi}_j, \tilde{\varphi}_j \rangle$$

that is finite since Ω_0 is trace class and the spectrum of K^*K is decreasing. Let

now consider Υ :

$$\Upsilon = \begin{bmatrix} \Omega_0 & \Omega_0 K^* \\ K\Omega_0 & \Sigma_n + K\Omega_0 K^* \end{bmatrix}.$$

Let $e_j = (e_{1j}, e_{2j})$ be a basis in $\mathcal{X} \times \mathcal{Y}$, the trace of Υ is:

$$\begin{aligned} \text{tr}(\Upsilon) &= \sum_j \langle \Upsilon e_j, e_j \rangle \\ &= \sum_j (\langle \Omega_0 e_{1j}, e_{1j} \rangle + \langle \Omega_0 K^* e_{2j}, e_{1j} \rangle + \langle K\Omega_0 e_{1j}, e_{2j} \rangle \\ &\quad + \langle (\Sigma_n + K\Omega_0 K^*) e_{2j}, e_{2j} \rangle). \end{aligned}$$

By the previous part of this proof and since Ω_0 is trace-class, the infinite sum of the first and last terms are finite. We only have to consider the two terms in the center: $\sum_j (\langle \Omega_0 K^* e_{2j}, e_{1j} \rangle + \langle K\Omega_0 e_{1j}, e_{2j} \rangle) = 2 \sum_j \langle \Omega_0^{\frac{1}{2}} K^* e_{2j}, \Omega_0^{\frac{1}{2}} e_{1j} \rangle$ and

$$\begin{aligned} 2 \sum_j \langle \Omega_0^{\frac{1}{2}} K^* e_{2j}, \Omega_0^{\frac{1}{2}} e_{1j} \rangle &\leq 2 \sum_j \|\Omega_0 e_{1j}\| \|K^* e_{2j}\| \\ &\leq 2 \sup_j \|K^* e_{2j}\| \sum_j \langle \Omega_0 e_{1j}, e_{1j} \rangle \end{aligned}$$

that is finite since Ω_0 is trace class and K^* is bounded. The necessity of Υ_{yy} being trace-class to have Υ trace-class is evident and this complete the proof.

Proof of Theorem 1

(i). Let $(\tilde{x}, \tilde{y}) \in \mathcal{X} \times \mathcal{Y}$. Assumptions 1 implies that there exist $\tilde{y}_1 \in \mathcal{R}(K)$ and $\tilde{y}_2 \in \overline{\mathcal{R.K.H.S.}(\Sigma_n)}$ such that $\tilde{y} = \tilde{y}_1 + \tilde{y}_2$. Therefore, \tilde{y}_1 and \tilde{y}_2 are independent and there exists \tilde{x} such that $\tilde{y}_1 = K\tilde{x}$. For all $(\varphi, \psi) \in \mathcal{X} \times \mathcal{Y}$

$$\begin{aligned} \langle (\tilde{x}, \tilde{y}), (\varphi, \psi) \rangle &= \langle \tilde{x}, \varphi \rangle + \langle \tilde{y}_1 + \tilde{y}_2, \psi \rangle \\ &= \langle \tilde{x}, \varphi \rangle + \langle K\tilde{x}, \psi \rangle + \langle \tilde{y}_2, \psi \rangle \\ &= \langle \tilde{x}, \varphi + K^* \psi \rangle + \langle \tilde{y}_2, \psi \rangle \end{aligned}$$

and $\langle \tilde{x}, \varphi + K^* \psi \rangle + \langle \tilde{y}_2, \psi \rangle$ is normally distributed with mean $\langle x_0, \varphi + K^* \psi \rangle$ and variance $\langle \Omega_0(\varphi + K^* \psi), (\varphi + K^* \psi) \rangle + \langle \Sigma_n \psi, \psi \rangle$. Hence, we have proved that the joint measure Π on $\mathcal{X} \times \mathcal{Y}$ is gaussian. The mean m_{xy} is defined through $\langle m_{xy}, (\varphi, \psi) \rangle = \mathbb{E}_\Pi \langle (\tilde{x}, \tilde{y}), (\varphi, \psi) \rangle$ and since

$\langle x_0, \varphi + K^* \psi \rangle = \langle (x_0, Kx_0), (\varphi, \psi) \rangle$ we get $m_{xy} = (x_0, Kx_0)$. From the definition of Υ , we get $\langle \Upsilon(\varphi, \psi), (\varphi, \psi) \rangle = \langle \Omega_0 \varphi, \varphi \rangle + \langle (\Sigma_n + K\Omega_0 K^*) \psi, \psi \rangle$ that concludes the proof.

(ii). Let Q be the projection of Π on $(\mathcal{Y}, \mathcal{F})$ with mean function m_Q and covariance operator R_Q . Since Π is gaussian, the projection must be gaussian. Moreover, $\forall \psi \in \mathcal{Y}$, $\langle m_Q, \psi \rangle = \langle m_{xy}, (0, \psi) \rangle = \langle Kx_0, \psi \rangle$ and

$$\begin{aligned} \langle R_Q \psi, \psi \rangle &= \langle \Upsilon(0, \psi), (0, \psi) \rangle \\ &= \langle (\Omega_0 0 + \Omega_0 K^* \psi, (\Sigma_n + K\Omega_0 K^*) \psi + K\Omega_0 0), (0, \psi) \rangle \\ &= \langle (\Sigma_n + K\Omega_0 K^*) \psi, \psi \rangle. \end{aligned}$$

Hence, $m_Q = m_y$ and $R_Q = \Upsilon_{yy}$. This implies $Q \equiv P$ since there is an unique correspondence between a gaussian measure and its covariance operator and mean element.

Proof of Theorem 2

For any true value $x_* \in \mathcal{X}$, the Bayes estimation error can be decomposed as:

$$\begin{aligned} \mathbb{E}_\alpha(x|\hat{Y}) - x_* &= \Omega_0 K^* (\alpha_n I + \Sigma_n + K\Omega_0 K^*)^{-1} K(x_* - x_0) \\ &\quad + \Omega_0 K^* (\alpha_n I + \Sigma_n + K\Omega_0 K^*)^{-1} U - (x_* - x_0) \\ &= - \overbrace{[I - \Omega_0 K^* (\alpha_n I + K\Omega_0 K^*)^{-1} K](x_* - x_0)}^{\mathcal{A}} + \underbrace{\Omega_0 K^* [(\alpha_n I + \Sigma_n + K\Omega_0 K^*)^{-1} - (\alpha_n I + K\Omega_0 K^*)^{-1}] K(x_* - x_0)}_{\mathcal{B}} \\ &\quad + \underbrace{\Omega_0 K^* (\alpha_n I + \Sigma_n + K\Omega_0 K^*)^{-1} U}_{\mathcal{C}}. \end{aligned} \tag{5.1}$$

Term \mathcal{A} looks very similar to the regularization bias of the solution of a functional equation. More clearly, under Assumption 3 and by taking the norm in \mathcal{X} :

$$\begin{aligned} \mathcal{A} &= [I - \Omega_0 K^* (\alpha_n I + K\Omega_0 K^*)^{-1} K] \Omega_0^{\frac{1}{2}} \delta_* \\ &= \Omega_0^{\frac{1}{2}} [I - \Omega_0^{\frac{1}{2}} K^* (\alpha_n I + K\Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}}] \delta_*, \\ \mathbb{E} \|\mathcal{A}\|^2 &\leq \|\Omega_0^{\frac{1}{2}}\|^2 \|(I - \Omega_0^{\frac{1}{2}} K^* (\alpha_n I + K\Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}})\|^2 \|\delta\|^2. \end{aligned}$$

We notice that $(I - \Omega_0^{\frac{1}{2}} K^* (\alpha_n I + K\Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}})$ is equal to $[I - (\alpha_n I +$

$\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{-1} \Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}}]$. The latter is the regularization bias associated to the regularized solution of the ill-posed inverse problem $K \Omega_0^{\frac{1}{2}} \delta_* = r$ computed using Tikhonov regularization scheme. Assumption 4 guarantees identification of its solution. It converges to zero when $\alpha_n \rightarrow 0$ and then the norm $\|(I - \Omega_0^{\frac{1}{2}} K^* (\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}})\|^2$ is bounded. Its speed of convergence to zero depends on the regularity of δ_* and consequently of $(x_* - x_0)$. If $\delta_* \in \Phi_\beta$, it is at most of order α_n^β , see Proposition 3.12 in Carrasco, Florens and Renault (2007). Then $\mathbb{E} \|\mathcal{A}\|^2 = \mathcal{O}_p(\alpha_n^\beta)$.

For term \mathcal{B} we have $\mathbb{E} \|\mathcal{B}\|^2 = \|\Omega_0 K^* (\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1} (\Sigma_n) (\alpha_n I + K \Omega_0 K^*)^{-1} K (x_* - x_0)\|^2$ and it is less than or equal to

$$\|\Omega_0 K^*\|^2 \|(\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1}\|^2 \|\Sigma_n\|^2 \|(\alpha_n I + K \Omega_0 K^*)^{-1} K (x_* - x_0)\|^2$$

where the first norm is bounded and the second and the third ones are $\mathcal{O}_p(\frac{1}{\alpha_n^2})$ and $\mathcal{O}_p(\|\Sigma_n\|^2)$ respectively. The last norm can be written as $\|(\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}} \delta_*\|^2$, and, by using the hypothesis that $\delta_* \in \Phi_\beta$

$$\|(\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}} \delta_*\|^2 = \frac{1}{\alpha^2} \|\alpha (\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}} (\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}} \rho\|^2,$$

for some $\rho \in \mathcal{X}$ and it is at least of order $\frac{1}{\alpha^2} \alpha^{\beta+1}$. As a consequence of the fact that, with a Tikhonov regularization, a degree of smoothness greater than or equal to 2 may be useless, we get $\|(\alpha_n I + K \Omega_0 K^*)^{-1} K (x_* - x_0)\|^2 \sim \mathcal{O}_p(\frac{1}{\alpha_n^2} \alpha_n^{(\beta+1) \wedge 2})$.

To find speed of convergence of term \mathcal{C} we re-write it as:

$$\begin{aligned} \mathcal{C} &= \underbrace{\Omega_0 K^* [(\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1} - (\alpha_n I + K \Omega_0 K^*)^{-1}] U}_{\mathcal{C}a} + \\ &\quad \underbrace{\Omega_0 K^* (\alpha_n I + K \Omega_0 K^*)^{-1} U}_{\mathcal{C}b}. \end{aligned}$$

By standard computation it is trivial to show that $\mathbb{E} \|\mathcal{C}a\|^2 \sim \mathcal{O}_p(\frac{1}{\alpha_n^3} \|\Sigma_n\|^2 \text{tr} \Sigma_n)$ and $\mathbb{E} \|\mathcal{C}b\|^2 \sim \mathcal{O}_p(\frac{1}{\alpha_n} \text{tr} \Sigma_n)$, since $\mathbb{E} \|U\|^2 = \text{tr} \Sigma_n$. Term $\mathbb{E} \|\mathcal{C}a\|^2$ is negligible with respect to the terms $\|\mathcal{B}\|^2$ and $\mathbb{E} \|\mathcal{C}b\|^2$.

Proof of Theorem 3

By recalling expression (3.3), $\forall \varphi \in \mathcal{X}$ we can rewrite V_α as

$$V_\alpha \varphi = \underbrace{\Omega_0 - \Omega_0 K^* (\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0}_{\mathcal{D}} \varphi + \underbrace{\Omega_0 K^* (\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0 \varphi - \Omega_0 K^* (\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1} K \Omega_0 \varphi}_{\mathcal{G}}. \quad (5.2)$$

Since Ω_0 is a positive definite self-adjoint operator, it can be decomposed as $\Omega_0 = \Omega_0^{\frac{1}{2}} \Omega_0^{\frac{1}{2}}$. Term \mathcal{D} is treated as term \mathcal{A} in (5.1), so that, if $\varphi \in \mathcal{X}$ is such that $\Omega_0^{\frac{1}{2}} \varphi \in \mathcal{R}(\Omega_0^{\frac{1}{2}} K^* K \Omega_0^{\frac{1}{2}})^{\frac{\beta}{2}}$, $\|\mathcal{D}\varphi\|^2 = \mathcal{O}_p(\alpha_n^\beta)$. Term \mathcal{G} can be rewritten as $\Omega_0 K^* (\alpha_n I + \Sigma_n + K \Omega_0 K^*)^{-1} \Sigma_n (\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0^{\frac{1}{2}} \Omega_0^{\frac{1}{2}} \varphi$ and treated as term \mathcal{B} in (5.1), then $\|\mathcal{G}\|^2 = \mathcal{O}_p(\frac{1}{\alpha_n^4} \|\Sigma_n\|^2 \alpha_n^{(\beta+1)\wedge 2})$.

Proof of Corollary 1

Let consider term $\mathcal{C}b$ in the proof of Theorem 2, *i.e.* $\mathcal{C}b = \Omega_0 K^* (\alpha_n I + K \Omega_0 K^*)^{-1} U$. The expectation of its norm $\mathbb{E}\|\mathcal{C}b\|^2$ is bounded by $\|\Omega_0^{\frac{1}{2}}\|^2 \mathbb{E}\|\Omega_0^{\frac{1}{2}} K^* (\alpha_n I + K \Omega_0 K^*)^{-1} U\|^2$. The first norm is finite and the second one is equal to $\text{tr}(\Sigma_n (\alpha_n I + K \Omega_0 K^*)^{-1} K \Omega_0 K^* (\alpha_n I + K \Omega_0 K^*)^{-1})$ if Σ_n and $K \Omega_0 K^*$ have the same eigenfunctions. Let $\Sigma_n = \frac{1}{n} \Sigma$, ρ_j^2 denote the eigenvalues of Σ and λ_j^2 denote the eigenvalues of $K \Omega_0 K^*$, then

$$\begin{aligned} \mathbb{E}\|\mathcal{C}b\|^2 &\leq \|\Omega_0^{\frac{1}{2}}\|^2 \frac{1}{n} \sum_j \frac{\rho_j^2 \lambda_j^2}{(\alpha_n + \lambda_j^2)^2} \\ &\leq \|\Omega_0^{\frac{1}{2}}\|^2 \frac{1}{n} \sup_j \left(\frac{\lambda_j^{2+2\gamma}}{(\alpha_n + \lambda_j^2)^2} \right) \sum_j \rho_j^2 \lambda_j^{-2\gamma} \\ &\sim \mathcal{O}_p\left(\frac{1}{n} \alpha_n^{\gamma-1}\right) \end{aligned}$$

for $\gamma \in [0, 1]$. The optimal α_n is obtained by equating this rate to that one of term $\|\mathcal{A}\|^2$ in the proof of Theorem 2; the optimal speed follows. The upper bound $\frac{3\beta-1}{2}$ of γ ensures that the other term in $\mathbb{E}\|\mathbb{E}_\alpha(x|\hat{Y}) - x_*\|^2$ converges to zero.

Proof of Theorem 4

By using Chebishev's Inequality for a sequence ε_n with $\varepsilon_n \rightarrow 0$, we have

$$\mu_\alpha^{\hat{Y}}\{x \in \mathcal{X}; \|x - x_*\| \geq \varepsilon_n\} \leq \frac{\|\mathbb{E}_\alpha(x|\hat{Y}) - x_*\|^2 + \text{tr}(V_\alpha(x|\hat{Y}))}{\varepsilon_n^2}.$$

By Theorem 2 we know that $\|\mathbb{E}_\alpha(x|\hat{Y}) - x_*\|^2$ converges to 0 and we know its rate of contraction. In order to compute the trace of the variance, we use the decomposition in (5.2), hence $tr(V_\alpha(x|\hat{Y})) = tr(\mathcal{D}) + tr(\mathcal{G})$. By using properties and the definition of the trace function, we get

$$\begin{aligned}
tr(\mathcal{D}) &= tr[\Omega_0^{\frac{1}{2}}(I - \Omega_0^{\frac{1}{2}}K^*(\alpha_n I + K\Omega_0 K^*)^{-1}K\Omega_0^{\frac{1}{2}})\Omega_0^{\frac{1}{2}}] \\
&= tr[\alpha_n(\alpha_n I + \Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}})^{-1}\Omega_0] \\
&= \sum_j \frac{\alpha_n}{\alpha_n + \rho_j^2} < \Omega_0 \varphi_j, \varphi_j > \\
&\leq \sup_j \left(\frac{\alpha_n \rho_j^{2\kappa}}{\alpha_n + \rho_j^2} \right) \sum_j \frac{< \Omega_0 \varphi_j, \varphi_j >}{\rho_j^{2\kappa}} \\
&\sim \mathcal{O}(\alpha_n^\kappa)
\end{aligned}$$

under the assumption that $\sum_j \frac{< \Omega_0 \varphi_j, \varphi_j >}{\rho_j^{2\kappa}} < \infty$, where $(\rho_j^2, \varphi_j)_j$ is the eigensystem associated to $\Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}}$. Then, it converges to 0. The $tr(\mathcal{G})$ is less or equal to $tr(K\Omega_0^2 K^*(\alpha_n I + K\Omega_0 K^*)^{-1})tr(\Sigma_n(\alpha_n I + \Sigma_n + K\Omega_0 K^*)^{-1})$ and, in a similar way as for $tr(\mathcal{D})$, we can prove that $tr(\mathcal{G}) \sim \mathcal{O}(\alpha_n^\kappa \frac{1}{\alpha} tr \Sigma_n)$. This concludes the proof.

Proof of Theorem 5

To prove Theorem 5 we use Corollary 8.22 in Engl, Hanke and Neubauer(2000). We give a slightly modified version of it:

Corollary 3. *Let \mathcal{X}_s , $s \in \mathbb{R}$, be a Hilbert scale induced by L and let $T : \mathcal{X} \rightarrow \mathcal{Y}$ be a bounded operator satisfying $\|x\|_{-a} \sim \|Tx\|$ on \mathcal{X} for some $a > 0$. Then for $B := TL^{-s}$, $s \geq 0$ and $|\nu| \leq 1$ we have $\|x\|_{-\nu(a+s)} \sim \|(B^*B)^{\frac{\nu}{2}}x\|$. Moreover, $\mathcal{R}((B^*B)^{\frac{\nu}{2}}) = \mathcal{X}_{\nu(a+s)}$.*

We rewrite the bias $\mathbb{E}_s(x|\hat{Y}) - x_*$ as

$$\begin{aligned}
\mathbb{E}_s(x|\hat{Y}) - x_* &= \overbrace{-[I - \Omega_0 K^*(\alpha_n L^{2s} + K\Omega_0 K^*)^{-1}K](x_* - x_0)}^{\mathcal{K}} + \\
&\quad \underbrace{\Omega_0 K^*[(\alpha_n L^{2s} + \Sigma_n + K\Omega_0 K^*)^{-1}K - (\alpha_n L^{2s} + K\Omega_0 K^*)^{-1}K](x_* - x_0)}_{\mathcal{J}} \\
&\quad + \underbrace{\Omega_0 K^*(\alpha_n L^{2s} + \Sigma_n + K\Omega_0 K^*)^{-1}U}_{\mathcal{M}}.
\end{aligned}$$

Then, by using Assumption 5 (ii) $\mathbb{E}||\mathcal{K}||^2 \leq ||\Omega_0^{\frac{1}{2}}[I - (\alpha_n\Omega_0^{-s} + \Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}})^{-1}\Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}}]\Omega_0^{\frac{\beta}{2}}\rho_*||^2$ if Ω_0 is such that $\Omega_0^{\frac{1}{2}}K^*(\alpha_nL^{2s} + K\Omega_0K^*)^{-1} = (\alpha_n\Omega_0^{-s} + \Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}})^{-1}\Omega_0^{\frac{1}{2}}K^*$, *i.e.* $\Omega_0^{-s+\frac{1}{2}}K^* = \Omega_0^{\frac{1}{2}}K^*L^{2s}$. Let $B = K\Omega_0^{\frac{s+1}{2}}$, we rewrite

$$\begin{aligned}\mathbb{E}||\mathcal{K}||^2 &= ||\Omega_0^{\frac{s+1}{2}}(I - (\alpha_nI + B^*B)^{-1}B^*B)\Omega_0^{\frac{\beta-s}{2}}\rho_*||^2 \\ &= ||\Omega_0^{\frac{s+1}{2}}(I - (\alpha_nI + B^*B)^{-1}B^*B)(B^*B)^{\frac{\beta-s}{2(a+s)}}v||^2 \\ &\sim ||(B^*B)^{\frac{\beta+1}{2(a+s)}}\alpha_n(\alpha_nI + B^*B)^{-1}v||^2 \\ &\sim \mathcal{O}_p(\alpha_n^{\frac{\beta+1}{(a+s)}})\end{aligned}$$

where the second line follows from the fact that $\mathcal{R}(\Omega_0^{\frac{\beta-s}{2}}) \equiv \mathcal{X}_{\beta-s} \equiv \mathcal{R}((B^*B)^{\frac{\beta-s}{2(a+s)}})$, by Corollary 3, then $\Omega_0^{\frac{\beta-s}{2}}\rho_* = (B^*B)^{\frac{\beta-s}{2(a+s)}}v$, for some $v \in \mathcal{X}$. The third equivalence too follows from Corollary 3. We conclude that $\mathbb{E}||\mathcal{K}||^2 \sim \mathcal{O}_p(\alpha_n^{\frac{\beta+1}{a+s}})$.

In a similar way, for term \mathcal{J} we have:

$$||\mathcal{J}|| \leq ||\Omega_0K^*(\alpha_nL^{2s} + \Sigma_n + K\Omega_0K^*)^{-1}|| ||\Sigma_n|| ||(\alpha_nL^{2s} + K\Omega_0K^*)^{-1}K\Omega_0^{\frac{1}{2}}\delta_*||$$

where the first norm is of order $\frac{1}{\alpha_n}$ and

$$\begin{aligned}||(\alpha_nL^{2s} + K\Omega_0K^*)^{-1}K\Omega_0^{\frac{1}{2}}\delta_*|| &= ||K\Omega_0^{\frac{1}{2}}(\alpha_n\Omega_0^{-s} + \Omega_0^{\frac{1}{2}}K^*K\Omega_0^{\frac{1}{2}})^{-1}\delta_*|| \\ &= ||B(\alpha_nI + B^*B)^{-1}\Omega_0^{\frac{s+\beta}{2}}v|| \\ &\sim ||(B^*B)^{\frac{2s+\beta+a}{2(a+s)}}(\alpha_nI + B^*B)^{-1}v|| \\ &\sim \mathcal{O}_p(\frac{1}{\alpha_n}\alpha_n^{\frac{2s+\beta+a}{2(a+s)}}).\end{aligned}$$

Thus, $\mathbb{E}||\mathcal{J}||^2 \sim \mathcal{O}_p\left(\frac{1}{\alpha_n^2}||\Sigma_n||^2\alpha_n^{\frac{\beta-a}{(a+s)}}\right)$. The last term can be decomposed as

$$\begin{aligned}\mathcal{M} &= \underbrace{\Omega_0K^*[(\alpha_nL^{2s} + \Sigma_n + K\Omega_0K^*)^{-1} - (\alpha_nL^{2s} + K\Omega_0K^*)^{-1}]U}_{\mathcal{Ma}} + \\ &\quad \underbrace{\Omega_0K^*(\alpha_nL^{2s} + K\Omega_0K^*)^{-1}U}_{\mathcal{Mb}},\end{aligned}$$

and $\mathbb{E}||\mathcal{M}a||^2$ is less or equal then

$$\begin{aligned}
& ||\Omega_0 K^* (\alpha_n L^{2s} + K\Omega_0 K^*)^{-1}||^2 ||\Sigma_n||^2 ||(\alpha_n L^{2s} + \Sigma_n + K\Omega_0 K^*)^{-1}||^2 \mathbb{E}||U||^2 \\
& \leq ||\Omega_0^{\frac{s+1}{2}} (\alpha_n I + \Omega_0^{\frac{s+1}{2}} K^* K\Omega_0^{\frac{s+1}{2}})^{-1} \Omega_0^{\frac{s+1}{2}} K^*||^2 ||\Sigma_n||^2 ||(\alpha_n L^{2s} + \Sigma_n + K\Omega_0 K^*)^{-1}||^2 \text{tr} \Sigma_n \\
& \sim ||(B^* B)^{\frac{a+2s+1}{2(a+s)}} (\alpha_n I + B^* B)^{-1}||^2 ||\Sigma_n||^2 ||(\alpha_n L^{2s} + \Sigma_n + K\Omega_0 K^*)^{-1}||^2 \text{tr} \Sigma_n \\
& \sim \mathcal{O}_p\left(\frac{1}{\alpha_n^2} ||\Sigma_n||^2 \text{tr} \Sigma_n \alpha_n^{\frac{1-a}{a+s}}\right).
\end{aligned}$$

The expectation of the squared norm of the term $\mathcal{M}b$ is:

$$\begin{aligned}
\mathbb{E}||\mathcal{M}b||^2 &= \mathbb{E}||\Omega_0^{\frac{1}{2}} (\alpha_n \Omega_0^{-s} + \Omega_0^{\frac{1}{2}} K^* K\Omega_0^{\frac{1}{2}})^{-1} \Omega_0^{\frac{1}{2}} K^* U||^2 \\
&= \mathbb{E}||\Omega_0^{\frac{s+1}{2}} (\alpha_n I + B^* B)^{-1} B^* U||^2 \\
&\leq ||(B^* B)^{\frac{2s+a+1}{2(a+s)}} (\alpha_n I + B^* B)^{-1}||^2 \mathbb{E}||U||^2 \\
&\sim \mathcal{O}_p(\alpha_n^{\frac{1-a}{a+s}} \text{tr} \Sigma_n).
\end{aligned}$$

Thus $||\mathcal{M}b||^2 \sim \mathcal{O}_p(\alpha_n^{\frac{1-a}{a+s}} \text{tr} \Sigma_n)$.

Next, we consider the norm of the variance operator V_s applied to an element $\varphi \in \mathcal{X}$ such that $\Omega_0^{\frac{1}{2}} \varphi \in \mathcal{R}(\Omega_0^{\frac{\beta}{2}})$. Then, φ is such that $\Omega_0^{\frac{s+1}{2}} \varphi \in \mathcal{R}(\Omega_0^{\frac{\beta-s}{2}})$ and we have the decomposition

$$\begin{aligned}
V_s \varphi &= \overbrace{[\Omega_0 - \Omega_0 K^* (\alpha_n L^{2s} + K\Omega_0 K^*)^{-1} K\Omega_0] \varphi}^{\mathcal{W}} \\
&\quad + \underbrace{\Omega_0 K^* [(\alpha_n L^{2s} + K\Omega_0 K^*)^{-1} - (\alpha_n L^{2s} + \Sigma_n + K\Omega_0 K^*)^{-1}] K\Omega_0 \varphi}_{\mathcal{Z}}.
\end{aligned}$$

Computations for $||\mathcal{W}||$ are similar to that ones for term $||\mathcal{K}||$ above and computation for $||\mathcal{Z}||$ to that one for term $||\mathcal{J}||$, therefore: $||\mathcal{W}||^2 \sim \mathcal{O}_p(\alpha_n^{\frac{\beta+1}{a+s}})$ and $||\mathcal{Z}||^2 \sim \mathcal{O}_p\left(\frac{1}{\alpha_n^2} ||\Sigma_n||^2 \alpha_n^{\frac{\beta-a}{a+s}}\right)$. The result follows.

Appendix B: Numerical Simulations

In all the simulations we take the Tikhonov regularized posterior mean $\mathbb{E}_\alpha(x|\hat{Y})$ as point estimator of the solution of the inverse problem (1.1).

Functional equation with a parabola as solution

We take $\mathcal{X} = \mathcal{Y} = L_\pi^2$ with π the uniform measure on $[0, 1]$. The data

generating process is

$$\begin{aligned}
 \hat{Y} &= \int_0^1 x(s)(s \wedge t)ds + U, & x_* &= -3s^2 + 3s \\
 U &\sim \mathcal{GP}(0, \Sigma_n), & \Sigma_n &= n^{-1} \int_0^1 \exp\{-(s-t)^2\}ds \\
 x &\sim \mathcal{GP}(x_0, \Omega_0), & x_0 &= -2.8s^2 + 2.8s \\
 \Omega_0\varphi(t) &= \omega_0 \int_0^1 \exp(-(s-t)^2)\varphi(s)ds.
 \end{aligned} \tag{6.1}$$

The results are shown in Figure 6.1. In Panels (a), (b) and (c) of this figure we represent estimation of x_* by using three different prior means x_0 (represented by the dash-dotted magenta line) and different values for Ω_0 : $x_0 = -2.8s^2 + 2.8s$ and $\Omega_0\varphi(t) = 2 \int_0^1 \exp(-(s-t)^2)\varphi(s)ds$ in panel (a), $x_0 = -2s^2 + 2s$ and $\Omega_0\varphi(t) = 40 \int_0^1 ((s \wedge t) - st)\varphi(s)ds$ in panel (b), $x_0 = -2.22s^2 + 2.67s - 0.05$ and $\Omega_0\varphi(t) = 100 \int_0^1 (0.9(s-t)^2 - 1.9|s-t| + 1)\varphi(s)ds$ in panel (c). The black dotted line represents the $\mathbb{E}_\alpha(x|\hat{Y})$ and the solid red line represents the true function x_* . The blue dotted line represents the classical solution obtained with a Tikhonov method. The regularization parameter α has been set to $2.e - 03$, the sample size is $n = 1000$ and the discretization step is 0.001.

In Panels (a), (b) and (c), results of a Monte Carlo experiment with 100 iterations are shown. The specification of the prior distribution changes as in the previous panels. The dotted line represents the mean of the regularized posterior means obtained for each iteration.

Density Estimation

Let $\mathcal{X} = \mathcal{Y} = L_\pi^2$, with π the uniform measure on $[-3, 3]$. The true density f_* is the density of a standard gaussian measure on \mathbb{R} . Let ξ_1, \dots, ξ_n be an *i.i.d.* sample from f_* used to estimate the cumulative distribution function \hat{F} and the sampling variance Σ_n (as defined in Example 2). The operator K is known.

The prior mean is $f_0 = \frac{1}{\sqrt{2\pi}\sigma} \exp\{-\frac{1}{2\sigma^2}(\xi - \theta)^2\}$, the prior variance is $\Omega_0\varphi(t) = \omega_0 \int_{-3}^3 \exp(-(s-t)^2)\varphi(s)\frac{1}{6}ds$ and parameters $(\sigma, \theta, \omega_0)$ have been differently set to see the effect of prior changes on the estimated solution. The regularization parameter α_n has been set equal to 0.05 and the sample size is of $n = 1000$. The results for different specification of the parameters in the prior distribution are shown in Figure 6.2. In panels (a) and (d) the true density (continuous black line),

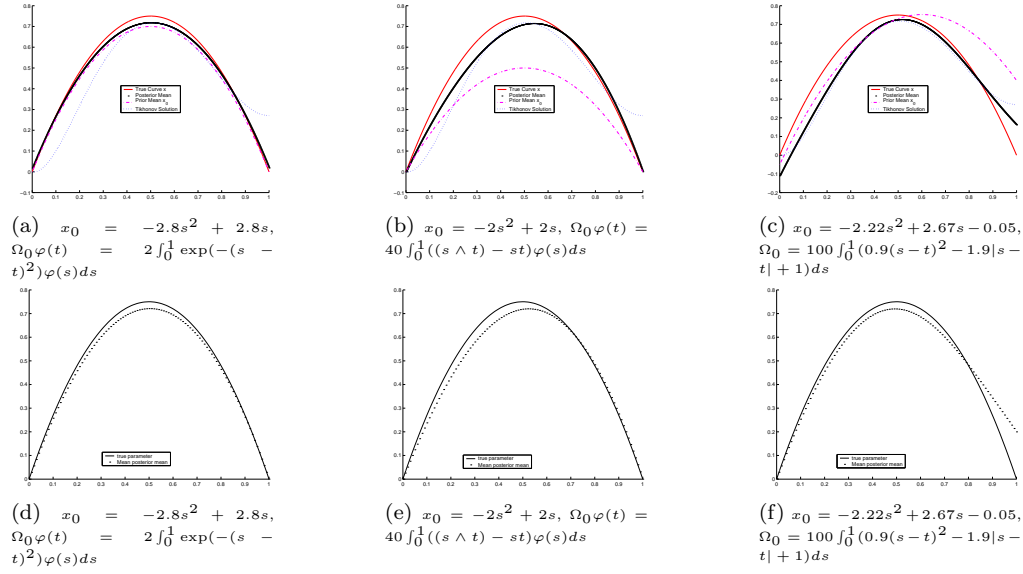


Figure 6.1: Regularized posterior mean for different prior specifications and comparison with the classical Tikhonov solution

the prior mean (dotted blue line) and the regularized posterior mean estimator (dashed-dotted red line) are drawn; panels (b) and (e) show the comparison between our estimator and the classical Tikhonov solution (dotted blue line). Panels (c) and (f) represent a sample of curves drawn from the prior distribution together with the prior mean (continuous line) and the true density (dotted line).

References

- Agliari, A. and Parisetti, C.C. (1988). A g-reference informative prior: a Note on Zellner's g-prior. *The Statistician*, Vol.37, **3**, 271 - 275.
- Baker, C.R. (1973). Joint measures and cross-covariance operators. *Transactions of the American Mathematical Society*, **186**, 273-289.
- Carrasco, M., and Florens, J.P. (2000). Generalization of GMM to a continuum of moment conditions. *Econometric Theory* **16**, 797-834.
- Carrasco, M., Florens, J.P., and Renault, E. (2007). Linear inverse problems in structural econometrics: estimation based on spectral decomposition and

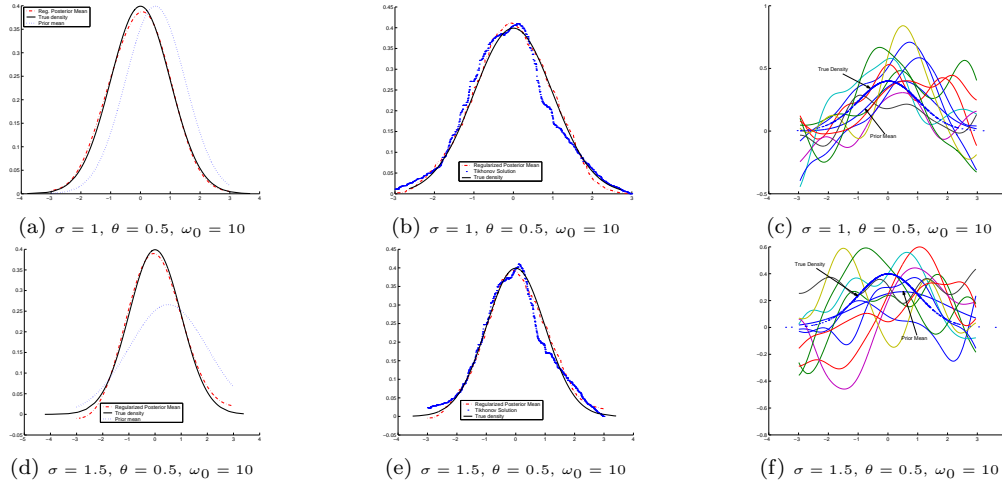


Figure 6.2: Regularized posterior distribution for a density function. Comparison among different specification of the prior distribution and with the classical Tikhonov solution.

regularization. *Handbook of Econometrics*, J.J. Heckman and E. Leamer, eds., **6B**, Elsevier, North Holland.

Diaconis, F. and Freedman, D. (1986). On the consistency of bayes estimates. *Ann. Statist.* **14**, 1-26.

Engl, H.W., Hanke, M. and Neubauer, A. (2000). *Regularization of inverse problems*, Kluwer Academic, Dordrecht.

Florens, J.P., Mouchart, M. and Rolin, J.M. (1990). *Elements of Bayesian statistics*, Dekker, New York.

Florens, J.P. and Simoni, A. (2009). Regularizing priors for linear inverse problems. Preprint. Available at http://simoni.anna.googlepages.com/Regularizing_Priors.pdf

Franklin, J.N. (1970). Well-posed stochastic extension of ill-posed linear problems. *J. Math. Anal. Appl.* **31**, 682 - 716.

Hiroshi, S. and Yoshiaki, O. (1975). Separabilities of a gaussian measure. *Ann. Inst. H. Poincaré Probab. Statist.* tome 11, **3**, 287 - 298.

Kaipio, J. and Somersalo, E. (2004). *Statistical and computational inverse problems*, Applied Mathematical Series, vol.160, Springer, Berlin.

- Krein, S.G. and Petunin, J.I. (1966). Scales of Banach spaces. *Russian Math. Surveys*, **21**, 85 - 160.
- Kress, R. (1999). *Linear integral equation*, Springer.
- Lehtinen, M.S., Päivärinta, L. and Somersalo, E. (1989). Linear inverse problems for generalised random variables. *Inverse Problems*, **5**, 599-612.
- Mandelbaum, A. (1984). Linear estimators and measurable linear transformations on a Hilbert space. *Z. Wahrscheinlichkeitstheorie*, **3**, 385-98.
- Neveu, J. (1965). *Mathematical foundations of the calculus of probability*, San Francisco: Holden-Day.
- Prenter, P.M. and Vogel, C.R. (1985). Stochastic inversion of linear first kind integral equations. I. Continuous theory and the stochastic generalized inverse. *J. Math. Anal. Appl.* **106**, 202 - 212.
- Simoni, A. (2009), 'Bayesian analysis of linear inverse problems with applications in economics and finance', *PhD Dissertation* - Université de Science Sociales, Toulouse. Available at http://simoni.anna.googlepages.com/PhD_Dissertation.pdf
- Van der Vaart, A.W. and Van Zanten, J.H. (2008a). Rates of contraction of posterior distributions based on gaussian process priors. *Ann. Statist.* **36**.
- Van der Vaart, A.W. and Van Zanten, J.H. (2008b). Reproducing kernel Hilbert spaces of gaussian priors. *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh. IMS Collections*, **3**, 200 - 222. Institute of Mathematical Statistics.
- Vapnik, V.N. (1998). *Statistical learning theory*, John Wiley & Sons, Inc.
- Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distribution. In *Bayesian Inference and Decision Techniques: essays in honor of Bruno de Finetti*, Goel, P.K. and Zellner, A. (eds.) pp. 233-243 (Amsterdam, North Holland).

Toulouse School of Economics, Toulouse, FRANCE.

E-mail: (florens@cict.fr)

Università Bocconi, Milano, Italy.

E-mail: (anna.simoni@unibocconi.it)