Volume 15, No. 4 (2006), pp. Allerton Press, Inc.

MATHEMATICAL METHODS OF STATISTICS

REGULARIZATION OF INVERSE PROBLEMS WITH UNKNOWN OPERATOR

C. Marteau

CMI, Université Aix-Marseille 1
39, rue F. Joliot-Curie, F-13 453 Marseille Cedex, France
E-mail: clement.marteau@cmi.univ-mrs.fr

In this paper, we study statistical inverse problems. We are interested in the case where the operator is not exactly known. Using the penalized blockwise Stein's rule, we construct an estimator that produces sharp asymptotic oracle inequalities in different settings. In particular, we consider the case, where the set of bases is not associated with the singular value decomposition. The representation matrix of the operator is not diagonal and the regularization problem becomes more difficult.

Key words: blockwise Stein's rule, exact minimax constant, oracle inequalities, singular value decomposition, statistical inverse problem.

2000 Mathematics Subject Classification: 62G05, 62G20.

1. Introduction

Inverse problems consist in recovering an unknown signal f using indirect observation Y. We consider in this paper the statistical linear inverse problem framework. Let H, K be separable Hilbert spaces and $A\colon H \to K$ a compact operator. Assume that we observe

$$Y = Af + \varepsilon\xi, \tag{1.1}$$

where $f \in H$ and ε is the noise level. The quantity ξ is assumed to be a Gaussian white noise (see Hida [18] for more detail). Notation (1.1) means in this case that, for any function $g \in K$, we can observe

$$\langle Y, g\rangle = \langle Af, g\rangle + \varepsilon\langle \xi, g\rangle, \tag{1.2}$$

where $\langle \xi, g\rangle \sim \mathcal{N}(0, \|g\|^2)$. Given $g_1, g_2 \in K$, the associated covariance between $\langle \xi, g_1\rangle$ and $\langle \xi, g_2\rangle$ is the scalar product $\langle g_1, g_2\rangle$. Using (1.1), our aim is to estimate the function f.

Since A is a compact operator, the solution of (1.1) does not continuously depend on the data. The problem is ill-posed. Only approximations of f obtained via regularization methods are available. A classical way of regularization is related to the singular value decomposition (SVD) of A (see, for instance, Baumeister [1], Kress [21] or Engl *et al.* [13]). The operator A^*A is compact and self-adjoint. Call $(b_k^2)_{k\geq 1}$ the sequence of eigenvalues and assume that A^*A admits an orthonormal system of eigenfunctions $(\phi_k)_{k\geq 1}$. Then construct an image basis $(\psi_k)_{k\geq 1}$ satisfying, for all $k \in \mathbb{N}$:

$$\begin{cases} A\phi_k = b_k\psi_k, \\ A^*\psi_k = b_k\phi_k. \end{cases} \tag{1.3}$$

The system $(\psi_k)_{k\geq 1}$ is orthonormal. For all integer k, replace g by ψ_k in (1.2) and set $\theta_k = \langle f, \phi_k\rangle$. The model (1.1) can be written in the sequence space form:

$$y_k = b_k\theta_k + \varepsilon\xi_k, \qquad k \in \mathbb{N}, \tag{1.4}$$

where the ξ_k are independent standard Gaussian random variables. In the L_2 sense, recovering f is equivalent to recovering the sequence $\theta = (\theta_k)_{k\geq 1}$.

Since A^*A is compact, the sequence $(b_k)_{k\geq 1}$ vanishes as k tends to infinity. For large values of k, the signal $b_k\theta_k$ is thus attenuated compared to the noise $\varepsilon\xi_k$. The difficulty of the problem in such a situation is related to the behavior of the eigenvalues. The faster the sequence $(b_k)_{k\geq 1}$ decreases, the more difficult the problem is. In this paper, only mildly ill-posed problems are considered: the sequence $(b_k)_{k\geq 1}$ is polynomially decreasing. If the sequence $(b_k)_{k\geq 1}$ is exponentially decreasing, the problem is said to be severely ill-posed. This particular case will not be studied here.

In this framework, very interesting results were obtained in the last two decades. We mention, for instance, Ermakov [14], Johnstone and Silverman [20], Fan [15], Mair and Ruymgaart [22], Efromovich [11], Nemirovski [24], Golubev and Khasminskii [16], Tsybakov [26] or Cavalier *et al.* [5].

In all the papers mentioned above, the operator A is assumed to be exactly known. This assumption is of major importance and may not be satisfied in many situations. Consider the example of convolution operator defined on $L^2(0,1)$ by:

$$Af\colon [0,1] \to \mathbb{R}, \qquad t \mapsto Af(t) = \int_0^1 K(x-t)f(x)\,dx,$$

where the kernel K belongs to $L^2(0,1)$. The Fourier basis is associated with the singular value decomposition. In this situation, the eigenvalues correspond to the Fourier coefficients of K. If the kernel is unknown even up to a parameter, no estimator can be constructed. Nevertheless, the sequence of eigenvalues may be approximated via independent observations on the kernel K. Recently, some authors were interested in the quality of estimation in such a situation. In the model selection context, Cavalier and Hengartner [6] dealt, for instance, with two sets of data:

$$\begin{cases} y_k = b_k\theta_k + \varepsilon\xi_k, \\ x_k = b_k + \sigma\eta_k, \qquad \forall k \in \mathbb{N}. \end{cases} \tag{1.5}$$

For all $k \in \mathbb{N}$, the η_k denote i.i.d. standard Gaussian random variables independent of the ξ_k and $\sigma > 0$ is the noise level. The sequence $(x_k)_{k\geq 1}$ corresponds to observations on the eigenvalues $(b_k)_{k\geq 1}$. Since $b_k \to 0$ as $k \to +\infty$, the main difficulty is to control the error in $(x_k)_{k\geq 1}$. When k is large, there is mainly noise in x_k. In this case x_k^{-1} is not a good estimator for b_k^{-1}.

There exist also some restrictions on the use of the SVD. If an operator appears in two different expansions, the same bases will be used without care about the object of interest. The bases $(\phi_k)_{k\geq 1}$ and $(\psi_k)_{k\geq 1}$ are indeed equally suitable for A: the representation matrix is diagonal. Nevertheless, the basis $(\phi_k)_{k\geq 1}$ may not be appropriate for representing f. Moreover, the SVD is not always available following the structure of A or may be difficult to compute. The wavelet-vaguelette decomposition (WVD) introduced in Donoho [10] is an interesting alternative to this problem. It combines the simplicity of the SVD framework and the representation efficiency of wavelet bases. However, except for some particular classes of operators, the vaguelettes may be difficult to obtain.

In this paper, we study a more general approach. We would like to make different choices for the bases $(\phi_k)_{k\geq 1}$ and $(\psi_k)_{k\geq 1}$. In this general framework, the operator will be represented by a nondiagonal matrix. This approach has already been studied, for instance, in Mathé and Pereverzev [23], Mair and Ruymgaart [22] or Cohen *et al.* [9]. We also assume the operator to be noisy and consider the following observation:

$$X = A + \sigma\eta,$$

where η is a perturbation operator and σ a noise level. In this setting, Efromovich and Koltchinskii [12] developed an adaptive projection method.

The paper Hoffmann and Reiss [19] is closely related. Using a Galerkin projection approach, they constructed a threshold estimator that attains the minimax rate of convergence on Besov spaces. In particular, they were interested in the case $\sigma > \varepsilon$ and proved that the minimax rate of convergence is related to $\max(\sigma, \varepsilon)$.

Following the principle of unbiased risk estimation, we would like to obtain sharp results in this setting. Given a family of estimators Λ, we want to construct an adaptive estimator that mimics the linear oracle on Λ for any $f \in H$, i.e.,

$$\mathbb{E}_f\|f^\star - f\|^2 \leq (1 + o(1)) \inf_{\tilde{f}\in\Lambda} \mathbb{E}_f\|\tilde{f} - f\|^2 \qquad \text{as} \quad \varepsilon \to 0. \tag{1.6}$$

Inequality (1.6) means that $f^\star$ is asymptotically the best one in this family.

Our aim is to understand the influence of the structure of the matrix A and the noise $\sigma\eta$ on the results. More specifically, we would like to know which kind of assumptions may lead to results similar to (1.6).

This paper is organized as follow. In Section 2, we construct an estimator based on the well-known unbiased risk estimation method. Section 3 contains the main assumptions and main results. Section 4 is devoted to sharp minimax inequalities. In Section 5, we introduce an example of operator satisfying the assumptions of Section 3. Sections 6 and 7 contain the proofs and technical lemmas.

2. Construction of the Estimator

2.1. Linear estimation. When the SVD is chosen to solve (1.1), the selected representation is suitable for A but not always for f. Here, we want more flexibility. Consider $(\phi_k)_{k\geq 1}$ and $(\psi_k)_{k\geq 1}$ a set of orthonormal bases of H and K, respectively, not necessarily associated with the SVD. This set has to be appropriate for f and A. The sequence θ should belong, for instance, to an ellipsoid in L_2 (see Section 4). In Section 3, we introduce some specific assumptions concerning the associated structure of the representation matrix. In Section 5, we present an example, where the chosen bases may be both convenient for f and A.

From now on, the operator A is represented by an infinite matrix which will be denoted by $A = (a_{kl})_{k,l\in\mathbb{N}}$. We use the same notation for both the operator and the matrix but the meaning will be clear from the context.

For all $k \in \mathbb{N}$, replace g by ψ_k in (1.2) to obtain the observations:

$$\langle Y, \psi_k\rangle \overset{\triangle}{=} y_k = \langle Af, \psi_k\rangle + \varepsilon\langle \xi, \psi_k\rangle = \sum_{l=1}^{+\infty} a_{kl}\theta_l + \varepsilon\xi_k, \qquad k \in \mathbb{N}. \tag{2.1}$$

The ξ_k are independent standard Gaussian random variables. In the SVD setting (1.4) or (1.5), each y_k is sufficient to estimate θ_k. In our framework, the approach is rather different. Each y_k gives some information on all the coefficients of the function f.

Following Cohen *et al.* [9] or Hoffmann and Reiss [19], we construct our estimator in two steps: inversion and smoothing.

The inversion step is based on the well-known projection scheme. Projection estimation has been intensively studied in the numerical and statistical analysis. We mention, for instance, Kress [21], Hackbush [17] or Mathé and Pereverzev [23]. Since A is a compact operator, it is not continuously invertible. Therefore, we approximate the infinite matrix A by a sequence $(A_n)_{n\in\mathbb{N}}$ of $n \times n$ matrices: see Böttcher [3] or Efromovich and Koltchinskii [12] for more detail.

For all $n \in \mathbb{N}$, denote by $Y_{(n)}$ the vector ${}^t(y_1, \ldots, y_n)$. Set

$$A_n = \Pi_n A P_n, \tag{2.2}$$

where P_n and Π_n denote the orthogonal projections from H on $H_n = \mathrm{span}(\phi_1, \ldots, \phi_n)$, the subspace of H spanned by $\{\phi_j : j = 1, \ldots, n\}$, and from K to $K_n = \mathrm{span}(\psi_1, \ldots, \psi_n)$, respectively. The corresponding representation matrix is the upper $n \times n$ submatrix of A. From now on, we assume that for all $n \in \mathbb{N}$, A_n is non-singular. The matrix A_n^{-1} always exists. Define

$$\hat{\theta}_n = A_n^{-1} Y_{(n)}. \tag{2.3}$$

This is the classical linear projection estimator. There exist simple choices for n that lead to good minimax efficiency. However, these choices are often related to some a priori information on f as regularity or l^2-norms. Adaptive choice was proposed, for instance, by Efromovich and Kolchinskii [12].

Since our aim is to obtain sharp results, we complete the previous step by smoothing $\hat{\theta}_n$. Let $\lambda = (\lambda_k)_{k\geq 1}$ be a filter, i.e., a real sequence taking values between 0 and 1. Define by F_λ the matrix with entries $f_{jk} = 0$ if $j \neq k$ and $f_{jj} = \lambda_j$, for $1 \leq j, k \leq n$. The corresponding linear estimator will be defined by:

$$\hat{\theta}_{\lambda,n} = F_\lambda A_n^{-1} Y_{(n)} = \sum_{k=1}^{n} \lambda_k \langle A_n^{-1} Y_{(n)}, \phi_k \rangle \phi_k. \tag{2.4}$$

This approach is close to the one of Cohen *et al.* [9] or Hoffmann and Reiss [19].

Now, the operator is supposed to be noisy. Consider the following observation matrix:

$$X = A + \sigma\eta,$$

where η is a random matrix with entries $(\eta_{kl})_{k,l\in\mathbb{N}}$. The η_{kl} are supposed to be i.i.d. standard Gaussian random variables independent of the ξ_k and $\sigma > 0$ a noise level. The case A and η diagonal exactly corresponds to the setting of Cavalier and Hengartner [6]. For all $n \in \mathbb{N}$, set

$$X_n = \Pi_n X P_n. \tag{2.5}$$

Here, we naturally use X_n^{-1} instead of A_n^{-1}. In this situation, the choice of n is also related to the control of the noise in X_n. Indeed, A is compact and not continuously invertible. In this case, X_n^{-1} may not be a good estimator for A_n^{-1} for large values of n.

Given an operator (matrix) T, let $\|T\|$ denotes its operator norm. We consider the following stopping rule:

$$M = \min\left\{ l \leq N_\sigma \colon \|X_l^{-1}\|^2 \geq \frac{1}{\sigma^2 l^2 \log^{1+\tau} \frac{1}{\sigma}} \right\} - 1, \tag{2.6}$$

where $\tau > 0$. The quantity N_σ ensures that M is not too large. Typically, choose $N_\sigma = \sigma^{-2}$. Define also:

$$M_0 = \min\left\{ l \colon \|A_l^{-1}\|^2 \geq \frac{1}{\sigma^2 l^2 \log^{1+2\tau} \frac{1}{\sigma}} \right\} - 1, \tag{2.7}$$

and

$$M_1 = \min\left\{ l \colon \|A_l^{-1}\|^2 \geq \frac{1}{\sigma^2 l^2 \log^{1+\frac{\tau}{2}} \frac{1}{\sigma}} \right\}. \tag{2.8}$$

The bandwidth M is stochastic but Lemma 1 provides that,

$$M_0 < M < M_1$$

with a large probability. In order to control the noise in X_n, we choose $n \leq M$. If $\sigma = 0$, we formally set $M = +\infty$.

In this paper, we assume σ to be $o(\varepsilon)$ or $O(\varepsilon)$ as $\varepsilon \to 0$. The noise in the operator is smaller than or of the same order as in the observations. Interesting results were obtained by Hoffmann and Reiss [19] in the case where $\sigma > \varepsilon$. In particular, they proved that the rate of convergence is related to $\max(\varepsilon, \sigma)$. Our work could certainly be extended to this situation, but the proofs may be rather technical. For the sake of convenience, we assume throughout the paper that $\sigma \leq c\varepsilon$ for some $c > 0$.

2.2. Adaptive estimation. Let λ be a filter, $n \in \mathbb{N}$, and $\hat{\theta}_{\lambda,n}$ the corresponding linear estimator defined in (2.4). The associated quadratic risk is given by:

$$R_\varepsilon^n(\lambda, \theta) = \mathbb{E}_\theta \|F_\lambda A_n^{-1} Y_{(n)} - f\|^2 \tag{2.9}$$
$$= \sum_{k=1}^n \left[(\lambda_k \langle A_n^{-1} \Pi_n A f, \phi_k \rangle - \theta_k)^2 + \varepsilon^2 \lambda_k^2 \sum_{l=1}^n \langle A_n^{-1} \psi_l, \phi_k \rangle^2 \right] + \sum_{k>n} \theta_k^2.$$

Use the following decomposition:

$$A_n^{-1} \Pi_n A = A_n^{-1} \Pi_n A P_n + A_n^{-1} \Pi_n A P_n^\perp = P_n + G_n, \tag{2.10}$$

where $G_n = A_n^{-1} \Pi_n A P_n^\perp$ and $P_n^\perp$ denotes the orthogonal complement projection of P_n. Following Efromovich and Koltchinskii [12], we call G_n the projection error operator. If the set of bases corresponds to the SVD one, $\Pi_n A P_n^\perp$ is equal to 0. The representation matrix A is then diagonal, the operator G_n vanishes, and we obtain the classical quadratic risk of a linear estimator.

We want to select a pair (λ, n) in an adaptive way. A sufficiently large n will not have a great influence on the quality of estimation since the preliminary estimator $\hat{\theta}_n$ defined in (2.3) will be smoothed (see Section 4 for more detail). The quantity n will be fixed later, in Section 2.3. The choice of λ is however a critical step.

Our goal is to construct a filter which will be the best one among a given family Λ, i.e., which will have the smallest risk. If θ is known, the best filter for a fixed n is

$$\lambda_0 = \operatorname*{argmin}_{\lambda \in \Lambda} R_\varepsilon^n(\lambda, \theta). \tag{2.11}$$

It is called the oracle filter corresponding to the family Λ. In order to approximate the optimal filter, we minimize an estimate of the quadratic risk. This well-known idea was developed in Stein [25]. It was also intensively studied by Cavalier *et al.* [5] in the model selection framework. This subsection is devoted to the construction of an estimator for $R_\varepsilon^n(\lambda, \theta)$.

In order to simplify the notation, we denote by $(b_{kl})_{k,l \in \mathbb{N}}$ the entries of the matrix A_n^{-1}. This sequence depends on n but we do not take this into account. First remark that:

$$\mathbb{E}_\theta \left[\langle A_n^{-1} Y_{(n)}, \phi_k \rangle^2 - \varepsilon^2 \sum_{l=1}^n b_{kl}^2 \right] = \langle (P_n + G_n) f, \phi_k \rangle^2. \tag{2.12}$$

Using (2.12), we propose the following estimator for $R^n_\varepsilon(\lambda, \theta)$:

$$U(\lambda, Y) = \sum_{k=1}^n (1-\lambda_k)^2 \left(\langle A_n^{-1} Y_{(n)}, \phi_k \rangle^2 - \varepsilon^2 \sum_{l=1}^n b_{kl}^2 \right) + \varepsilon^2 \sum_{k=1}^n \lambda_k^2 \sum_{l=1}^n b_{kl}^2 .$$

Remark that this estimator is biased since for all integer k, θ_k is approximated by $\langle (P_n + G_n) f, \phi_k \rangle$. These quantities can be estimated by $\langle A_n^{-1} Y_{(n)}, \phi_k \rangle$. The bias term is of order $\|G_n f\|^2$. It can easily be controlled via standard assumption (see Section 4). See also Kress [21] or Efromovich and Koltchinskii [12] for a complete discussion on the operator G_n.

Now, the operator A is supposed to be unknown. Thus we estimate A_n^{-1} by X_n^{-1} (defined in (2.5)). It produces the following estimator of the risk:

$$U(\lambda, X, Y) = \sum_{k=1}^n (1-\lambda_k)^2 \left(\langle X_n^{-1} Y_{(n)}, \phi_k \rangle^2 - \varepsilon^2 \sum_{l=1}^n x_{kl}^2 \right) + \varepsilon^2 \sum_{k=1}^n \lambda_k^2 \sum_{l=1}^n x_{kl}^2, \tag{2.13}$$

where the $x_{kl} = x_{kl}^n$ are the entries of the inverse matrix X_n^{-1}. Since we use squared estimators, some σ^2-correction would be expected. However, introducing such an additional term only increases the difficulty of the proofs and does not improve the theoretical results.

In order to simplify the notation, set for all k in $\{1, \ldots, n\}$:

$$\tilde{y}_k = \langle X_n^{-1} Y_{(n)}, \phi_k \rangle = \langle X_n^{-1} A_n (P_n + G_n) f, \phi_k \rangle + \varepsilon \sum_{l=1}^n x_{kl} \xi_l . \tag{2.14}$$

The random variables $\tilde{y}_k$, $k = 1, \ldots, n$, are correlated if the matrix X_n^{-1} is not diagonal. This is different from the singular value decomposition framework (1.5). The degree of correlation is essentially related to the structure of the matrix A_n^{-1}.

2.3. Blockwise Stein's rule estimator. Our aim is to obtain an oracle inequality similar to (1.6) on the class of the monotone filters:

$$\Lambda_{\text{mon}} = \big\{ \lambda \in \ell_2 \colon 1 \geq \lambda_1 \geq \cdots \geq \lambda_m \geq \cdots \geq 0 \big\}.$$

By analogy with Cavalier and Tsybakov [7], we will proceed step by step. Define the set of the blockwise constant filters by:

$$\Lambda^* = \big\{ \lambda \in l^2 \colon 0 \leq \lambda_k \leq 1,\ \lambda_k = \lambda_{K_j},\ k \in [K_j, K_{j+1} - 1], \\ j = 0, \ldots, J-1 \quad \text{and} \quad \lambda_k = 0 \quad \text{for} \quad k > N \big\},$$

where $J, N, (K_j)_{j=0,\ldots,J-1}$ are integers such that $K_0 = 1$ and $K_J = N+1$. For all $j \in \{1, \ldots, J\}$, set $I_j = \{k \in [K_{j-1}, K_j - 1]\}$ and $T_j = K_j - K_{j-1}$ the length of the block I_j. The set Λ^* is entirely determined by $(T_j)_{j=1,\ldots,J}$ and N. In the next section we propose different possible choices of blocks.

Set $n = N \wedge M \stackrel{\triangle}{=} \inf(N, M)$ in $U(\lambda, X, Y)$ (where M is defined in (2.6)). First minimize this functional in Λ^*. The filter $\tilde{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda^*} U(\lambda, X, Y)$ is given by

$$\tilde{\lambda}_k = \begin{cases} \left(1 - \dfrac{\tilde{\sigma}_j^2}{\|\tilde{y}\|_{(j)}^2}\right)_+, & k \in I_j, \quad j = 1, \ldots, J, \\ 0, & k > n = N \wedge M, \end{cases}$$

where $\tilde{\sigma}_j^2 = \varepsilon^2 \sum_{k\in I_j} \sum_{l=1}^n x_{kl}^2 \ \mathbf{1}_{\{k\leq n\}}$ and $\|\tilde{y}\|_{(j)}^2 = \sum_{k\in I_j} \tilde{y}_k^2 \ \mathbf{1}_{\{k\leq n\}}$. This filter has the following properties:

1. If $\tilde{\sigma}_j^2$ is of order $\|\tilde{y}\|_{(j)}^2$, the quality of estimation may not be very good. In this case, the filter $\tilde{\lambda}$ is close to 0 on the block I_j.
2. Now, if $\tilde{\sigma}_j^2$ is negligible compared to $\|\tilde{y}\|_{(j)}^2$, the estimation has a chance to be good. In this case, $\tilde{\lambda}$ is close to 1.

Only the blocks where estimation has a chance to be good are taken into account. To increase this effect, introduce a penalty in $\tilde{\lambda}$. It produces the penalized blockwise Stein's rule filter:

$$\lambda_k^\star = \begin{cases} \left(1 - \dfrac{\tilde{\sigma}_j^2(1+\varphi_j)}{\|\tilde{y}\|_{(j)}^2}\right)_+, & k \in I_j, \quad j = 1\dots J, \\ 0, & k > N\wedge M, \end{cases} \tag{2.15}$$

where $\varphi_j > 0$ for all $j = 1, \dots, J$. The associated estimator is

$$\theta^\star = F_{\lambda^\star} X_n^{-1} Y_{(n)} \tag{2.16}$$

with $n = N \wedge M$. If $\sigma = 0$, $M = +\infty$ and $n = N$. Under some regularity assumptions on f and A, $\theta^\star$ is asymptotically the best one among Λ_{mon}, see Section 4. We exactly obtain the estimator constructed by Cavalier and Tsybakov [7] if the matrix A is diagonal and $\sigma = 0$.

Remark that $\theta^\star$ has been constructed in a general setup. The result presented in the next section covers both the SVD setting with noisy eigenvalues of Cavalier and Hengartner [6] and the general framework of Efromovich and Koltchinskii [12] and Hoffmann and Reiss [19].

3. Main Results

3.1. NOTATION AND ASSUMPTIONS. To obtain the first oracle inequality, some notation and assumptions are required. Let $\mathcal{B}$ and $\mathcal{M}$ be the events defined in Lemma 2 (see Section 6). For all $j \in \{1, \dots, J\}$, define

$$C_j = \left\{ \|(P_n + G_n)f\|_{(j)}^2 < \varphi_j \frac{\tilde{\sigma}_j^2}{8} \right\} \cap \mathcal{B} \cap \mathcal{M}. \tag{3.1}$$

The following two assumptions concern the structure of the representation matrix A_n^{-1} defined in (2.2).

Assumption A1: There exists a positive constant $b_\star$ such that, for some $\beta \geq 0$,

$$\sum_{l=1}^m \langle A_m^{-1}\psi_l, \phi_k\rangle^2 = b_\star k^{2\beta}(1+o(1)) \qquad \text{as} \quad k, m \to +\infty.$$

In particular, $\|A_m^{-1}\| = O(m^{2\beta})$ as $m \to +\infty$.

Assumption A2: There exists a constant $c_1 \geq 1$ independent of ε, such that, for all $j \in \{2, \dots, J\}$ and for all $m \in \{N\wedge M_0, \dots, N \wedge M_1\}$,

$$T_j \sum_{\substack{l=1 \\ l\neq k}}^m \langle A_m^{-1}\psi_l, \phi_k\rangle^2 \leq c_1 \sum_{l=1}^m \langle A_m^{-1}\psi_l, \phi_k\rangle^2, \qquad \forall k \in I_j, \quad k \leq m.$$

Assumption A1 allows for a better understanding of the problem. In the special case of the SVD setting (1.4), this assumption is very standard and corresponds to mildly ill-posed problems (see Engl *et al.* [13]). The parameter β quantifies the degree of ill-posedness of the problem.

In the SVD setting (see (1.4) or (1.5)), Assumption A2 is directly verified. Indeed, the non-diagonal terms of the representation matrix A_m^{-1} are zero. This assumption means that the matrix A_m^{-1} has to be close to a diagonal one. We want the main terms to be concentrated on the diagonal. Assumption A2 is in fact related to the degree of correlation of the variables $\tilde{y}_k$ in each block I_j (see (2.14)). If the correlation is too large, it will be difficult to detect the signal θ. Remark that Assumption 3.4 of Hoffmann and Reiss [19] is closely related. They require some sparsity concerning the structure of the representation matrix A.

Assumption A3: If the matrices A and η are not diagonal, the penalty is chosen so that

$$\varphi_j > c_3 \frac{\sigma^2}{\varepsilon^2} \|(P_n + G_n) f\|^2, \qquad \forall j \in \{1, \dots, J\}.$$

Here c_3 denotes a positive constant independent of ε. It can be explicitly computed (see the proof of Lemma 3 in Section 6 for more detail).

Assumption A3 enables us to control the quantity $P(\{\lambda^\star_{K_j-1} > 0\} \cap C_j)$. On C_j the signal is negligible compared to $\tilde{\sigma}_j^2 \varphi_j$. We cannot expect a good estimation. The probability that $\lambda^\star_{K_{j-1}} > 0$ on C_j has to be very small. Lemma 3 in Section 6 provides an upper bound for this quantity. It is larger than in Cavalier and Tsybakov [7]. Indeed, it is more difficult for $\theta^\star$ to detect that the signal is too small if the operator is noisy and non-diagonal. Remark that the penalty will not be the same for SVD or non-diagonal setting. In Section 4, we discuss the different possible values of the penalty considering the structure of the operator and the noise levels σ and ε.

The construction of $\theta^\star$ should be modified in the particular case where $\sigma \geq \varepsilon$. Indeed, it is clear that a penalty satisfying A3 in this situation may lead to a poor efficiency.

3.2. Main result. For the sake of convenience, we will present our result for a specific class of blocks. Following Nemirovskii [24] or Cavalier and Tsybakov [7], we use weakly geometrically increasing blocks. Set $\nu_\varepsilon = \log 1/\varepsilon$ and $\rho_\varepsilon = \log^{-1} \nu_\varepsilon$. The size of blocks is defined by

$$\begin{cases} T_1 = \lceil \nu_\varepsilon \rceil, \\ T_j = \lceil \nu_\varepsilon (1+\rho_\varepsilon)^{j-1} \rceil, \quad j > 1, \end{cases} \tag{3.2}$$

and we set $N = \varepsilon^{-2}$. The penalty should be chosen so that $\varphi_j \geq (\nu_\varepsilon \rho_\varepsilon)^{-\gamma}$, where $0 < \gamma < \frac{1}{2}$. In Section 3.3, we present some other possible choices of blocks and related penalties.

The following proposition is the main result of this paper. We will use it in Section 4 to obtain sharp minimax results subject to some regularity assumptions on θ.

Proposition 1. *Assume that* A1–A3 *hold and* $\|\theta\| \leq r$, *for some* $r > 0$. *Let* $\theta^\star$ *be the estimator defined in* (2.15), (2.16), *and* $B > 0$. *For all* $j = 1, \dots, J$, *there exists* $g_j = g_j(\|\theta\|, \varphi_j, B)$ *such that*

$$\begin{aligned}(3.3)\quad \mathbb{E}_\theta\|\theta^\star - \theta\|^2 \leq & \max_{j=1,\dots,J}(1+g_j) \inf_{\lambda \in \Lambda_{\text{mon}}} R_\varepsilon^N(\lambda, \theta) \\ & + c\left(1 + \frac{\sigma^2}{\varepsilon^2}\log^{1+\tau}\frac{1}{\sigma}\right)\varepsilon^2 \nu_\varepsilon^{2\beta+1} + c\frac{1+B^{-1}}{1-B}\mathbb{E}_\theta\|G_n f\|^2 + \Gamma(\theta) + c\Omega,\end{aligned}$$

where G_n *is defined in* (2.10), $c > 0$, $\Omega = M_1 e^{-\log^{1+\tau} 1/\sigma}$, *and*

$$(3.4)\qquad \Gamma(\theta) = \mathbb{E}_\theta\left|\sum_{k=1}^n \varepsilon^2(\lambda_k^0)^2 \sum_{l=1}^n \left(\langle A_n^{-1}\psi_l, \phi_k\rangle^2 - \langle A_M^{-1}\psi_l, \phi_k\rangle^2\right) + \sum_{k=M_0}^\infty \theta_k^2\right|$$

with $\lambda^0 = \operatorname{argmin}_{\lambda \in \Lambda_{\text{mon}}} R_\varepsilon^N(\lambda, \theta)$.

The proof of this proposition is given in Section 7. The functionals g_j are defined in (7.15) and (7.19). We will show that for some particular cases, $g_j \to 0$ as $\varepsilon \to 0$ uniformly in $j = 1, \dots, J$. Moreover, many residual terms appear in inequality (3.3). The goal is to show that they are negligible compared to $R_\varepsilon^N(\lambda^0, \theta)$. This will be done later, in Section 4, by considering specific classes of functions.

The term $\Gamma(\theta)$ corresponds to the rest of the risk $R_\varepsilon^N(\lambda^0, \theta)$ truncated at the order M. If M is large enough, $\Gamma(\theta)$ will be negligible compared to $R_\varepsilon^N(\lambda^0, \theta)$. The two terms Ω and $\Gamma(\theta)$ appear with the noise in the operator. These quantities were introduced for the first time in Cavalier and Hengartner [6].

Remark that the oracle inequality (3.3) is obtained for a fixed N. Since N is large enough, it has not a great influence on the quality of estimation. This is proved in Section 4 using Assumption A1.

3.3. Choice of blocks and penalties. In Cavalier and Tsybakov [8], some other choices of blocks are presented in the SVD setting with $\beta = 0$. Such blocks can be used in this framework. We only recall here the available choices.

- **Constant size blocks depending on** ε. The size is defined by

$$T_j = \left\lceil C_* S\left(\frac{1}{\varepsilon}\right)\log^{1+\tau}\frac{1}{\varepsilon}\right\rceil,$$

  where $S(\frac{1}{\varepsilon}) = \log\log\frac{1}{\varepsilon}$, $C_* > 0$, and the penalty satisfies $\varphi_j \geq S^{-1/2}(\frac{1}{\varepsilon})$.
- **Increasing blocks independent of** ε. The size is defined by one of the three following expressions:

$$T_j = \lceil C_1 j^\rho \rceil, \qquad T_j = \lceil C_1 \exp(j^\rho)\rceil, \qquad T_j = \lceil C_1 j \mu(j) \log j\rceil,$$

  and the penalty should be chosen so that $\varphi_j \geq \mu(j)^{-1/2}$, where $\mu(j) = \log\log(j+20)$.

In these particular cases, Assumption A1 should be verified only for $j \in \{n_\varepsilon, \dots, J\}$, where $n_\varepsilon \to +\infty$ as $\varepsilon \to 0$. The term n_ε is the same as in Cavalier and Tsybakov [7].

4. Minimax Adaptation

In this section, we apply the results of Section 3 to show that our estimator has good minimax properties. Assume that the sequence θ belongs to an ellipsoid:

$$\Theta = \Theta(v, Q) = \left\{ \theta \colon \sum_{k=1}^{+\infty} v_k^2 \theta_k^2 \leq Q \right\},$$

where $v = (v_k)_{k\in\mathbb{N}} \subset \mathbb{R}$ and $Q > 0$. The minimax risk on Θ is

$$r_\varepsilon(\Theta) = \inf_{\tilde\theta} \sup_{\theta\in\Theta} \mathbb{E}_\theta \|\tilde\theta - \theta\|^2.$$

The infimum is taken over all the estimators of θ based on the observations (1.1). In the sequel, we write $v \simeq (k^\alpha)_{k\geq 1}$ for some $\alpha \in \mathbb{N}$ if we can find positive constants d_0 and d_1 such that $d_0 k^\alpha \leq v_k \leq d_1 k^\alpha$ for all $k \in \mathbb{N}$.

Proposition 1 is very general. It involves weak assumptions on the operator A and on the noise σ. In order to obtain precise minimax results, additional restrictions are needed. We will consider in this section three different settings:

- the SVD case with noisy eigenvalues,
- the non-diagonal case, where A is completely known ($\sigma = 0$),
- the non-diagonal case with noise in the operator.

We only present the results for the weakly geometrically increasing blocks. The proofs in the other cases introduced in Section 3.3 exactly follow the same lines. We leave them to the interested reader.

4.1. The SVD case. Assume that the set of bases $(\phi_k)_{k\geq 1}$ and $(\psi_k)_{k\geq 1}$ exactly corresponds to the SVD one. The matrix η is supposed to be diagonal. For all $k \in \mathbb{N}$, we observe:

$$\left\{ \begin{array}{l} y_k = b_k \theta_k + \varepsilon \xi_k, \\ x_k = b_k + \sigma \eta_k. \end{array} \right. \tag{4.1}$$

This is exactly the setting of Cavalier and Hengartner [6].

Theorem 1. *Let $\Theta = \Theta(v, Q)$ be an ellipsoid with monotone non-decreasing $v \simeq (k^\alpha)_{k\geq 1}$, $\alpha > 1/2$, and $Q \in [0, Q_0]$ for some $Q_0 > 0$. Assume that Assumption* A1 *holds. Choose $\varphi_j = (\nu_\varepsilon \rho_\varepsilon)^{-\gamma}$ for $0 < \gamma < 1/2$. Then, the estimator $\theta^\star$ defined by* (2.15), (2.16) *satisfies*:

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta \|\theta^\star - \theta\|^2 = (1 + o(1)) r_\varepsilon(\Theta) \qquad as \quad \varepsilon \to 0. \tag{4.2}$$

The parameter α represents the smoothness of the functions contained in Θ. Our estimator is adaptive, since it does not depend on this parameter.

Theorem 4.2 of Cavalier and Hengartner [6] is slightly different from inequality (4.2) since the framework was the model selection one. They only consider finite families of estimators.

The same result as (4.2) was obtained in Cavalier and Tsybakov [7] with $\sigma = 0$. From an asymptotic point of view, the only difference appears in the quantity $o(1)$. It is significantly smaller if $\sigma = 0$ (see the proof of Proposition 1 in Section 7). Regularization is easier without noise in the operator.

Proof. In this setting, the operator G_n vanishes. The representation matrix A is diagonal, Assumption A2 holds. Assumption A3 does not concern the SVD case. By a direct application of Proposition 1,

$$\mathbb{E}_\theta\|\theta^\star - \theta\|^2 \leq \max_{j=1,\dots,J}(1+g_j) \inf_{\lambda\in\Lambda_{\mathrm{mon}}} R_\varepsilon^N(\lambda,\theta) + c\varepsilon^2\left(1+\frac{\sigma^2}{\varepsilon^2}\log^{1+\tau}\frac{1}{\sigma}\right)\nu_\varepsilon^{2\beta+1} + \Gamma(\theta) + \Omega, \tag{4.3}$$

where g_j is defined in (7.15) and (7.19) for all $j = 1,\dots,J$. The idea of the proof is first to show that $g_j \to 0$ as $\varepsilon \to 0$, and then that the three residual terms in (4.3) are negligible compared to $r_\varepsilon(\Theta)$.

Let $j \in \{1,\dots,J\}$. Using Cavalier and Tsybakov [7], $\Delta_j \leq (\rho_\varepsilon\nu_\varepsilon)^{-1}(1+\rho_\varepsilon)^{-j}$ and

$$\frac{(\varphi_j^2 + c\Delta_j)}{\varphi_j} \leq \left[(\rho_\varepsilon\nu_\varepsilon)^{-\gamma} + (\rho_\varepsilon\nu_\varepsilon)^{\gamma-1}(1+\rho_\varepsilon)^{-j}\right] \to 0 \qquad \text{as} \quad \varepsilon \to 0.$$

Since $K_j \geq \nu_\varepsilon(1+\rho_\varepsilon)^j$ and $\sigma = O(\varepsilon)$ as $\varepsilon \to 0$,

$$\frac{\sigma^2}{\varepsilon^2}\frac{\|\theta\|_{(j)}}{\varphi_j}\log^{1+\tau}\frac{1}{\sigma} \leq cK_j^{-2\alpha}(\nu_\varepsilon\rho_\varepsilon)^\gamma \log^{1+\tau}\left(\frac{1}{\sigma}\right)\frac{\sigma^2}{\varepsilon^2} \to 0 \qquad \text{as} \quad \varepsilon \to 0. \tag{4.4}$$

Thus, $\max_{j=1,\dots,J} g_j \to 0$ with a good choice of B. By a direct application of Pinsker's Theorem (see Belitser and Levit [2]) and simple calculation,

$$\inf_{\lambda\in\Lambda_{\mathrm{mon}}} \sup_{\theta\in\Theta} R_\varepsilon^N(\lambda,\theta) = (1+o(1))\sup_{\theta\in\Theta} R_\varepsilon(\hat\lambda,\theta) = (1+o(1))r_\varepsilon(\Theta),$$

where $\hat\lambda$ is Pinsker's estimator and $R_\varepsilon(\lambda,\theta)$ denotes the classical quadratic risk of a linear estimator.

We now focus on the residual terms in (4.3). Under our assumptions, Belitser and Levit [2] show that for all $\varepsilon > 0$,

$$r_\varepsilon(\Theta) = O(\varepsilon^{\frac{4\alpha}{2\beta+2\alpha+1}}) \qquad \text{as} \quad \varepsilon \to 0. \tag{4.5}$$

The term Ω is clearly negligible compared to $r_\varepsilon(\Theta)$. Now remark that

$$\Gamma(\theta) \leq \sum_{i=M_0}^\infty \varepsilon^2(\lambda_i^0)^2 b_i^{-2} + \sum_{i=M_0}^\infty \theta_i^2 \leq \sum_{i=M_0}^\infty \frac{\varepsilon^2 b_i^{-2}\theta_i^2}{\varepsilon^2 b_i^{-2} + \theta_i^2} + \sum_{i=M_0}^\infty \theta_i^2 \leq 2\sum_{i=M_0}^\infty \theta_i^2.$$

Clearly, $\Gamma(\theta) = o(r_\varepsilon)$. Thus

$$\sup_{\theta\in\Theta}\mathbb{E}_\theta\|\theta^\star - \theta\|^2 = (1+o(1))r_\varepsilon(\Theta).$$

This concludes the proof. □

4.2. Non-diagonal case with $\sigma = 0$. The case $\sigma = 0$ corresponds to the situation, where A is completely known. We observe

$$y_k = \sum_{l=1}^{\infty} a_{kl}\theta_l + \varepsilon\xi_k, \qquad k \in \mathbb{N}.$$

The results here are somewhat less precise than in Theorem 1. Indeed, a generalization of Pinsker's Theorem to this framework has not been established yet.

Theorem 2. *Let $\Theta = \Theta(v, Q)$ be an ellipsoid with monotone non-decreasing $v \simeq (k^\alpha)_{k\geq 1}$, $Q > 0$, and $\alpha > 1/2$. Assume that Assumptions* A1 *and* A2 *hold and that $\|G_m\| \to 0$ as $m \to +\infty$. Choose $\varphi_j = (\nu_\varepsilon\rho_\varepsilon)^{-\gamma}$ for $0 < \gamma < 1/2$. Then, the estimator $\theta^\star$ defined by* (2.15) *and* (2.16) *satisfies*

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta\|\theta^\star - \theta\|^2 = (1 + o(1)) \inf_{\lambda,m} \sup_{\theta\in\Theta} R_\varepsilon^m(\lambda, \theta)$$

as $\varepsilon \to 0$, where the infimum is taken over all filters $\lambda \in \Lambda_{\mathrm{mon}}$ and bandwidths m.

Proof. The noise level σ is zero. Assumption A3 is satisfied since $\varphi_j > 0$. A direct application of Proposition 1 provides:

$$\mathbb{E}_\theta\|\theta^\star - \theta\|^2 \leq \max_{j=1,\ldots,J}(1 + g_j) \inf_{\lambda\in\Lambda_{\mathrm{mon}}} R_\varepsilon^N(\lambda, \theta) + c\varepsilon^2\nu_\varepsilon^{2\beta+1} + c\mathbb{E}_\theta\|G_n f\|^2,$$

where

$$g_j = \frac{c(\varphi_j^2 + 4\Delta_j)}{\varphi_j}(1 + o(1)) \qquad \text{as} \quad \varepsilon \to 0.$$

As in Theorem 1, $g_j \to 0$ as $\varepsilon \to 0$. By Assumption A1, there exists $t(\varepsilon) \to 0$ as $\varepsilon \to 0$, independent of k and m such that

$$\left| \sum_{l=1}^{m} \langle A_m^{-1}\psi_l, \phi_k\rangle^2 - b_\star^2 k^{2\beta} \right| \leq b_\star^2 k^{2\beta} t(\varepsilon), \qquad \forall k \geq \nu_\varepsilon, \quad \forall m \geq \tilde{m},$$

where $\tilde{m}$ is the information complexity of the problem (see Mathé and Pereverzev [23] for more detail). Since $\|G_m\| \to 0$ as $m \to +\infty$, it is then easy to see that,

$$\inf_\lambda R_\varepsilon^N(\lambda, \theta) \leq (1 + o(1)) \inf_{\lambda,m} R_\varepsilon^m(\lambda, \theta) + c\varepsilon^2\nu_\varepsilon^{2\beta+1} \qquad \text{as} \quad \varepsilon \to 0$$

uniformly in $\theta \in \Theta(a, Q)$. The residual term $\varepsilon^2\nu_\varepsilon^{2\beta+1}$ is negligible. Indeed, the minimax rate of convergence is of the same order as in the SVD case (see Efromovich and Koltchinskii [12]). In particular, the projection estimator attains the optimal rate of convergence and is linear and monotone. To conclude the proof, just remark that for all m,

$$\|G_m f\|^2 = \|G_m P_m^\perp f\|^2 \leq \|G_m\|^2 \|P_m^\perp f\|^2 \leq C^2 \sum_{k>m} \theta_k^2.$$

By simple calculation the quantity $\mathbb{E}_\theta\|G_n f\|^2$ is negligible. $\square$

Our estimator $\theta^\star$ attains the minimax rate of convergence on the ellipsoid Θ. Moreover, it is asymptotically the best one compared to the family of estimators defined in (2.4).

If we only assume that the operator G_n is uniformly bounded, the estimator $\theta^\star$ only attains the minimax rate of convergence on $\Theta(v,Q)$. This assumption is very standard. This is a necessary condition for the convergence of the projection method (see Kress [21], Böttcher [3]). The assumption $\|G_m\| \to 0$ is more restrictive. It guarantees that the projection error $A_n^{-1}\Pi_n Af - f$ can be controlled by the familiar bias $P_n^\perp f$ (see (2.9) and (2.10)). We refer to Kress [21] or Efromovich and Koltchinskii [12] for more detail.

4.3. The non-diagonal case. We finish this section with a general setting. Assume we have at our disposal the observations:

$$\begin{cases} Y = Af + \varepsilon\xi, \\ X = A + \sigma\eta. \end{cases}$$

This is the setting of Efromovich and Koltchinskii [12] and Hoffmann and Reiss [19]. We obtain the following result:

Theorem 3. *Let $\Theta = \Theta(v,Q)$ be an ellipsoid with monotone non-decreasing $v \simeq (k^\alpha)_{k\geq 1}$, $Q \in [0, Q_0]$, and $\alpha > 1/2$. Assume that Assumptions* A1 *and* A2 *hold and $\|G_m\| \to 0$ as $m \to +\infty$.*

(i) *Choose $\varphi_j > 16Q_0\sigma^2/\varepsilon^2$. Then, there exists a positive constant c such that*

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta\|\theta^\star - \theta\|^2 \leq c r_\varepsilon(\Theta)\log^{1+\tau}\frac{1}{\varepsilon}. \tag{4.6}$$

The estimator $\theta^\star$ attains the minimax rate of convergence up to a log term.

(ii) *Assume $\sigma\log^{1+\tau}(1/\sigma) = o(\varepsilon)$ as $\varepsilon \to 0$. Choose $\varphi_j = (\nu_\varepsilon\rho_\varepsilon)^{-\gamma}$ for $0 < \gamma < 1/2$. Then,*

$$\sup_{\theta\in\Theta} \mathbb{E}_\theta\|\theta^\star - \theta\|^2 = (1+o(1))\inf_{\lambda,m}\sup_{\theta\in\Theta} R_\varepsilon^m(\lambda,\theta) \qquad \text{as} \quad \varepsilon \to 0.$$

Proof. Using simple algebra and Assumption A1

$$\Gamma(\theta) \leq \mathbb{E}_\theta \sum_{k=M_0}^{+\infty}\left(\lambda_k^0\varepsilon^2\sum_{l=1}^{+\infty} b_{kl}^2 + \theta_k^2\right) + 2b_\star^2\varepsilon^2 t(\varepsilon)\mathbb{E}_\theta\sum_{k=\nu_\varepsilon}^{M_1}(\lambda_k^0)^2 k^{2\beta} + c\varepsilon^2\nu_\varepsilon^{2\beta+1},$$

where $\lambda^0 = \mathrm{argmin}_{\lambda\in\Lambda_{\mathrm{mon}}} R_\varepsilon^N(\lambda,\theta)$. Therefore, the term $\Gamma(\theta)$ is negligible for ε small enough.

The end of the proof follows the same lines as for Theorems 1 and 2. The only difference is in the expression of g_j (see (7.15) and (7.19)). Consider the quantity:

$$\frac{Q_0}{\varphi_j^2}\frac{\sigma^2}{\varepsilon^2}\log^{1+\tau}\frac{1}{\sigma}.$$

If $\sigma \log^{1+\tau}(1/\sigma) = o(\varepsilon)$, it vanishes with $\varphi_j = (\nu_\varepsilon \rho_\varepsilon)^{-\gamma}$. Assumption A3 is satisfied for ε small enough and $g_j \to 0$ as $\varepsilon \to 0$. The estimator $\theta^\star$ produces sharp minimax results. Otherwise, we choose $\varphi_j > 16 Q_0 \sigma^2/\varepsilon^2$. Assumption A3 is satisfied, but g_j does not vanish as $\varepsilon \to 0$. Our estimator attains the minimax rate of convergence up to a log term. □

The assumption $\sigma \log^{1+\tau} 1/\sigma = o(\varepsilon)$ as $\varepsilon \to 0$ means that the noise in the operator is smaller than ε from an asymptotic point of view. One may think either of numerical measurements of the operator with a high quality or of independent observations with a larger number of data.

The degree of ill-posedness of the problem is supposed to be unknown. For this reason, a log term appears in the right-hand side of (4.6). If β is known, we can see in the proofs that the log term is not needed anymore. Indeed, choose $M = M_0$. In this case, M is not stochastic and quantities as $\varepsilon \|X_n^{-1} A_n f\|^2$ are easier to control.

In case β is known and $\sigma = O(\varepsilon)$, our estimator $\theta^\star$ produces the same rates of convergence as the projection estimator constructed in Efromovich and Koltchinskii [12]. Indeed, using Lepski's method, they proposed an adaptive estimator that attains the minimax rate of convergence on ellipsoids.

The framework of Hoffmann and Reiss [19] is a little bit different. They considered functions that belong to Besov spaces and constructed a threshold estimator $\tilde{f}$ that attains the minimax rate of convergence up to a log term. In particular, it satisfies:

$$\sup_{f \in V_p^s(Q)} E_f \|\tilde{f} - f\|^2 \leq c \max(\sigma, \varepsilon)^{-\frac{2s}{2s+2\beta+1}},$$

where $V_p^s(Q)$ is a Besov ball. The rate of convergence is thus related to the largest noise. They proved that this rate is optimal. This could certainly be generalized to our framework. However, some modifications in the construction of $\theta^\star$ are required in this situation. Indeed, a penalty term satisfying Assumption A3 may lead to very bad rates of convergence for $\varepsilon = o(\sigma)$ as $\sigma \to 0$. In this particular case, the solution would be to replace ε by $\max(\varepsilon, \sigma)$ in the construction of $\theta^\star$.

4.4. Conclusion. This paper generalizes the results of Cavalier and Tsybakov [7]. Here, two different problems have been treated.

The first one concerns non-diagonal representation matrices. Our estimator produces sharp minimax results subject to some assumptions on the sparsity of A. This seems to be the price to pay when using blockwise estimator in this setting (see, for example, Assumption 3.4 of Hoffmann and Reiss [19]). This model can be compared to direct observation with correlated noise. The structure of the matrix A_n^{-1} is related to the degree of correlation. Our results could be certainly extended to inverse problems with correlated data. In this case, Assumption A2 should be replaced by a correlation assumption.

The second problem concerns the noise in the operator. In the SVD representation, this noise has no real influence on the construction of $\theta^\star$. Problems appear when considering non-diagonal perturbation matrices. The regularization problem is perturbed by the estimation of A_M^{-1}. In this case, the penalty should be chosen large enough in order to control the noise in the operator.

We can expect a bad quality of recovery for large values of $\|\theta\|$, even in the SVD case. Indeed, the quantity $\sigma^2/\varepsilon^2 \|\theta\|^2$ explicitly appears in (3.3). This property is not specific to the penalized blockwise Stein's rule estimator: we refer, for instance, to the results of Efromovich and Koltchinskii [12], Cavalier and Hengartner [6] and Hoffmann and Reiss [19].

5. Example

This example is inspired by Cavalier [4] who studied the fractional integration in the non-periodic framework. We refer to Zygmund [27] for more detail on the fractional integration operator.

First consider the operator $d^{-\beta}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ defined by

$$d^{-\beta}f\colon \mathbb{R} \to \mathbb{R}, \qquad x \to d^{-\beta}f(x) = \int_{-\infty}^{x} \frac{(x-t)^{\beta-1}}{\Gamma(\beta)} f(t)\,dt.$$

This is the periodic version of the fractional integration. It can be proved that $d^{-\beta}e^{i2\pi k\cdot}(x) = (i2\pi k)^{-\beta}e^{i2\pi kx}$ for all $k \in \mathbb{N}$ and $x \in \mathbb{R}$. The Fourier basis is thus associated with the singular value decomposition with eigenvalues $((i2\pi k)^{-\beta})_{k\in\mathbb{N}}$.

One may imagine that the degree of ill-posedness (i.e., the parameter β) is unknown. Suppose that we can send each ϕ_k as an input function f (as in (1.1)) and observe independently the corresponding Y_k:

$$Y_k = d^{-\beta}\phi_k + \varepsilon\eta,$$

where η is a Gaussian white noise. Since for all $k \in \mathbb{N}$, $\langle Y_k, \phi_k\rangle = b_k + \varepsilon\eta_k$, we exactly obtain a sequence of observations on the eigenvalues as in (1.5) with $\sigma = \varepsilon$. Such an approach can easily be extended to every convolution operator (see Cavalier and Hengartner [6] for more detail).

Now consider the non-periodic version of the fractional integration. Let $D^{-\beta}$: $L^2(0,1) \to L^2(0,1)$ defined by:

$$D^{-\beta}f\colon [0,1] \to \mathbb{R}, \qquad x \to D^{-\beta}f(x) = \int_{0}^{x} \frac{(x-t)^{\beta-1}}{\Gamma(\beta)} f(t)\,dt.$$

We are interested in functions that belong to some Sobolev balls:

$$f \in W(\alpha, Q) = \left\{ f\colon \int_0^1 (f^{(\alpha)}(t))^2 dt \le Q \right\}.$$

Here, we consider the case where f is not 1-periodic to avoid boundary effects. In this case, the Fourier basis is not suited for representing f, spline bases are preferable. In order to represent $D^{-\beta}$, choose the spline basis defined in Cavalier [4] as $(\phi_k)_{k\in\mathbb{N}}$ and $(\psi_k)_{k\in\mathbb{N}}$.

The associated representation matrix A can be written as $A = DS$, where D denotes a diagonal matrix with eigenvalues $d_k = k^{-\alpha}(1 + o(1))$ as $k \to +\infty$ and $S = I + V$, where I is the identity matrix. The matrix V cannot be computed explicitly. We do not obtain a finite bound in the associated Hilbert–Schmidt norm. Nevertheless, we may conjecture that V represents a compact operator. Efromovich and Koltchinskii [12] proved that for a matrix associated with such a decomposition

$$\|G_m\| \to 0 \qquad \text{as} \quad m \to +\infty.$$

Moreover, using the properties of the spline bases, we can show in this case that, uniformly in k,

$$\sum_{l=1}^{m} \langle A_m^{-1}\psi_l, \phi_k\rangle^2 \to \sum_{l=1}^{+\infty} \langle A^{-1}\psi_l, \phi_k\rangle^2 \qquad \text{as} \quad m \to +\infty \tag{5.1}$$

and

$$\sum_{l=1,\ l\neq k}^{m} \langle A_m^{-1}\psi_l, \phi_k\rangle^2 \to \sum_{l=1,\ l\neq k}^{+\infty} \langle A^{-1}\psi_l, \phi_k\rangle^2 \qquad \text{as} \quad m \to +\infty. \tag{5.2}$$

The structure of the matrices A_m^{-1} is asymptotically the same as that of A^{-1}. In this case, Assumptions A1 and A2 hold for ε small enough since the main terms of the representation matrix A^{-1} are concentrated on the diagonal (for more detail, see the proof and, in particular, equations (34) and (35) of Cavalier [4]).

More generally, every representation matrix of the form $A = D(I+V)$, where D is diagonal and V is compact is a good candidate for satisfying Assumption A2, provided A^{-1} possesses a quasi-diagonal structure. This corresponds to compact perturbation of the SVD representation.

6. Technical Lemmas

For all $j = 1, \ldots, J$, define the quantities:

$$\Delta_j = \max_{m=N\wedge M_0,\ldots,N\wedge M_1} \frac{\max_{k\in I_j} \sum_{l=1}^{m} \langle A_m^{-1}\psi_l, \phi_k\rangle^2}{\sum_{k\in I_j} \sum_{l=1}^{m} \langle A_m^{-1}\psi_l, \phi_k\rangle^2} \tag{6.1}$$

and for all $h \in H$,

$$l_j(h) = \begin{cases} \|h\|_{(j)}^2 & \text{in the SVD setting,} \\ \|h\|^2 & \text{else.} \end{cases} \tag{6.2}$$

The following lemma provides a control for the stochastic bandwidth M.

Lemma 1. *Let M, M_0, and M_1 be defined in Section* 2 *by* (2.6)–(2.8). *Then*

$$P\big(\{M < M_0\} \cup \{M > M_1\}\big) = O(\Omega) \qquad \textit{as} \quad \sigma \to 0.$$

Proof. Remark that $A_{M_1}^{-1} = (I + \sigma A_{M_1}^{-1}\eta_{M_1})X_{M_1}^{-1}$. This implies

$$\|A_{M_1}^{-1}\| \le \|X_{M_1}^{-1}\|\,\|I + \sigma A_{M_1}^{-1}\eta_{M_1}\|.$$

The probability that M is greater than M_1 is

$$\begin{aligned} P(M > M_1) &\le P\left(\|X_{M_1}^{-1}\|^2 \le \frac{1}{\sigma^2 M_1^2 \log^{1+\tau}\frac{1}{\sigma}}\right) \\ &\le P\left(\|I + \sigma A_{M_1}^{-1}\eta_{M_1}\| \ge \|A_{M_1}^{-1}\|\sigma M_1 \log^{\frac{1+\tau}{2}}\frac{1}{\sigma}\right) \\ &\le P\left(\|\eta_{M_1}\| \ge M_1 \log^{\frac{1+\tau}{2}}\frac{1}{\sigma} - M_1 \log^{\frac{1}{2}+\frac{\tau}{4}}\frac{1}{\sigma}\right) \le c e^{-\log^{1+\tau}\frac{1}{\sigma}}. \end{aligned}$$

Indeed, we use (2.6), (2.8), and Lemma 6.1 of Efromovich and Koltchinskii [12]. Now, remark that, using (2.7),

$$P(M < M_0) \leq P\left(\bigcup_{l=1}^{M_0}\left\{\|X_l^{-1}\|^2 \geq \frac{1}{\sigma^2 l^2 \log^{1+\tau}\frac{1}{\sigma}}\right\}\right).$$

Using (2.8) and Lemma 2, $P(M > M_1) \leq ce^{-\log^{1+\tau}\frac{1}{\sigma}} \Rightarrow P(\mathcal{B}^c) = O(\Omega)$ and

$$\|X_l^{-1}\|\mathbf{1}_{\mathcal{B}} \leq 2\|A_l^{-1}\|, \qquad \forall l \leq M_1.$$

Thus, $P(M < M_0) = O(\Omega)$ as $\sigma \to 0$. $\square$

Lemma 2. *Define the events*

$$\mathcal{B} = \bigcap_{m=1}^{M}\left\{\sigma\|A_m^{-1}\|\|\eta_m\| \leq \frac{1}{2\sqrt{m}}\right\} \quad \text{and} \quad \mathcal{M} = \{M_0 < M < M_1\}. \tag{6.3}$$

Let n be the quantity defined in Section 2.3, $h \in H$ and $B > 0$. Let $k \in \{1, \ldots, n\}$. The following relations hold:

(i) $P(\mathcal{B}^c) = O(\Omega)$,

(ii) $\|X_n^{-1}\|\mathbf{1}_{\mathcal{B}} \leq 2\|A_n^{-1}\|$,

(iii) $\mathbb{E}_\theta\|X_n^{-1}A_n h\|_{(j)}^2\mathbf{1}_{\mathcal{B}} \leq (1+B)\|h\|_{(j)}^2 + \left(1+\frac{1}{B}\right)l_j(h)\frac{\sigma^2}{\varepsilon^2}c_\sigma\mathbb{E}_\theta\sigma_j^2 + \frac{c\Omega}{M_0}$,

(iv) $P\left(\frac{1}{1+c/\sqrt{M_0}}\sum_{l=1}^n b_{kl}^2 \leq \sum_{l=1}^n x_{kl}^2\mathbf{1}_{\mathcal{M}\cap\mathcal{B}} \leq \left(1+\frac{8}{\sqrt{M_0}}\right)\sum_{l=1}^n b_{kl}^2\right) = 1 - O(\Omega)$,

where $c > 0$, $\sigma_j^2 = \varepsilon^2\sum_{k\in I_j}\sum_{l=1}^n\langle A_n^{-1}\psi_l, \phi_k\rangle^2$, and $c_\sigma = \log^{1+\tau}1/\sigma$.

Proof. Using Lemma 6.1 of Efromovich and Koltchinskii [12], for all $m \in \mathbb{N}$,

$$P\left(\sigma\|A_m^{-1}\|\,\|\eta_m\| \geq \frac{1}{2\sqrt{m}}\right) \leq \exp\left[-\frac{1}{32\sigma^2 m^2\|A_m^{-1}\|^2}\right].$$

Thus, using Lemma 1,

$$P(\mathcal{B}^c) \leq M_1\exp\left[-\frac{1}{32\sigma^2(M_1-1)^2\|A_{(M_1-1)}^{-1}\|^2}\right] + O(\Omega) = O(\Omega).$$

For the second inequality use, as in Cavalier and Hengartner [6], a Taylor expansion:

$$\begin{aligned}\|X_n^{-1}\|\mathbf{1}_{\mathcal{B}} &\leq \|(I+\sigma A_n^{-1}\eta_n)^{-1}\|\,\|A_n^{-1}\|\mathbf{1}_{\mathcal{B}}\\ &= \|I - \sigma A_n^{-1}\eta_n + \sigma^2(A_n^{-1}\eta_n)^2 - \ldots\|\,\|A_n^{-1}\|\mathbf{1}_{\mathcal{B}} \leq 2\|A_n^{-1}\|.\end{aligned}$$

This yields (ii). Now, remark that we can find a matrix R_n such that

$$X_n^{-1}A_n = (I+\sigma A_n^{-1}\eta_n)^{-1} = I - \sigma A_n^{-1}\eta_n + \sigma A_n^{-1}\eta_n R_n, \tag{6.4}$$

where $\|R_n\|^2 \mathbf{1}_{\mathcal{B}} \le 1/n$. Let $h \in H$. A direct computation provides:

$$\begin{aligned}\|X_n^{-1} A_n h\|_{(j)}^2 \mathbf{1}_{\mathcal{B}} &\le (1+B)\|h\|_{(j)}^2 + 2(1+B^{-1})\|\sigma A_n^{-1}\eta_n h\|_{(j)}^2 \mathbf{1}_{\mathcal{B}} \\ &\quad + 2(1+B^{-1})\|\sigma A_n^{-1}\eta_n R_n h\|_{(j)}^2 \mathbf{1}_{\mathcal{B}} \\ &= (1+B)\|h\|_{(j)}^2 + 2(1+B^{-1})(U_j + V_j),\end{aligned}$$

for all $B > 0$. We begin with the study of U_j:

$$\begin{aligned}\mathbb{E}_\theta U_j &\triangleq \sigma^2 \mathbb{E}_\theta \sum_{k\in I_j,\ k\le n} \left(\sum_{l=1}^n \sum_{p=1}^n b_{kl}\eta_{lp} h_p \right)^2 \mathbf{1}_{\mathcal{B}} \\ &\le \sigma^2 \mathbb{E}_\theta \sum_{k\in I_j,\ k\le n} \sum_{l=1}^n \sum_{p=1}^n h_p^2 b_{kl}^2 \max_{m=1,\dots,M_1} (s_k^m)^2 \ \mathbf{1}_{\mathcal{B}},\end{aligned}$$

where for all $m \in \{1,\dots,M_1\}$, the s_k^m are standard Gaussian random variables. Thus, $\max_{m=1,\dots,M_1}(s_k^m)^2 \le \log^{1+\tau} 1/\sigma$ with probability $1-\Omega$ and

$$U_j \le l_j(h)\frac{\sigma^2}{\varepsilon^2}\log^{1+\tau}\frac{1}{\sigma}\mathbb{E}_\theta \sigma_j^2 + c\frac{\sigma^2}{\varepsilon^2}\bar{\sigma}_j^2\Omega \le l_j(h)\frac{\sigma^2}{\varepsilon^2}\log^{1+\tau}\frac{1}{\sigma}\mathbb{E}_\theta\sigma_j^2 + \frac{c\varepsilon^2}{M_0},$$

where

$$\bar{\sigma}_j^2 = \max_{m=N\wedge M_0,\dots,N\wedge M_1} \sum_{k\in I_j}\sum_{l=1}^m \langle A_m^{-1}\psi_l, \phi_k\rangle^2. \tag{6.5}$$

Indeed, we use the definition of M_0 to show that $\bar{\sigma}_j^2\Omega \le c\varepsilon^2/M_0$ for some $c > 0$. In the same way:

$$\begin{aligned}\mathbb{E}_\theta V_j &\triangleq \mathbb{E}_\theta \|\sigma A_n^{-1}\eta_n R_n h\|_{(j)}^2 \mathbf{1}_{\mathcal{B}} \\ &\le \sigma^2 \mathbb{E}_\theta \sum_{k\in I_j}\sum_{p=1}^n \left(\sum_{l=1}^n b_{kl}\eta_{lp}\right)^2 \|R_n\|^2 l_j(h)\mathbf{1}_{\mathcal{B}} \le l_j(h)\frac{\sigma^2}{\varepsilon^2}c_\sigma \mathbb{E}_\theta\sigma_j^2 + \frac{c\varepsilon^2}{M_0},\end{aligned}$$

where $c_\sigma = \log^{1+\tau} 1/\sigma$. The proof of (iv) follows the same lines. Just remark that

$$\begin{aligned}\sum_{l=1}^n x_{kl}^2 &\le (1+B)\sum_{l=1}^n b_{kl}^2 + 2(1+B^{-1})\sigma^2 \sum_{l=1}^n \langle A_n^{-1}\eta_n A_n^{-1}\psi_l, \phi_k\rangle^2 \\ &\quad + 2(1+B^{-1})\sigma^2 \sum_{l=1}^n \langle A_n^{-1}\eta_n R_n A_n^{-1}\psi_l, \phi_k\rangle^2.\end{aligned} \tag{6.6}$$

Using the same principle, one obtains the result with a good choice of B. For the lower bound, use $A_n^{-1} = X_n^{-1}(I + \sigma A_n^{-1}\eta_n)$. This completes the proof. □

The following lemma provides an upper bound for the probability that $\lambda_{K_j}^\star > 0$ on the event C_j. The proof follows the same lines as in Cavalier and Tsybakov [7].

The main difficulty is to control the correlation between the variables $\tilde{y}_k$ in each block I_j.

Lemma 3. *Set $n_0 = M_0 \wedge N$ and $n_1 = M_1 \wedge N$ and let C_j be the event defined in* (3.1)*. Under the assumptions of Proposition* 1*, for all $j \in \{2, \dots, J\}$,*

$$P\big(\{\lambda^\star_{K_{j-1}} > 0\} \cap C_j\big) \leq T_j \exp\left[-\frac{C\varphi_j^2}{\Delta_j(1+2\sqrt{\varphi_j})^2}\right] + T_j(n_1 - n_0)\exp\left[-\frac{C\varphi_j^2\varepsilon^2/\sigma^2}{\Delta_j(1+2\sqrt{\varphi_j})^2}\right] + O(\Omega),$$

where C is independent of ε and σ. The second exponential term vanishes if $\sigma = 0$.

Proof. From the definitions of $\tilde{y}_k$ in (2.14) and $\lambda^\star_{K_j}$ in (2.15),

$$(6.7) \quad P\big(\{\lambda^\star_{K_{j-1}} > 0\} \cap C_j\big) = P\bigg(\Big\{\sum_{k\in I_j} \tilde{y}_k^2 \geq \tilde{\sigma}_j^2(1+\varphi_j)\Big\} \cap C_j\bigg)$$

$$= P\bigg(\bigg\{\varepsilon^2 \sum_{k\in I_j}\bigg(\sum_{l=1}^n x_{kl}\xi_l\bigg)^2 + 2\varepsilon\sum_{k\in I_j}\sum_{l=1}^n \langle X_n^{-1}A_n(P_n+G_n)f, \phi_k\rangle x_{kl}\xi_l + \sum_{k\in I_j}\langle X_n^{-1}A_n(P_n+G_n)f, \phi_k\rangle^2 \geq \tilde{\sigma}_j^2(1+\varphi_j)\bigg\} \cap C_j\bigg)$$

$$\leq p_1\bigg(\frac{(1-\mu)\tilde{\sigma}_j^2\varphi_j}{4\varepsilon^2}\bigg) + p_2\bigg(\frac{\mu\tilde{\sigma}_j^2\varphi_j}{8\varepsilon}\bigg) + P\bigg(\bigg\{\|X_n^{-1}A_n h_n\|^2_{(j)} > \frac{3\tilde{\sigma}_j^2\varphi_j}{4}\bigg\} \cap C_j\bigg),$$

where $h_n = (P_n+G_n)f$, $\mu = 2\sqrt{\varphi_j}/(1+2\sqrt{\varphi_j})$,

$$p_1(t) \triangleq P\bigg(\bigg\{\sum_{k\in I_j}\Big[\Big(\sum_{l=1}^n x_{kl}\xi_l\Big)^2 - \sum_{l=1}^n x_{kl}^2\Big] \geq t\bigg\} \cap C_j\bigg),$$

and

$$p_2(t) \triangleq P\bigg(\bigg\{\sum_{l=1}^n\Big(\sum_{k\in I_j}\langle X_n^{-1}A_n(P_n+G_n)f, \phi_k\rangle x_{kl}\Big)\xi_l \geq t\bigg\} \cap C_j\bigg),$$

for all $t \geq 0$. We begin with the evaluation of p_1. For all $k \in I_j$,

$$\bigg(\sum_{l=1}^n x_{kl}\xi_l\bigg)^2 = x_{kk}^2\xi_k^2 + \bigg(\sum_{\substack{m=1\\m\neq k}}^n x_{km}\xi_m\bigg)^2 + 2x_{kk}\xi_k\bigg(\sum_{\substack{m=1\\m\neq k}}^n x_{km}\xi_m\bigg) \triangleq A_1 + A_2 + A_3.$$

Thus, for all $t \geq 0$:

$$p_1(t) \leq P\big(A_1 \geq t/3\big) + P\big(A_2 \geq t/3\big) + P\big(A_3 \geq t/3\big).$$

Lemma 3 of Cavalier and Tsybakov [7] and Lemma 2 provide:

$$
\begin{aligned}
P\biggl(A_1 \geq \frac{(1-\mu)\tilde\sigma_j^2\varphi_j}{12\varepsilon^2}\biggr) &= P\biggl(\biggl\{\sum_{k\in I_j} x_{kk}^2(\xi_k^2-1) \geq \frac{(1-\mu)\tilde\sigma_j^2\varphi_j}{6\varepsilon^2}\biggr\}\cap C_j\biggr)\\
&\leq \mathbb{E}_\theta \exp\biggl[-\frac{(1-\mu)^2\tilde\sigma_j^4\varphi_j^2}{c\varepsilon^4(\sum_{k\in I_j} x_{kk}^4 + \frac{(1-\mu)\tilde\sigma_j^2\varphi_j}{2\varepsilon^2}\max_{k\in I_j} x_{kk}^2)}\biggr]\\
&\leq \exp\biggl[-\frac{C\varphi_j^2}{\Delta_j(1+2\sqrt{\varphi_j})}\biggr] + O(\Omega).
\end{aligned}
$$

Using (6.4), (6.6), and Assumption A2, for all $k\in I_j$,

$$
T_j\sum_{\substack{l=1\\ l\neq k}}^n x_{kl}^2\ \mathbf{1}_{C_j} \leq 2c\max_{k\in I_j}\sum_{l=1}^n b_{kl}^2 + \frac{c}{M_0}\sum_{l=1}^M b_{kl}^2 T_j \leq c\max_{k\in I_j}\sum_{l=1}^n b_{kl}^2, \tag{6.8}
$$

with probability $1-\Omega$. Therefore,

$$
\begin{aligned}
P\biggl(A_2 \geq \frac{\tilde\sigma_j^2\varphi_j}{12\varepsilon^2}\biggr) &= P\biggl(\biggl\{\sum_{k\in I_j}\biggl(\sum_{\substack{m=1\\ m\neq k}}^n x_{km}\xi_m\biggr)^2 - \sum_{\substack{m=1\\ m\neq k}}^n x_{km}^2 \geq \frac{\tilde\sigma_j^2\varphi_j(1-\mu)}{2\varepsilon^2}\biggr\}\cap C_j\biggr)\\
&\leq \sum_{k\in I_j} P\biggl(\biggl\{\sum_{\substack{m=1\\ m\neq k}}^n x_{km}^2(s_k^2-1) \geq \frac{\tilde\sigma_j^2\varphi_j(1-\mu)}{2\varepsilon^2 T_j}\biggr\}\cap C_j\biggr)\\
&\leq \mathbb{E}_\theta\sum_{k\in I_j}\exp\biggl[-\frac{(1-\mu)^2\tilde\sigma_j^4\varphi_j^2}{c\varepsilon^4 T_j^2(\sum_{m=1, m\neq k}^n x_{km}^2)^2}\biggr]\ \mathbf{1}(C_j)\\
&\leq T_j\exp\biggl[-\frac{C\varphi_j^2}{\Delta_j(1+2\sqrt{\varphi_j})}\biggr] + O(\Omega),
\end{aligned}
$$

where conditioned on η, the $(s_k)_{k\in I_j}$ are standard Gaussian random variables. The bound for $P(A_3\geq t)$ follows in the same way.

Using Lemma 3 of Cavalier and Tsybakov [7] and Assumption A2, one obtains the same bound for p_2. In particular, remark that

$$
\begin{aligned}
&\sum_{l=1}^n\biggl(\sum_{k\in I_j,\ k\leq n} x_{kl}\langle X_n^{-1}A_nh_n, \phi_k\rangle\biggr)^2\\
&\qquad \leq 2\biggl(\max_{k\in I_j,\ k\leq n} x_{kk}^2 + \sum_{k\in I_j,\ k\leq n}\sum_{\substack{l=1\\ l\neq k}}^n x_{kl}^2\biggr)\|X_n^{-1}A_nh_n\|_{(j)}^2.
\end{aligned}
$$

Then use (6.8) and Lemma 2.

We now bound the last probability in (6.7). In the SVD case, a simple calculation shows that it vanishes using the definition of C_j in (3.1). Otherwise, using Lemma 2,

$$\begin{aligned}
&P\bigg(\Big\{\|X_n^{-1}A_nh_n\|_{(j)}^2 \geq \frac{3\varphi_j\tilde{\sigma}_j^2}{4}\Big\}\cap C_j\bigg)\\
&\quad\leq \sum_{m=n_0}^{n_1} P\bigg(\Big\{\|X_m^{-1}A_mh_m\|_{(j)}^2 \geq \frac{3\varphi_j\tilde{\sigma}_j^2}{4}\Big\}\cap C_j\bigg) + O(\Omega)\\
&\quad\leq \sum_{m=n_0}^{n_1} P\bigg(2\|\sigma A_m^{-1}\eta_m h_m\|_{(j)}^2 + 2\|\sigma A_m^{-1}\eta_m R_m h_m\|_{(j)}^2 \geq \frac{\varphi_j\tilde{\sigma}_j^2}{4}\bigg) + O(\Omega)\\
&\quad\leq P_1 + P_2 + O(\Omega),
\end{aligned}$$

where $h_m = (P_m + G_m)f$ for all $m \in \mathbb{N}$. Then write as in Lemma 2:

$$P_1 \triangleq \sum_{m=n_0}^{n_1} P\bigg(\sigma^2 \sum_{k\in I_j}\bigg(\sum_{l=1}^{m}\sum_{p=1}^{m}\langle A_m^{-1}\psi_l,\phi_k\rangle\eta_{lp}h_p\bigg)^2 \geq \frac{\varphi_j\tilde{\sigma}_j^2}{16}\bigg).$$

The bound is obtained using the same methods as for p_1. By Assumption A3, with $c_3 = 16(1+8/\sqrt{M_0})$, we can find a constant $c \in]0,1[$ such that

$$\frac{\tilde{\sigma}_j^2\varphi_j}{16} - \sigma_j^2\frac{\sigma^2}{\varepsilon^2}\|(P_n+G_n)f\|^2 > \frac{\tilde{\sigma}_j^2\varphi_j c}{16},$$

with probability $1-\Omega$ on the event $\mathcal{B}$. The bound for P_2 follows exactly in the same way. $\square$

Lemma 4. *Let C_j be the event defined in* (3.1), *set $c_\sigma = \log^{1+\tau} 1/\sigma$ and $h_n = (P_n+G_n)f$. There exists a positive constant c such that, for all $j \in \{1,\dots,J\}$ and $B > 0$,*

$$\begin{aligned}
&\mathbb{E}_\theta\bigg[\frac{\tilde{\sigma}_j^2\|h_n\|_{(j)}^2}{\tilde{\sigma}_j^2 + \|X_n^{-1}A_nh_n\|_{(j)}^2}\bigg]\mathbf{1}(\bar{C}_j)\\
&\qquad\leq \bigg(1 + cB + (1+B^{-1})\frac{c_\sigma}{\varphi_j}l_j(h_n)\frac{\sigma^2}{\varepsilon^2}\bigg)\mathbb{E}_\theta\bigg[\frac{\tilde{\sigma}_j^2\|h_n\|_{(j)}^2}{\tilde{\sigma}_j^2 + \|h_n\|_{(j)}^2}\bigg]\mathbf{1}(\bar{C}_j) + c\bar{\sigma}_j^2\Omega,
\end{aligned}$$

where $l_j(h_n)$ is defined in (6.2) *and $\bar{\sigma}_j^2$ in* (6.5).

Proof. Let $j \in \{1,\dots,J\}$ be fixed. Define

$$\bar{R}_{\varepsilon,\sigma}^j(\lambda,\theta) = \sum_{k\in I_j}\bigg[(1-\lambda_k)^2(h_n^k)^2 + \varepsilon^2\sum_{l=1}^{n}x_{kl}^2\lambda_k^2\bigg]\mathbf{1}(\bar{C}_j), \tag{6.9}$$

$$\hat{R}_{\varepsilon,\sigma}^j(\lambda,\theta) = \sum_{k\in I_j}\bigg[\big(\lambda_k\langle X_n^{-1}h_n,\phi_k\rangle - h_n^k\big)^2 + \varepsilon^2\sum_{l=1}^{n}x_{kl}^2\lambda_k^2\bigg]\mathbf{1}(\bar{C}_j), \tag{6.10}$$

where $h_n^k = \langle h_n, \phi_k \rangle$ for all $k \in \{1, \dots, n\}$. These two quantities correspond to different approximations of the mean squared risk restricted to the block j. Remark that

$$\mathbb{E}_\theta[\inf_{\lambda \in \Lambda^*} \bar{R}^j_{\varepsilon,\sigma}(\lambda, \theta)] = \mathbb{E}_\theta\left[\frac{\tilde{\sigma}_j^2 \|h_n\|^2_{(j)}}{\tilde{\sigma}_j^2 + \|h_n\|^2_{(j)}}\right]\mathbf{1}(\bar{C}_j), \tag{6.11}$$

and

$$\mathbb{E}_\theta[\inf_{\lambda \in \Lambda^*} \hat{R}^j_{\varepsilon,\sigma}(\lambda, \theta)] = \mathbb{E}_\theta\left[\frac{\tilde{\sigma}_j^2 \|h_n\|^2_{(j)}}{\tilde{\sigma}_j^2 + \|X_n^{-1} A_n h_n\|^2_{(j)}}\right]\mathbf{1}(\bar{C}_j). \tag{6.12}$$

For all $\lambda \in \Lambda^*$ and $B > 0$, using the elementary inequality $2ab \le B^{-1}a^2 + Bb^2$ and Lemma 2, we have

$$\begin{aligned} &\left|\mathbb{E}_\theta \bar{R}^j_{\varepsilon,\sigma}(\lambda, \theta) - \mathbb{E}_\theta \hat{R}^j_{\varepsilon,\sigma}(\lambda, \theta)\right| \\ &\le \mathbb{E}_\theta\left[\sum_{k \in I_j} \left\{B(1-\lambda_k)^2 (h_n^k)^2 + (1+B^{-1})\lambda_k^2 \langle (X_n^{-1}A_n - I)h_n, \phi_k\rangle^2\right\}\right]\mathbf{1}(\bar{C}_j) \\ &\le cB\mathbb{E}_\theta \bar{R}^j_{\varepsilon,\sigma}(\lambda, \theta) + c(1+B^{-1})c_\sigma \frac{\sigma^2}{\varepsilon^2} l_j(h_n)\mathbb{E}_\theta \tilde{\sigma}_j^2\ \mathbf{1}(\bar{C}_j) + \bar{\sigma}_j^2 O(\Omega), \end{aligned} \tag{6.13}$$

where $c > 0$. Then remark that

$$\mathbb{E}_\theta \tilde{\sigma}_j^2\ \mathbf{1}(\bar{C}_j) \le c\mathbb{E}_\theta \frac{\tilde{\sigma}_j^2 \|(P_n + G_n)f\|^2_{(j)}}{\tilde{\sigma}_j^2 + \|(P_n + G_n)f\|^2_{(j)}}\left(1 + \frac{8}{\varphi_j}\right)\mathbf{1}(\bar{C}_j), \tag{6.14}$$

and apply inequality (6.13) to $\hat{\lambda}_j = \operatorname{argmin}_{\lambda \in \Lambda^*} \hat{R}^j_{\varepsilon,\sigma}(\lambda, \theta)$ to conclude the proof. □

The following lemma provides an upper bound for the two residual terms appearing in (7.3).

Lemma 5. *Let $\bar{C}_j$ be the event defined in* (3.1) *and set $c_\sigma = \log^{1+\tau} 1/\sigma$. There exists a positive constant c such that*

$$\text{(i)} \qquad \mathbb{E}_\theta \|(X_n^{-1}A_n - I)P_n f\|^2_{(j)} \mathbf{1}(\bar{C}_j) \le cc_\sigma l_j(\theta)\frac{\sigma^2}{\varepsilon^2}\mathbb{E}_\theta \tilde{\sigma}_j^2\ \mathbf{1}(\bar{C}_j) + \bar{\sigma}_j^2 O(\Omega),$$

$$\begin{aligned} \text{(ii)} \qquad &\mathbb{E}_\theta \sum_{k \in I_j,\ k \le n} \langle (X_n^{-1}A_n - I)f, \phi_k\rangle \tilde{y}_k (1 - \bar{\lambda}_j)\,\mathbf{1}(\bar{C}_j) \\ &\qquad \le c\Big(1 + c_\sigma \frac{\sigma^2}{\varepsilon^2} l_j(\theta)\Big)\mathbb{E}_\theta \tilde{\sigma}_j^2\ \mathbf{1}(\bar{C}_j) + \bar{\sigma}_j^2 O(\Omega). \end{aligned}$$

Proof. The proof uses the same techniques as in Lemmas 2–4. □

7. Proof of Proposition 1

Remark that

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 = \sum_{j=1}^J \mathbb{E}_\theta \|\theta^\star - \theta\|^2_{(j)} + \sum_{k > N} \theta_k^2.$$

For all $j \in \{1, \dots, J\}$, let C_j be the event defined in (3.1) and set

$$\bar{C}_j = \left\{ \|(P_n + G_n)f\|^2_{(j)} \geq \varphi_j \frac{\tilde{\sigma}_j^2}{8} \right\} \cap \mathcal{B} \cap \mathcal{M}. \tag{7.1}$$

The events $\mathcal{B}$ and $\mathcal{M}$ are defined in Lemma 2 (see Section 6). First we bound the risk separately on C_j and $\bar{C}_j$.

7.1. Bound of the risk on $\bar{C}_j$. Let $j \in \{1, \dots, J\}$ be fixed. First we assume that the penalty satisfies

$$\Delta_j \leq \frac{1 - \varphi_j}{4} \frac{1}{(1 + 8/\sqrt{M_0})^2}, \tag{7.2}$$

and set

$$\bar{\theta}_k = \tilde{y}_k \left(1 - \frac{\tilde{\sigma}_j^2 (1 + \varphi_j)}{\|\tilde{y}\|^2_{(j)}} \right) \triangleq \tilde{y}_k \bar{\lambda}_k(\tilde{y}).$$

Using the decomposition of Stein [25], we have

$$\begin{aligned} \mathbb{E}_\theta \|\theta^\star - \theta\|^2_{(j)} &= \mathbb{E}_\theta \|\theta^\star - X_n^{-1} A_n (P_n + G_n) f\|^2_{(j)} \\ &\quad + \mathbb{E}_\theta \|(X_n^{-1} A_n - I) f + X_n^{-1} A_n G_n f\|^2_{(j)} \\ &\quad + 2\mathbb{E}_\theta \sum_{k \in I_j,\ k \leq n} \big[\langle (X_n^{-1} A_n - I) f + X_n^{-1} A_n G_n f, \phi_k \rangle \tilde{y}_k (1 - \lambda_k^\star) \big]. \end{aligned} \tag{7.3}$$

Set $h_n = (P_n + G_n) f$. Applying Lemma 5 of Cavalier and Tsybakov [7], we get

$$\begin{aligned} \mathbb{E}_\theta \|\theta^\star - X_n^{-1} A_n h_n\|^2_{(j)} \ \mathbf{1}(\bar{C}_j) &\leq \mathbb{E}_\theta \|\bar{\theta} - X_n^{-1} A_n h_n\|^2_{(j)} \ \mathbf{1}(\bar{C}_j) \\ &= \sum_{k \in I_j} \mathbb{E}_\theta \Big\{ \big(\tilde{y}_k - \langle X_n^{-1} A_n h_n, \phi_k \rangle \big)^2 + \big(\tilde{y}_k (\bar{\lambda}_k - 1) \big)^2 \\ &\qquad - 2 \big[\big(\tilde{y}_k - \langle X_n^{-1} A_n h_n, \phi_k \rangle \big) \tilde{y}_k (1 - \bar{\lambda}_k) \big] \Big\} \mathbf{1}_{\{k \leq n\}} \ \mathbf{1}(\bar{C}_j). \end{aligned} \tag{7.4}$$

Using (2.14)

$$\mathbb{E}_\theta \big(\tilde{y}_k - \langle X_n^{-1} A_n h_n, \phi_k \rangle \big)^2 \ \mathbf{1}_{\{k \leq n\}} \ \mathbf{1}(\bar{C}_j) = \varepsilon^2 \mathbb{E}_\theta \sum_{l=1}^n x_{kl}^2 \ \mathbf{1}_{\{k \leq n\}} \ \mathbf{1}(\bar{C}_j),$$

and applying Lemma 1 of Stein [25], we obtain

$$\begin{aligned} S_j = \mathbb{E}_\theta \|\bar{\theta} - X_n^{-1} A_n h_n\|^2_{(j)} \ \mathbf{1}(\bar{C}_j) &= \mathbb{E}_\theta [\tilde{\sigma}_j^2] \ \mathbf{1}(\bar{C}_j) \\ &+ \mathbb{E}_\theta \sum_{k \in I_j \ k \leq n} \left\{ [\tilde{y}_k (\bar{\lambda}_k - 1)]^2 - \varepsilon^2 \sum_{l=1}^n x_{kl}^2 \left(1 - \bar{\lambda}_j - \tilde{y}_k \frac{\partial \bar{\lambda}_k}{\partial \tilde{y}_k} (\tilde{y}) \right) \right\} \mathbf{1}(\bar{C}_j). \end{aligned}$$

After some algebra, using (7.2), Jensen's inequality, and (iv) of Lemma 2, we get

$$
\begin{aligned}
S_j &\le \mathbb{E}_\theta \tilde\sigma_j^2 \ \mathbf{1}(\bar C_j) - (1-\varphi_j^2)\mathbb{E}_\theta\left(\frac{\tilde\sigma_j^4}{\|\tilde y\|^2_{(j)}}\right)\mathbf{1}(\bar C_j) \\
&\quad + 4\varepsilon^2 \mathbb{E}_\theta\left(\max_{k\in I_j,\ k\le n}\sum_{l=1}^n x_{kl}^2 \frac{(1+\varphi_j)\tilde\sigma_j^2}{\|\tilde y\|^2_{(j)}}\right)\mathbf{1}(\bar C_j) \\
&\le \mathbb{E}_\theta\left[\tilde\sigma_j^2 - \frac{\tilde\sigma_j^4}{\|\tilde y\|^2_{(j)}}\big(1-\varphi_j^2 - 4(1+8/\sqrt{M_0})^2\Delta_j(1+\varphi_j)\big)\right]\mathbf{1}(\bar C_j) + c\bar\sigma_j^2\Omega.
\end{aligned}
$$

Conditioning on X_n, $\mathbb{E}_\theta[\|\tilde y\|^2_{(j)} \mid X_n] = \tilde\sigma_j^2 + \|X_n^{-1}A_n h_n\|^2_{(j)}$. We can apply Jensen's inequality since (7.2) holds:

$$
\begin{aligned}
S_j &\le \mathbb{E}_\theta\left[\tilde\sigma_j^2 - \frac{\tilde\sigma_j^4(1-\varphi_j^2 - 4(1+8/\sqrt{M_0})^2\Delta_j(1+\varphi_j))}{\mathbb{E}_\theta[\|\tilde y\|^2_{(j)}/X_n]}\right]\mathbf{1}(\bar C_j) + c\bar\sigma_j^2\Omega \\
&\le \mathbb{E}_\theta\left(c\frac{(\varphi_j^2+4\Delta_j)}{\varphi_j} + \frac{\|X_n^{-1}A_n h_n\|^2_{(j)}}{\|h_n\|^2_{(j)}}\right)\frac{\tilde\sigma_j^2\|h_n\|^2_{(j)}}{\tilde\sigma_j^2 + \|X_n^{-1}A_n h_n\|^2_{(j)}}\mathbf{1}(\bar C_j).
\end{aligned}
$$

Using (7.3), Lemmas 4 and 5, and the same techniques as in Lemma 2, we eventually obtain:

$$
\begin{aligned}
&(7.5)\quad \mathbb{E}_\theta\|\theta^\star - \theta\|^2_{(j)}\mathbf{1}(\bar C_j) \\
&\quad \le \left(1 + c\frac{(\varphi_j^2+4\Delta_j)}{\varphi_j}\right)\left(1 + B + B^{-1}\frac{l_j(h_n)}{\varphi_j}\frac{\sigma^2}{\varepsilon^2}c_\sigma\right)^2 \mathbb{E}_\theta\frac{\tilde\sigma_j^2\|h_n\|^2_{(j)}}{\tilde\sigma_j^2 + \|h_n\|^2_{(j)}}\mathbf{1}(\bar C_j) \\
&\qquad + c\mathbb{E}_\theta\|X_n^{-1}A_n G_n f\|^2_{(j)}\mathbf{1}_{\mathcal{B}} + \bar\sigma_j^2 O(\Omega),
\end{aligned}
$$

where $c_\sigma = \log^{1+\tau} 1/\sigma$ and c is a positive constant.

Now consider the case, where inequality (7.2) is not satisfied. The penalty φ_j is too large. Using (2.15) and Lemmas 2 and 5,

$$
\mathbb{E}_\theta\|\theta^\star - \theta\|^2_{(j)}\mathbf{1}(\bar C_j) \le 2\mathbb{E}_\theta\left[\sum_{k\in I_j}(\tilde y_k - \theta_k)^2 + \sum_{k\in I_j}\tilde y_k^2(1-\lambda_k^\star)^2\right]\mathbf{1}(\bar C_j) \tag{7.6}
$$

$$
\begin{aligned}
&\le c\Big(1 + c_\sigma l_j(\theta)\frac{\sigma^2}{\varepsilon^2} + 1 + \varphi_j\Big)\mathbb{E}_\theta\frac{\tilde\sigma_j^2\|h_n\|^2_{(j)}}{\tilde\sigma_j^2 + \|h_n\|^2_{(j)}}\mathbf{1}(\bar C_j) \\
&\quad + \bar\sigma_j^2\Omega + c\mathbb{E}_\theta\|X_n^{-1}A_n G_n f\|^2_{(j)}\mathbf{1}_{\mathcal{B}}.
\end{aligned} \tag{7.7}
$$

Sharp results for $\theta^\star$ will be obtained only if inequality (7.2) is satisfied. According to the structure of the representation matrix, it will not always be the case. Indeed, inequality (7.2) holds automatically if the penalty is small enough. But we require also that Assumption A3 holds. The choice of φ_j is therefore a trade off between a good quality of recovering and a control of the noise in the operator.

7.2. Bound of the risk on C_j. Define

$$A_j = \left\{\|\tilde{y}\|^2_{(j)} \leq \tilde{\sigma}^2_j(1+\varphi_j)\right\} = \{\lambda^\star_{K_{j-1}} = 0\}.$$

On $\bar{A}_j \cap C_j$, the function $\lambda^\star_k(\tilde{y})$ is positive but Lemma 3 provides that $P(\bar{A}_j \cap C_j)$ is small. Thus $\theta^\star$ is equal to zero on C_j with large probability. Let $B \in]0,1[$. One can easily show that,

$$\|\theta\|^2_{(j)} - \|(P_n+G_n)f\|^2_{(j)} \leq (1+B^{-1})\|G_n f\|^2_{(j)} + B\|\theta\|^2_{(j)}.$$

Therefore

$$\begin{aligned}(7.8)\quad & \mathbb{E}_\theta\|\theta^\star - \theta\|^2_{(j)}\mathbf{1}(C_j) \\ &= \mathbb{E}_\theta\|\theta\|^2_{(j)}\mathbf{1}(C_j) + \mathbb{E}_\theta \sum_{k\in I_j,\ k\leq n}\left[(\lambda^\star_k\tilde{y}_k)^2 - 2\lambda^\star_k\tilde{y}_k\theta_k\right]\mathbf{1}(C_j\cap\bar{A}_j) \\ &\leq \frac{1}{1-B}\mathbb{E}_\theta\|(P_n+G_n)f\|^2_{(j)}\mathbf{1}(C_j) + \frac{1+B^{-1}}{1-B}\mathbb{E}_\theta\|G_n f\|^2_{(j)} \\ &\quad + R_1 - 2R_2.\end{aligned}$$

With the Cauchy–Schwarz and Young inequalities:

$$(7.9)\qquad |R_2| \triangleq \left|\mathbb{E}_\theta \sum_{k\in I_j,\ k\leq n} \lambda^\star_k\tilde{y}_k\theta_k\ \mathbf{1}(C_j\cap\bar{A}_j)\right| \leq \frac{1}{2}R_1 + \frac{r^2}{2}P(C_j\cap\bar{A}_j),$$

since $\|\theta\| \leq r$. Using (2.14) and the fact that $0 \leq \lambda^\star_k \leq 1$, we get

$$\begin{aligned}(7.10)\ R_1 &\triangleq \mathbb{E}_\theta \sum_{k\in I_j,\ k\leq n} (\lambda^\star_k\tilde{y}_k)^2\ \mathbf{1}(C_j\cap\bar{A}_j) \\ &\leq 2\mathbb{E}_\theta \sum_{k\in I_j}\left[\langle X_n^{-1}A_n(P_n+G_n)f,\phi_k\rangle^2 + \varepsilon^2\left(\sum_{l=1}^n x_{kl}\xi_l\right)^2\right]\mathbf{1}(C_j\cap\bar{A}_j) \\ &\leq 4r^2P(C_j\cap\bar{A}_j) + c\bar{\sigma}^2_j P(C_j\cap\bar{A}_j)^{1/2} + c\mathbb{E}_\theta\|X_n^{-1}A_nG_nf\|^2_{(j)}\mathbf{1}_{\mathcal{B}},\end{aligned}$$

where $\bar{\sigma}^2_j$ is defined in (6.5). Indeed,

$$\|X_n^{-1}A_n(P_n+G_n)f\|^2\mathbf{1}(C_j\cap\bar{A}_j) \leq 2r^2 + 2\|X_n^{-1}A_nG_nf\|^2_{(j)}\mathbf{1}_{\mathcal{B}}.$$

Moreover,

$$\begin{aligned}(7.11)\qquad \varepsilon^2\mathbb{E}_\theta\left(\sum_{l=1}^n x_{kl}\xi_l\right)^2\mathbf{1}(C_j\cap\bar{A}_j) &= \varepsilon^2\mathbb{E}_\theta\sum_{l=1}^n x^2_{kl}(s^n_k)^2\ \mathbf{1}(C_j\cap\bar{A}_j) \\ &\leq c\bar{\sigma}^2_j P(C_j\cap\bar{A}_j)^{1/2} + \bar{\sigma}^2_j O(\Omega),\end{aligned}$$

where $s_k^n \sim \mathcal{N}(0,1)$ conditioned on η and c is a positive constant. On the other hand, remark that

$$\mathbf{1}(C_j) \leq \Big(1 + \frac{\varphi_j}{8}\Big) \frac{\tilde{\sigma}_j^2}{\tilde{\sigma}_j^2 + \|(P_n + G_n)f\|_{(j)}^2} \mathbf{1}(C_j). \tag{7.12}$$

Using (7.8)–(7.10) and (7.12), we eventually find

$$\begin{aligned} &\mathbb{E}_\theta \|\theta^\star - \theta\|_{(j)}^2 \mathbf{1}(C_j) \\ &\quad \leq \Big(1 + \frac{\varphi_j}{8}\Big) \frac{1}{1-B} \mathbb{E}_\theta \left[\frac{\tilde{\sigma}_j^2 \|(P_M + G_M)f\|_{(j)}^2}{\tilde{\sigma}_j^2 + \|(P_n + G_n)f\|_{(j)}^2} \right] \mathbf{1}(C_j) \\ &\qquad + c\frac{1 + B^{-1}}{1-B} \mathbb{E}_\theta \|X_n^{-1} A_n G_n f\|_{(j)}^2 \mathbf{1}_{\mathcal{B}} + c\bar{\sigma}_j^2 P(C_j \cap \bar{A}_j)^{1/2} + c\bar{\sigma}_j^2 \Omega, \end{aligned} \tag{7.13}$$

for some $c > 0$.

7.3. Final bound of the risk. Let $j \in \{1, \dots, J\}$. Using (iv) of Lemma 2 and the same methods as in Lemma 4, we find:

$$\begin{aligned} \mathbb{E}_\theta \frac{\tilde{\sigma}_j^2 \|(P_n + G_n)f\|_{(j)}^2}{\tilde{\sigma}_j^2 + \|(P_n + G_n)f\|_{(j)}^2} \mathbf{1}_{\mathcal{B}\cap\mathcal{M}} &\leq \Big(1 + \frac{c}{\sqrt{M_0}}\Big) \mathbb{E}_\theta \frac{\sigma_j^2 \|(P_n + G_n)f\|_{(j)}^2}{\sigma_j^2 + \|(P_n + G_n)f\|_{(j)}^2} + c\Omega \\ &= (1 + c/\sqrt{M_0}) \inf_{\lambda\in\Lambda^\star} \bar{R}_\varepsilon^n(\lambda,\theta) + O(\Omega), \end{aligned}$$

where $\sigma_j^2 = \varepsilon^2 \sum_{k\in I_j} \sum_{l=1}^n b_{kl}^2$, and

$$\bar{R}_\varepsilon^n(\lambda,\theta) = \sum_{k=1}^n (1-\lambda_k)^2 \langle (P_n + G_n)f, \phi_k\rangle^2 + \varepsilon^2 \sum_{k=1}^n \sum_{l=1}^n \lambda_k^2 b_{kl}^2 + \sum_{k>n} \theta_k^2.$$

Using (7.5), (7.7), (7.13), and summing up over j, we obtain

$$\begin{aligned} \mathbb{E}_\theta \|\theta^\star - \theta\|^2 \mathbf{1}_{\mathcal{B}\cap\mathcal{M}} &\leq \max_{j=1,\dots,J} t_j \mathbb{E}_\theta \inf_{\lambda\in\Lambda^*} \bar{R}_\varepsilon^n(\lambda,\theta) + c\varepsilon^2 \\ &\quad + c\sum_{j=1}^{n-1} \sigma_j^2 P(C_j \cap \bar{A}_j)^{1/2} + \frac{1+B^{-1}}{1-B} \mathbb{E}_\theta \|G_n f\|^2 + O(\Omega). \end{aligned} \tag{7.14}$$

Indeed, the definition of M_0 provides $\bar{\sigma}_j^2 \Omega \leq c\varepsilon^2/M_0$. For all $j \in \{1, \dots, J\}$,

$$t_j \leq \left(c_0 + \frac{c(\varphi_j^2 + 4\Delta_j)}{\varphi_j}\right)\left(1 + cB + c(1 + B^{-1})\frac{c_\sigma}{\varphi_j} l_j(\theta) \frac{\sigma^2}{\varepsilon^2}\right)^2 \left(1 + \frac{c}{\sqrt{M_0}}\right), \tag{7.15}$$

where $c_\sigma = \log^{1+\tau} 1/\sigma$, $l_j(\theta)$ is defined in (6.2), $c_0 = 1$ if inequality (7.2) is satisfied and $c_0 \geq 1$ else. In this case, this constant can be explicitly computed.

To finish the proof, remark that, using (7.6),

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \mathbf{1}_{\mathcal{B}^c \cup \mathcal{M}^c} = O(\Omega).$$

Following the proof of Lemma 1 of Cavalier and Tsybakov [7], we get

$$\mathbb{E}_\theta[\inf_{\lambda\in\Lambda^*} \bar{R}^n_\varepsilon(\lambda,\theta)] \leq (1+\eta_\varepsilon)\mathbb{E}_\theta[\inf_{\lambda\in\Lambda_{\mathrm{mon}}} \bar{R}^n_\varepsilon(\lambda,\theta)] + c\varepsilon^2\nu_\varepsilon^{2\beta+1} + O(\Omega), \tag{7.16}$$

where $\eta_\varepsilon \to 0$ as $\varepsilon \to 0$ by Assumption A1. In view of Lemma 3 and with the choice of geometrically increasing blocks, we can find a positive constant c such that

$$\sum_{j=2}^{J} \bar{\sigma}_j^2 P(C_j \cap \bar{A}_j)^{1/2} \leq c\varepsilon^2. \tag{7.17}$$

The bound is obtained in the same way as in Cavalier and Tsybakov [7]. Now, remark that for all θ and λ satisfying the conditions of the proposition and $B>0$:

$$\begin{aligned}
&\mathbb{E}_\theta[\bar{R}^n_\varepsilon(\lambda,\theta) - R^n_\varepsilon(\lambda,\theta)] \\
&\quad= \mathbb{E}_\theta\Big[\sum_{k=1}^n (1-\lambda_k)^2\langle G_n f,\phi_k\rangle^2 + 2\sum_{k=1}^n (1-\lambda_k)^2\theta_k\langle G_n f,\phi_k\rangle\Big] \\
&\quad\leq (1+B^{-1})\mathbb{E}_\theta\|G_n f\|^2 + B\mathbb{E}_\theta\sum_{k=1}^M (1-\lambda_k)^2\theta_k^2 \\
&\quad\Rightarrow \mathbb{E}_\theta[\inf_{\lambda\in\Lambda_{\mathrm{mon}}} \bar{R}^n_\varepsilon(\lambda,\theta)] \leq (1+B)\mathbb{E}_\theta[\inf_{\lambda\in\Lambda_{\mathrm{mon}}} R^n_\varepsilon(\lambda,\theta)] + (1+B^{-1})\mathbb{E}_\theta\|G_n f\|^2.
\end{aligned}$$

With (7.14), (7.16), and (7.17), we eventually obtain:

$$\begin{aligned}
\mathbb{E}_\theta\|\theta^\star - \theta\|^2 &\leq \max_{j=1,\dots,J}(1+g_j)\mathbb{E}_\theta[\inf_{\lambda\in\Lambda_{\mathrm{mon}}} R^n_\varepsilon(\lambda,\theta)] \\
&\quad + c\bigg(1+\frac{\sigma^2}{\varepsilon^2}\log^{1+\tau}\frac{1}{\sigma}\bigg)\varepsilon^2\nu_\varepsilon^{2\beta+1} + c\frac{1+B^{-1}}{1-B}\mathbb{E}_\theta\|G_n f\|^2 + O(\Omega),
\end{aligned} \tag{7.18}$$

where $g_j = t_j(1+\eta_\varepsilon)\big((1+B)/(1-B)\big) - 1$, t_j is defined in (7.15), c is a positive constant, and $\eta_\varepsilon \to 0$ as $\varepsilon \to 0$ by Assumption A1. Hence, one can easily show that

$$\mathbb{E}_\theta[\inf_{\lambda\in\Lambda_{\mathrm{mon}}} R^n_\varepsilon(\lambda,\theta)] \leq \inf_{\lambda\in\Lambda_{\mathrm{mon}}} R^N_\varepsilon(\lambda,\theta) + \Gamma(\theta),$$

where $\Gamma(\theta)$ is defined in (3.4). This concludes the proof. □

Acknowledgment. I would like to thank Laurent Cavalier for having introduced me to inverse problems and for his helpful and constructive remarks.

References

[1] J. Baumeister, *Stable Solution of Inverse Problems*, Advanced Lectures in Math., Friedr. Vieweg and Sohn, Braunschweig, 1987.

[2] E. N. Belitser and B. Ya. Levit, *On minimax filtering on ellipsoids*, Math. Methods Statist., **4** (1995), 259–273.

[3] A. Böttcher, *Infinite Matrices and Projection Methods*, Lectures on Operator Theory and its Applications, AMS, 1996.

[4] L. Cavalier, *Estimation in a problem of fractional integration*, Inverse Problems, **20** (2004), 1445–1454.
[5] L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov, *Oracle inequalities for inverse problems*, Ann. Statist., **30** (2002), 843–874.
[6] L. Cavalier and N. W. Hengartner, *Adaptive estimation for inverse problems with noisy operators*, Inverse Problems, **21** (2005), 1345–1361.
[7] L. Cavalier and A. B. Tsybakov, *Sharp adaptation for inverse problems with random noise*, Probab. Theory Rel. Fields, **123** (2002), 323–354.
[8] L. Cavalier and A. B. Tsybakov, *Penalized blockwise Stein's method, monotone oracles and sharp adaptive estimation*, Math. Methods Statist., **3** (2001), 247–282.
[9] A. Cohen, M. Hoffmann, and M. Reiss, *Adaptive wavelet Galerkin methods for linear inverse problems*, SIAM J. Numer. Anal., **42** (2004), 1479–1501.
[10] D. L. Donoho, *Nonlinear solutions of linear inverse problems by wavelet-vaguelette decomposition*, J. Appl. Comput. Harmonic Analysis, **2** (1995), 101–126.
[11] S. Efromovich, *Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise*, IEEE Trans. Inform. Theory, **43** (1997), 1184–1191.
[12] S. Efromovich and V. Koltchinskii, *On inverse problems with unknown operators*, IEEE Trans. Inform. Theory, **47** (2001), 2876–2893.
[13] H. W. Engl, M. Hank, and A. Neubauer, *Regularization of Inverse Problems*, Kluwer, Dordrecht, 1996.
[14] M. S. Ermakov, *Minimax estimation of the solution of an ill-posed convolution type problem*, Problems of Inform. Transmission, **25** (1989), 191–200.
[15] J. Fan, *On the optimal rates of convergence for nonparametric deconvolutions problems*, Ann. Statist., **19** (1991), 1257–1272.
[16] G. K. Golubev and R. Z. Khasminskii, *Statistical approach to some inverse boundary problems for partial differential equations*, Problems of Inform. Transmission, **35** (1999), 51–66.
[17] W. Hackbusch, *Integral Equations. Theory and Numerical Treatment*, ISNM 120, Birkhäuser, Basel, 1995.
[18] T. Hida, *Brownian Motion*, Springer, New York, 1980.
[19] M. Hoffmann and M. Reiss, *Nonlinear estimation for linear inverse problems with error in the operator*, Preprint, 2005.
[20] I. M. Johnstone and B. W. Silverman, *Speed of estimation in positron emission tomography and related inverse problems*, Ann. Statist., **18** (1990), 251–280.
[21] R. Kress, *Linear Integral Equations*, Applied Math. Sci., Vol 82, Springer, Berlin, 1989.
[22] B. Mair and F. H. Ruymgaart, *Statistical inverse estimation in Hilbert scales*, SIAM J. Appl. Math., **56** (1996), 1424–1444.
[23] P. Mathé and S. V. Pereverzev, *Optimal discretization of inverse problems in Hilbert scales. Regularization and self-regularization of projection methods*, SIAM J. Numer. Anal., **38** (2001), 1333–2021.
[24] A. Nemirovski, *Topics in Non-Parametric Statistics*, Lectures on Probab. Theory and Statist., Ecole d'été de Probab. de St.Flour XXVIII, 1998.
[25] C. M. Stein, *Estimation of the mean of a multivariate normal distribution*, Ann. Statist., **9** (1981), 1135–1151.
[26] A. B. Tsybakov, *On the best rate of adaptive estimation in some inverse problems*, C. R. Acad. Sci. Paris, Ser. 1, **330** (2000), 835–840.
[27] A. Zygmund, *Trigonometric Series*, Vols. I,II, 2nd ed., Cambridge Univ. Press, New York, 1959.

[*Received March* 2006; *revised November* 2006]