Journal of Nonparametric Statistics
Vol. 22, No. 6, August 2010, 685–702

Taylor & Francis
Taylor & Francis Group

The Stein hull

Clément Marteau*

Institut de Mathématiques, Université de Toulouse, INSA - 135, Avenue de Rangueil, F-31 077 Toulouse Cedex 4, France

(*Received 17 October 2008; final version received 2 October 2009*)

We are interested in the statistical linear inverse problem $Y = Af + \epsilon\xi$, where A denotes a compact operator and $\epsilon\xi$ a stochastic noise. In this setting, the risk hull point of view provides interesting tools for the construction of adaptive estimators. It sheds light on the processes governing the behaviour of linear estimators. In this article, we investigate the link between some threshold estimators and this risk hull point of view. The penalised blockwise Stein rule plays a central role in this study. In particular, this estimator may be considered as a risk hull minimisation method, provided the penalty is well chosen. Using this perspective, we study the properties of the threshold and propose an admissible range for the penalty leading to accurate results. We eventually propose a penalty close to the lower bound of this range.

Keywords: inverse problem; oracle inequality; risk hull; penalised blockwise Stein rule

Mathematical Subject classification (2000): 62G05; 62G20

1. Introduction

This article deals with the statistical inverse problem

$$Y = Af + \epsilon\xi, \tag{1}$$

where H, K are Hilbert spaces and $A : H \to K$ denotes a linear operator. The function $f \in H$ is unknown and has to be recovered from a measurement of Af corrupted by some stochastic noise $\epsilon\xi$. Here, ϵ represents a positive noise level and ξ a Gaussian white noise (see, Hida 1980 for more details). In particular, for all $g \in K$, we can observe

$$\langle Y, g\rangle = \langle Af, g\rangle + \epsilon\langle \xi, g\rangle, \tag{2}$$

where $\langle \xi, g\rangle \sim \mathcal{N}(0, \|g\|^2)$. Denote by $A^\star$ the adjoint operator of A. In the sequel, A is supposed to be a compact operator. Such a restriction is very interesting from a mathematical point of view. The operator $(A^\star A)^{-1}$ is unbounded: the least square solution $\hat{f}_{\mathrm{LS}} = (A^\star A)^{-1}A^\star Y$ does not continuously depend on Y. The problem is said to be ill-posed.

*Email: clement.marteau@insa-toulouse.fr

ISSN 1048-5252 print/ISSN 1029-0311 online

DOI: 10.1080/10485250903388878
http://www.informaworld.com

In a statistical context, several studies of ill-posed inverse problems have been proposed in recent years. It would be however impossible to cite them all. For the interested reader, we may mention Fan (1991) and Ermakov (1989) for convolution operators, Johnstone and Silverman (1990) for the positron emission tomography problem, Donoho (1995) in a wavelet setting, or Bissantz, Hohage, Munk and Ryumgaart (2007) for a general statistical approach and some rates of convergence. We also refer to Engl, Hank and Neubauer (1996) for a survey in a numerical setting.

Using a specific representation (i.e. particular choices for g in Equation (2)) may help in the understanding of the model (1). In this sense, the classical singular value decomposition is a very useful tool. Since $A^\star A$ is compact and self-adjoint, the associated sequence of eigenvalues $(b_k^2)_{k\in\mathbb{N}}$ is strictly positive and converges to 0 as $k \to +\infty$. The sequence of eigenvectors $(\phi_k)_{k\in\mathbb{N}}$ is supposed, in the sequel, to be an orthonormal basis of H. For all $k \in \mathbb{N}$, set $\psi_k = b_k^{-1} A\phi_k$. The triple $(b_k, \phi_k, \psi_k)_{k\in\mathbb{N}}$ verifies

$$\begin{aligned} A\phi_k &= b_k\psi_k, \\ A^\star\psi_k &= b_k\phi_k, \end{aligned} \tag{3}$$

for all $k \in \mathbb{N}$. This representation leads to a simpler understanding of the model (1). Indeed, for all $k \in \mathbb{N}$, using Equation (3) and the properties of the Gaussian white noise,

$$y_k = \langle Y, \psi_k\rangle = \langle Af, \psi_k\rangle + \epsilon\langle \xi, \psi_k\rangle = b_k\langle f, \phi_k\rangle + \epsilon\xi_k, \tag{4}$$

where ξ_k are i.i.d. standard Gaussian variables. Hence, for all $k \in \mathbb{N}$, we can obtain from Equation (1) an observation on $\theta_k = \langle f, \phi_k\rangle$. In the ℓ^2-sense, $\theta = (\theta_k)_{k\in\mathbb{N}}$ and f represents the same mathematical object. The sequence space model (4) clarifies the effect of A on the signal f. Since A is compact, $b_k \to 0$ as $k \to +\infty$. For large values of k, the coefficients $b_k\theta_k$ are negligible compared with $\epsilon\xi_k$. In a certain sense, the signal is smoothed by the operator. The recovery becomes difficult in the presence of noise for large ‘frequencies’ (i.e. when k is large).

Our aim is to estimate the sequence $(\theta_k)_{k\in\mathbb{N}} = (\langle f, \phi_k\rangle)_{k\in\mathbb{N}}$. The linear estimation plays an important role in the inverse problem framework and is a starting point for several recovering methods. Let $(\lambda_k)_{k\in\mathbb{N}}$ be a real sequence with values in [0, 1]. In the following, this sequence will be called a filter. The associated linear estimator is defined by

$$\hat{f}_\lambda = \sum_{k=1}^{+\infty} \lambda_k b_k^{-1} y_k \phi_k.$$

In the sequel, $\hat{f}_\lambda$ may be sometimes identified with $\hat{\theta}_\lambda = (\lambda_k b_k^{-1} y_k)_{k\in\mathbb{N}}$. The meaning will be clear from the context. The error related to $\hat{f}_\lambda$ is measured by the quadratic risk $\mathbb{E}_\theta\|\hat{f}_\lambda - f\|^2$. Given a family of estimators T, we would like to construct an estimator $\theta^\star$ comparable with the best possible one contained in T (called the oracle), via the inequality

$$\mathbb{E}_\theta\|\theta^\star - \theta\|^2 \le (1 + \vartheta_\epsilon)\mathbb{E}_\theta\|\theta_T - \theta\|^2 + C\epsilon^2, \tag{5}$$

with $\vartheta_\epsilon, C > 0$. The quantity $C\epsilon^2$ is a residual term. The inequality (5) is said to be sharp if $\vartheta_\epsilon \to 0$ as $\epsilon \to 0$. In this case, $\theta^\star$ asymptotically mimics the behaviour of θ_T. Oracle inequalities play an important, though recent role in statistics. They provide a precise and non-asymptotic measure on the performances of $\theta^\star$, which does not require *a priori* informations on the signal. In several situations, oracle results lead to interesting minimax rates of convergence. This theory has given rise to a considerable amount of literature. We mention in particular Donoho (1995), Barron, Birgé and Massart (1999), Cavalier, Golubev, Picard and Tsybakov (2002) or Candés (2006) for a survey.

The risk hull minimisation (RHM) principle, initiated in Cavalier and Golubev (2006) for spectral cut-off (or projection) schemes, is an interesting approach for the construction of data-driven parameter choice rules. The principle is to identify the stochastic processes that control the behaviour of a projection estimator. Then, a deterministic criterion, called a hull, is constructed in order to contain these processes. We also mention Marteau (in press) for a generalisation of this method to some other regularisation approaches (Tikhonov, Landweber,...).

In this article, our aim is to establish a link between the RHM approach and some specific threshold estimators. We are interested in the family of blockwise constant filters. In this specific case, this approach leads to the penalised blockwise Stein rule studied for instance in Cavalier and Tsybakov (2002). This is a new perspective for this well-known threshold estimator. In particular, the risk hull point of view makes precise the role of the penalty through a simple and general assumption.

This article is organised as follows. In Section 2, we construct a hull for the family of blockwise constant filters. Section 3 establishes a link between the penalised blockwise Stein rule and the risk hull method, and investigates the performances of the related estimator. Section 4 proposes some examples and a discussion on the choice of the penalty. Some results on the theory of ordered processes and the proofs of the main results are gathered in Section 5.

2. A risk hull for blockwise constant filters

In this section, we recall the RHM approach for projection schemes. Then, we explain why an extension of the RHM method may be pertinent. A specific family of estimators is introduced and the related hull is constructed.

2.1. *The risk hull principle*

For all $N \in \mathbb{N}$, denote by $\hat{\theta}_N$ the projection estimator associated with the filter $(\mathbf{1}_{\{k\leq N\}})_{k\in\mathbb{N}}$. For each value of $N \in \mathbb{N}$, the related quadratic risk is

$$\mathbb{E}_\theta \|\hat{\theta}_N - \theta\|^2 = \sum_{k>N} \theta_k^2 + \mathbb{E}_\theta \sum_{k=1}^{N} (b_k^{-1} y_k - \theta_k)^2 = \sum_{k>N} \theta_k^2 + \epsilon^2 \sum_{k=1}^{N} b_k^{-2}. \tag{6}$$

The optimal choice for N is the oracle N^0 that minimises $\mathbb{E}_\theta \|\hat{\theta}_N - \theta\|^2$. It is a trade-off between the two sums (bias and variance) in the r.h.s. of Equation (6). This trade-off cannot be found without *a priori* knowledge on the unknown sequence θ. Data-driven choices for N are necessary.

The classical unbiased risk estimation (URE) approach consists in estimating the quadratic risk. One may use the functional

$$U(y, N) = -\sum_{k=1}^{N} b_k^{-2} y_k^2 + 2\epsilon^2 \sum_{k=1}^{N} b_k^{-2}, \quad \forall N \in \mathbb{N}.$$

The related adaptive bandwidth is defined as

$$\tilde{N} = \arg\min_{N\in\mathbb{N}} U(y, N).$$

Some oracle inequalities related to this approach have been obtained in different papers (see, for instance, Cavalier et al. 2002). Nevertheless, this approach suffers from some drawbacks, especially in the inverse problem framework.

Indeed, this method is based on the average behaviour of the projection estimators: $U(y, N)$ is an estimator of the quadratic risk. This is quite problematic in the inverse problem framework where the main quantities of interest often possess a great variability. This can be illustrated by a very simple example: $f = 0$. In this particular case, for all $N \in \mathbb{N}$, the loss of the related projection estimator $\hat{\theta}_N$ is

$$\|\hat{\theta}_N - \theta\|^2 = \epsilon^2 \sum_{k=1}^{N} b_k^{-2} + \eta_N, \quad \text{with } \eta_N = \epsilon^2 \sum_{k=1}^{N} b_k^{-2}(\xi_k^2 - 1) \ \ \forall N \in \mathbb{N}.$$

Since $b_k \to 0$ as $k \to +\infty$, the process $N \mapsto \eta_N$ possesses a great variability, which explodes with N. In this case, the behaviour of $\mathbb{E}_\theta \|\hat{\theta}_N - \theta\|^2$ and $\|\hat{\theta}_N - \theta\|^2$ are rather different. The variability is neglected when only considering the average behaviour of the loss. This leads in practice to wrong decisions for the choice of N. More generally, as soon as the signal-to-noise ratio is small, one may expect poor performances of the URE method. We refer to Cavalier and Golubev (2006) for a complete discussion illustrated by some numerical simulations.

From now on, the problem is to construct a data-driven bandwidth that takes into account this phenomenon. Instead of the quadratic risk, in Cavalier and Golubev (2006) it is proposed to consider a deterministic term $V(\theta, N)$, called a hull, satisfying

$$\mathbb{E}_\theta \sup_{N \in \mathbb{N}} [\|\hat{\theta}_N - \theta\|^2 - V(\theta, N)] \leq 0. \tag{7}$$

This hull bounds uniformly the loss in the sense of inequality (7). Ideally, it is constructed in order to contain the variability of the projection estimators. The related estimator is then defined as the minimiser of $V(y, N)$, an estimator of $V(\theta, N)$, on $\mathbb{N}$.

The theoretical and numerical properties of this estimator are presented and discussed in detail in Cavalier and Golubev (2006) in the particular case of spectral cut-off regularisation. In the same spirit, we mention Marteau (in press) for an extension of this method to wider regularisation schemes (Landweber, Tikhonov, ...).

2.2. *The choice of* $\boldsymbol{\Lambda}$

In order to construct an estimator leading to an accurate oracle inequality, one must consider both a family of filters Λ and a procedure in order to mimic the behaviour of the best element in Λ.

In this article we are interested in the risk hull principle. This point of view possesses indeed interesting theoretical properties. It makes the role of the stochastic processes involved in linear estimation more precise and leads to an accurate understanding of the problem.

Now, we address the problem of the choice of Λ. In the oracle sense, an ideal goal of adaptation is to obtain a sharp oracle inequality over all possible estimators. This is in most cases an unreachable task since this set is too large. The difficulty of the oracle adaptation increases with the size of the considered family. At a smaller scale, one may consider Λ_{mon}, the family of linear and monotone filters defined as

$$\Lambda_{\text{mon}} = \{\lambda = (\lambda_k)_{k \in \mathbb{N}} \in \ell^2 : 1 \geq \lambda_1 \geq \cdots \geq \lambda_k \geq \cdots \geq 0\},$$

The set Λ_{mon} contains the linear and monotone filters and covers most of the existing linear procedures as the spectral cut-off, Tikhonov, Pinsker or Landweber filters (see, for instance, Engl et al. 1996 or Bissantz et al. 2007). Some oracle inequalities have been already obtained on specific subsets of Λ_{mon} in Cavalier and Golubev (2006) and Marteau (in press), but we would like to consider the whole family at the same time.

The set Λ_{mon} is always rather large and obtaining an explicit estimator in this setting seems difficult. A possible alternative is to consider a set that contains elements presenting a behaviour similar to the best one in Λ_{mon}, but where an estimator could be explicitly constructed. In this sense, the collection of blockwise constant estimators may be a good choice. In the sequel, this family will be identified to the set

$$\Lambda^\star = \{\lambda \in l^2 : 0 \leq \lambda_k \leq 1, \lambda_k = \lambda_{K_j}, \forall k \in [K_j, K_{j+1}-1], j = 0, \dots, J, \lambda_k = 0, k > N\},$$

where J, N and $(K_j)_{j=0,\dots,J}$ are such that $K_0 = 1$, $K_J = N+1$ and $K_j > K_{j-1}$. In the following, we will also use the notations $I_j = \{k \in [K_{j-1}, K_j - 1]\}$ and $T_j = K_j - K_{j-1}$, for all $j \in \{1, \dots, J\}$.

In the following, most of the results are established for a general construction of $\Lambda^\star$. There exists several choices that may lead to interesting results. Typically, $N \to +\infty$ as $\epsilon \to 0$. It is chosen in order to capture most of the nonparametric functions with a controlled bias (see Equation (17) for an example). Concerning the size $(T_j)_{j=1,\dots,J}$ of the blocks, we refer to Cavalier and Tsybakov (2001) for several examples.

The family $\Lambda^\star$ can easily be handled. In particular, each block I_j can be considered independently of the other ones. This simplifies considerably the study of the considered estimators. Moreover, for all $\theta \in \ell^2$,

$$R(\theta, \lambda_{\text{mon}}) = \inf_{\lambda \in \Lambda_{\text{mon}}} R(\theta, \lambda) \quad \text{and} \quad R(\theta, \lambda^0) = \inf_{\lambda \in \Lambda^\star} R(\theta, \lambda) \tag{8}$$

are in fact rather close, subject to some reasonable constraints on the sequences $(b_k)_{k\in\mathbb{N}}$ and $(T_j)_{j=1,\dots,J}$ (see Section 4 or Cavalier and Tsybakov 2002 for more details).

The extension of the RHM principle to the family $\Lambda^\star$ presents other advantages. The related estimator corresponds indeed to a threshold scheme. Hence, we will be able to address the question of the choice of the threshold through the risk hull approach. This may be a new perspective for the blockwise constant adaptive approach, and more generally for this class of regularisation procedures.

2.3. *A risk hull for* $\Lambda^\star$

First, we introduce some notations. For all $j \in \{1, \dots, J\}$, let η_j be defined as follows:

$$\eta_j = \epsilon^2 \sum_{k \in I_j} b_k^{-2}(\xi_k^2 - 1). \tag{9}$$

The random variable η_j plays a central role in blockwise constant estimation. It corresponds to the main stochastic part of the loss in each block I_j. The hull proposed in Theorem 2.1 is constructed in order to contain these terms. Introduce also

$$\rho_\epsilon = \max_{j=1,\dots,J} \sqrt{\Delta_j} \quad \text{and} \quad \|\theta\|_{(j)}^2 = \sum_{k \in I_j} \theta_k^2 \;\; \forall j \in \{1, \dots, J\}, \tag{10}$$

with

$$\Delta_j = \frac{\max_{k \in I_j} \epsilon^2 b_k^{-2}}{\sigma_j^2} \quad \text{and} \quad \sigma_j^2 = \epsilon^2 \sum_{k \in I_j} b_k^{-2}.$$

We will see that $\rho_\epsilon \to 0$ as $\epsilon \to 0$ with appropriate choices of blocks and minor assumptions on the sequence $(b_k)_{k\in\mathbb{N}}$ (see Section 4 for more details).

From now on, we are able to present a hull for the family $\Lambda^\star$, that is, a deterministic sequence $(V(\theta,\lambda))_{\lambda\in\Lambda^\star}$ verifying

$$\mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\{\|\hat{\theta}_\lambda-\theta\|^2 - V(\theta,\lambda)\} \le 0.$$

The proof of the following result is postponed to Section 5.2.

THEOREM 2.1 *Let* $(\mathrm{pen}_j)_{j=1,\dots,J}$ *be a positive sequence verifying*

$$\sum_{j=1}^{J} \mathbb{E}[\eta_j - \mathrm{pen}_j]_+ \le C_1\epsilon^2, \tag{11}$$

for some positive constant C_1. *Then, there exists* $B>0$ *such that*

$$V(\theta,\lambda) = (1+B\rho_\epsilon)\left\{\sum_{j=1}^{J}[(1-\lambda_{K_j})^2\|\theta\|^2_{(j)} + \lambda^2_{K_j}\sigma_j^2 + 2\lambda_{K_j}\mathrm{pen}_j] + \sum_{k>N}\theta_k^2\right\} \\ + C_1\epsilon^2 + B\rho_\epsilon R(\theta,\lambda^0), \tag{12}$$

is a risk hull on $\Lambda^\star$.

Theorem 2.1 states in fact that the penalised quadratic risk,

$$R_{\mathrm{pen}}(\theta,\lambda) = \sum_{j=1}^{J}[(1-\lambda_{K_j})^2\|\theta\|^2_{(j)} + \lambda^2_{K_j}\sigma_j^2] + \sum_{k>N}\theta_k^2 + 2\sum_{j=1}^{J}\lambda_{K_j}\mathrm{pen}_j, \tag{13}$$

is, up to some constants and residual terms, a risk hull on the family $\Lambda^\star$. Hence, we will use $R_{\mathrm{pen}}(\theta,\lambda)$ as a criterion for the construction of a data-driven filter on $\Lambda^\star$, provided that inequality (11) is satisfied (see Section 3).

The construction of a hull can be reduced to the choice of a penalty $(\mathrm{pen}_j)_{j=1,\dots,J}$, provided Equation (11) is verified. A brief discussion concerning this assumption is presented in Section 3. Some examples of penalties are presented in Section 4.

3. Oracle inequalities

In Section 2, we have proposed a family of hulls indexed by the penalty $(\mathrm{pen}_j)_{j=1,\dots,J}$. In this section, we are interested in the performances of the estimators constructed from these hulls.

In the sequel, we set $\lambda_j=\lambda_{K_j}$ for all $j\in\{1,\dots,J\}$. This is a slight abuse of notation, but the meaning will be clear from the context. Then define

$$U_{\mathrm{pen}}(y,\lambda) = \sum_{j=1}^{J}[(\lambda_j^2 - 2\lambda_j)(\|\tilde{y}\|^2_{(j)} - \sigma_j^2) + \lambda_j^2\sigma_j^2 + 2\lambda_j\mathrm{pen}_j],$$

where

$$\|\tilde{y}\|^2_{(j)} = \epsilon^2\sum_{k\in I_j} b_k^{-2}y_k^2, \quad \forall j\in\{1,\dots,J\}.$$

The term $U_{\mathrm{pen}}(y,\lambda)$ is an estimator of the penalised quadratic risk $R_{\mathrm{pen}}(\theta,\lambda)$ defined in Equation (13). Recall that from Theorem 2.1, $R_{\mathrm{pen}}(\theta,\lambda)$ is, up to some constant and residual terms, a risk hull. Let $\theta^\star$ denote the estimator associated with the filter

$$\lambda^\star = \arg\min_{\lambda\in\Lambda^\star} U_{\text{pen}}(y,\lambda). \tag{14}$$

Using simple algebra, we can prove that the solution of Equation (14) is

$$\lambda_k^\star = \begin{cases} \left(1 - \left(\sigma_j^2 + \text{pen}_j\right)/\parallel \tilde{y} \parallel_{(j)}^2\right)_+, & k \in I_j, \quad j = 1,\ldots,J, \\ 0, & k > N. \end{cases} \tag{15}$$

This filter behaves as follows. For all $j \in \{1,\ldots,J\}$, $\lambda_j^\star$ compares the term $\|\tilde{y}\|_{(j)}^2$ with $\sigma_j^2 + \text{pen}_\text{j}$. When $\|\theta\|_{(j)}^2$ is 'small' (or even equal to 0), this comparison may lead to wrong decision. Indeed, $\|\tilde{y}\|_{(j)}^2$ is in this case close to $\sigma_j^2 + \eta_j$. The variance of the variables η_j is very large since $b_k \to 0$ as $k \to +\infty$. Fortunately, these variables are uniformly bounded by the penalty in the sense of Equation (11). Hence, $\lambda_j^\star$ should be close to 0 for 'small' $\|\theta\|_{(j)}^2$. Theorem 3.1 emphasises this heuristic discussion through a simple oracle inequality.

Remark that the particular case $\text{pen}_j = 0$ for all $j \in \{1,\ldots,J\}$ leads to the URE approach. Inequality (11) does not hold in this setting.

THEOREM 3.1 *Let $\theta^\star$ be the estimator associated with the filter $\lambda^\star$. Assume that inequality* (11) *holds. Then, there exists $C^\star > 0$ independent of ϵ such that, for all $\theta \in \ell^2$ and any* $0 < \epsilon < 1$,

$$\mathbb{E}_\theta\|\theta^\star - \theta\|^2 \le (1+\tau_\epsilon)\inf_{\lambda\in\Lambda^\star} R(\theta,\lambda) + C^\star\epsilon^2,$$

where $\tau_\epsilon \to 0$ as $\epsilon \to 0$ provided $\max_j \text{pen}_j/\sigma_j^2 \to 0$ *and $\rho_\epsilon \to 0$ as $\epsilon \to 0$.*

Although this result is rather general, the constraints on the blocks and the penalty are only expressed through one inequality (here Equation (11)). This is one of the advantages of the RHM approach.

For the particular choice $\text{pen}_j = \varphi_j\sigma_j^2$ leading to the penalised blockwise Stein rule, we obtain a simpler assumption than in Cavalier and Tsybakov (2002). This is an interesting outcome.

We conclude this section with an oracle inequality on Λ_{mon}, the family of monotone filters. We take advantage of the closeness between $\Lambda^\star$ and Λ_{mon} under specific conditions. For the sake of convenience, we restrict to one specific type of blocks.

Let $\nu_\epsilon = \lceil \log\epsilon^{-1}\rceil$ and $\kappa_\epsilon = \log^{-1}\nu_\epsilon$, where for all $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the minimal integer strictly greater than x. Define the sequence $(T_j)_{j=1,\ldots,J}$ by

$$T_1 = \lceil \nu_\epsilon \rceil, \quad T_j = \lceil \nu_\epsilon(1+\kappa_\epsilon)^{j-1}\rceil, \quad j > 1, \tag{16}$$

and the bandwidth J as

$$J = \min\{j : K_j > \bar{N}\}, \quad \text{with } \bar{N} = \max\left\{m : \sum_{k=1}^m b_k^{-2} \le \epsilon^{-2}\kappa_\epsilon^{-3}\right\}. \tag{17}$$

COROLLARY 3.2 *Assume that $(b_k) \sim (k^{-\beta})_{k\in\mathbb{N}}$ for some $\beta > 0$ and that the sequence $(\text{pen}_j)_{j=1,\ldots,J}$ satisfies inequality* (11). *Then , for any $\theta \in \ell^2$ and $0 < \epsilon < \epsilon_1$, we have*

$$\mathbb{E}_\theta\|\theta^\star - \theta\|^2 \le (1+\Gamma_\epsilon)\inf_{\lambda\in\Lambda_{mon}} R(\theta,\lambda) + C_2\epsilon^2,$$

where C_2, ϵ_1 denote positive constants independent of ϵ, and $\Gamma_\epsilon \to 0$ as $\epsilon \to 0$.

The proof is a direct consequence of Lemma 1 of Cavalier and Tsybakov (2002). It can be in fact extended to other constructions for blocks. One has only to verify that

$$\max_{j=1,\dots,J-1} \frac{\sigma_{j+1}^2}{\sigma_j^2} \leq 1 + \eta_\epsilon, \quad \text{for } 0 < \eta_\epsilon < \frac{1}{2}.$$

The inequality is sharp if $\eta_\epsilon \to 0$ as $\epsilon \to 0$. The interested reader can refer to Cavalier and Tsybakov (2001) for some examples of blocks.

The results obtained in this section hold for a wide range of penalties. This range is characterised and studied in the next section.

4. Some choices of penalty

In this section, we present two possible choices of penalty satisfying inequality (11). Then, we present a brief discussion on the range in which this sequence can be chosen. The goal of this section is not to say what could be a good penalty. This question is rather ambitious and may require more than a single paper. Our aim here is rather to present some hint on the way it could be chosen and on the related problem.

For the sake of convenience, we use in this section the same framework of Corollary 3.2. We assume that the sequence of eigenvalues possesses a polynomial behaviour, that is, $(b_k) \sim (k^{-\beta})_{k\in\mathbb{N}}$ for some $\beta > 0$. Concerning the set $\Lambda^\star$, we consider the weakly geometrically increasing blocks defined in Equations (16) and (17). All the results presented in the sequel hold for other constructions (see, for instance, Cavalier and Tsybakov 2001). We leave the proof to the interested reader. Concerning the sequence $(b_k)_{k\in\mathbb{N}}$, the relaxation of the assumption on polynomial behaviour is not straightforward. In particular, considering exponentially decreasing eigenvalues require a specific treatment in this setting.

Let u and v be two real sequences. Here, and in the sequel, for all $k \in \mathbb{N}$, we write $u_k \lesssim v_k$ if we can find a positive constant C independent of k such that $u_k \leq C v_k$, and $u_k \simeq v_k$ if both $u_k \lesssim v_k$ and $u_k \gtrsim v_k$. Since the sequence $(b_k)_{k\in\mathbb{N}}$ possesses a polynomial behaviour, we can write that for all $j \in \{1, \dots, J\}$

$$\sigma_j^2 = \epsilon^2 \sum_{k\in I_j} b_k^{-2} \simeq \epsilon^2 K_j^{2\beta}(K_{j+1} - K_j)$$

and

$$\Delta_j = \frac{K_{j+1}^{2\beta}}{K_j^{2\beta}(K_{j+1} - K_j)} \simeq (K_{j+1} - K_j)^{-1},$$

since $K_{j+1}/K_j \to 1$ as $j \to +\infty$.

4.1. *Examples*

The following lemma provides upper bounds on the term $\mathbb{E}_\theta[\eta_j - \text{pen}_j]_+$ and makes more explicit the behaviour of the penalty. It can be used to prove inequality (11) in several situations.

LEMMA 4.1 *For all $j \in \{1, \dots, J\}$ and δ such that* $0 < \delta < \epsilon^{-2} b_{K_j-1}^2/2$,

$$\mathbb{E}[\eta_j - \text{pen}_j]_+ \leq \delta^{-1} \exp\left\{-\delta \text{pen}_j + \delta^2 \Sigma_j^2 + 4\delta^3 \sum_{k\in I_j} \frac{\epsilon^6 b_k^{-6}}{(1 - 2\delta\epsilon^2 b_{K_j-1}^{-2})}\right\},$$

with

$$\Sigma_j^2 = \epsilon^4 \sum_{k \in I_j} b_k^{-4}, \quad \forall j \in \{1, \ldots, J\}. \tag{18}$$

Proof Let $j \in \{1, \ldots, J\}$ be fixed. First, remark that for all $\delta > 0$,

$$\begin{aligned} \mathbb{E}_\theta[\eta_j - \text{pen}_j]_+ &= \int_{\text{pen}_j}^{+\infty} P(\eta_j \geq t)\,\mathrm{d}t, \\ &= \int_{\text{pen}_j}^{+\infty} P(\exp(\delta\eta_j) \geq e^{\delta t})\,\mathrm{d}t, \\ &\leq \delta^{-1} \mathrm{e}^{-\delta \text{pen}_j} \mathbb{E}_\theta \exp(\delta\eta_j). \end{aligned}$$

Then, provided $0 < \delta < \epsilon^{-2} b_{K_{j-1}}^2/2$,

$$\mathbb{E}_\theta \exp(\delta\eta_j) \leq \exp\left\{ \delta^2 \Sigma_j^2 + 4\delta^3 \sum_{k \in I_j} \frac{\epsilon^6 b_k^{-6}}{(1 - 2\epsilon^2 \delta b_k^{-2})_+} \right\}.$$

This concludes the proof. ∎

The principle of RHM leads to an interesting choice. The only restriction on $(\text{pen}_j)_{j=1,\ldots,J}$ from the risk hull point of view is expressed through inequality (11) as follows:

$$\sum_{j=1}^{J} \mathbb{E}_\theta[\eta_j - \text{pen}_j]_+ \leq C_1 \epsilon^2,$$

for some positive constant C_1. Since $\mathbb{E}_\theta[\eta_j - u]_+ \leq \mathbb{E}_\theta \eta_j \mathbf{1}_{\{\eta_j \geq u\}}$ for all positive u, we may be interested in the penalty

$$\overline{\text{pen}}_j = (1+\alpha) U_j, \quad \text{with } U_j = \inf\{u : \mathbb{E}_\theta \eta_j \mathbf{1}_{\{\eta_j \geq u\}} \leq \epsilon^2\}, \ \ \forall j \in \{1, \ldots, J\}, \tag{19}$$

for some $\alpha > 0$. This penalty is an extension of the sequence proposed by Cavalier and Golubev (2006) for spectral cut-off schemes.

The next corollary establishes that the sequence $(\overline{\text{pen}}_j)_{j=1,\ldots,J}$ is a relevant choice for the penalty. We obtain a sharp oracle inequality for the related estimator. In particular, inequality (11) holds, i.e. the penalty contains the variability of the problem.

Corollary 4.2 *Let $\theta^\star$ be the estimator introduced in Equation* (15) *with the penalty* $(\overline{\text{pen}}_j)_{j=1,\ldots,J}$. *Then,*

$$\mathbb{E}_\theta \|\theta^\star - \theta\|^2 \leq (1 + \gamma_\epsilon) \inf_{\lambda \in \Lambda^\star} R(\theta, \lambda) + \frac{C_4}{\alpha} \epsilon^2, \tag{20}$$

where C_4 denotes a positive constant independent of ϵ and $\gamma_\epsilon = o(1)$ as $\epsilon \to 0$.

Proof We use the following lower bound on U_j:

$$U_j = \inf\{u : \mathbb{E}_\theta \eta_j \mathbf{1}_{\{\eta_j \geq u\}} \leq \epsilon^2\} \geq \sqrt{2\Sigma_j^2 \log(C\epsilon^{-4}\Sigma_j^2)}, \tag{21}$$

where for all $j \in \{1, \ldots, J\}$, Σ_j^2 is defined in Equation (18). The proof can be directly derived from the Lemma 1 of Cavalier and Golubev (2006). Then, thanks to Theorems 2.1 and 3.1, we only have to prove that inequality (11) holds since $\max_j \overline{\text{pen}}_j/\sigma_j^2$ converges to 0 as $\epsilon \to 0$. For all $j \in \{1,\dots,J\}$, using Lemma 4.1,

$$\mathbb{E}[\eta_j - \overline{\text{pen}}_j]_+ \le \frac{1}{\delta}\exp\left\{-\delta\overline{\text{pen}}_j + \delta^2\Sigma_j^2 + 4\delta^3\sum_{k\in I_j}\frac{\epsilon^6 b_k^{-6}}{(1-2\delta\epsilon^2 b_{K_{j-1}}^{-2})}\right\}, \tag{22}$$

for all $0<\delta<\epsilon^{-2}b_{K_j-1}^2/2$. Setting

$$\delta = \sqrt{\frac{\log(C\epsilon^{-4}\Sigma_j^2)}{2\Sigma_j^2}},$$

and using Equation (21), we obtain

$$\mathbb{E}[\eta_j - \overline{\text{pen}}_j]_+ \le \sqrt{\frac{2\Sigma_j^2}{\log(C\epsilon^{-4}\Sigma_j^2)}}\exp\left\{\frac{1}{2}\log(C\epsilon^{-4}\Sigma_j^2)\right\}\times\exp\{-(1+\alpha)\log(C\epsilon^{-4}\Sigma_j^2)\},$$

$$\le C\epsilon^2\sqrt{\frac{1}{\log(C\epsilon^{-4}\Sigma_j^2)}}\exp\{-\alpha\log(C\epsilon^{-4}\Sigma_j^2)\}.$$

Indeed, provided Equations (16) and (17) hold, $\delta b_{K_j-1}^{-2}$ and the last term in the right-hand side of the exponential in Equation (22) converge to 0 as $j\to+\infty$. Hence, we eventually obtain

$$\sum_{j=1}^{J}\mathbb{E}[\eta_j - \overline{\text{pen}}_j]_+ \le C\epsilon^2\sum_{j=1}^{J}\frac{1}{\log^{1/2}(CT_j)}\exp\{-\alpha\log(CT_j)\},$$

$$\le C\epsilon^2\sum_{j=1}^{J} j^{-1/2}\exp\{-\alpha\log(C\nu_\epsilon(1+\kappa_\epsilon)^j)\},$$

$$\le C\epsilon^2\sum_{j=1}^{+\infty} j^{-1/2}\exp\{-\alpha Dj\} < \frac{C\epsilon^2}{\alpha},$$

where D and C denote two positive constants independent of ϵ. This concludes the proof of Corollary 4.3. ∎

The penalty (19) is not explicit. Nevertheless, it can be computed using Monte–Carlo approximation: there are only J terms to compute. Remark that it is also possible to deal with the lower bound (21) which is explicit. The theoretical results are essentially the same since we use this bound in the proof of Corollary 4.3.

Now, we consider the penalty introduced in Cavalier and Tsybakov (2002). For all $j\in\{1,\dots,J\}$, it is defined as follows:

$$\text{pen}_j^{CT} = \Delta_j^{\gamma}\sigma_j^2, \quad \text{with } 0<\gamma<\frac{1}{2}.$$

Remark that with our assumptions on $\Lambda^\star$ and $(b_k)_{k\in\mathbb{N}}$,

$$\text{pen}_j^{CT} \simeq \epsilon^2 K_j^{2\beta}(K_{j+1}-K_j)^{1-\gamma} \gtrsim \overline{\text{pen}}_j.$$

This inequality entails that Equation (11) is satisfied. Hence, an oracle inequality similar to Equation (20) can be obtained for this sequence. This is the same result as in Cavalier and Tsybakov (2002). However, we construct a different proof, thanks to the RHM approach.

4.2. *The range*

Theorem 3.1 provides in fact an admissible range for the penalty. If we want a sharp oracle inequality, necessarily $\max_j \text{pen}_j/\sigma_j^2 \to 0$ as $\epsilon \to 0$. Hence, the penalty should not be too large. At the same time, we require from inequality (11) that the penalty contains in a certain sense the variables $(\eta_j)_{j=1,\dots,J}$. Hence, small penalties will not be convenient.

From inequality (11) and Lemma 4.1, the sequence $(\text{pen}_j)_{j=1,\dots,J}$ should at least fulfil $\text{pen}_j \gtrsim \Sigma_j$ for all $j \in \{1, \dots, J\}$. Since we require at the same time $\max_j \sigma_j^2/\text{pen}_j \to 0$ as $j \to +\infty$, an admissible penalty in the sense of Theorem 3.1 should satisfy

$$\Sigma_j \lesssim \text{pen}_\text{j} \lesssim \sigma_j^2, \quad \forall j \in \{1, \dots, J\}. \tag{23}$$

With our assumptions, this range is equivalent to

$$\epsilon^2 K_j^{2\beta}(K_{j+1} - K_j)^{1/2} \lesssim \text{pen}_j \lesssim \epsilon^2 K_{j+1}^{2\beta}(K_{j+1} - K_j), \quad \forall j \in \{1, \dots, J\}.$$

It is possible to prove that similar to Equation (20) oracle inequality holds for all penalty $(\text{pen}_j)_{j=1,\dots,J}$ satisfying Equation (23). This is in particular the case for $(\overline{\text{pen}}_j)_{j=1,\dots,J}$ and $(\text{pen}_j^{CT})_{j=1,\dots,J}$.

Using the same bounds as in the proof of Corollary 4.3, it seems difficult to obtain a sharp oracle inequality with the penalty $(\Sigma_j)_{j=1,\dots,J}$. Nevertheless, the range (23) is derived from upper bounds on the estimator $\theta^\star$ and may certainly be refined. A lower bound approach may perhaps produce interesting results (see for instance (Efromovich 2007)).

In order to conclude this discussion, it would be interesting to compare the two penalties presented in this section. Remark that $(\text{pen}_j)_{j=1,\dots,J}$ is closer to the lower bound of the range that $(\text{pen}_j)_{j=1,\dots,J}$. However, we do not claim that a penalty is better than another. This is an interesting but a very difficult question that should be addressed in a separate paper.

4.3. *Conclusion*

The main contribution of this article is an extension of the RHM method to the family $\Lambda^\star$ and a link between the penalised blockwise Stein rule and the risk hull approach. It is rather surprising that a threshold estimator may be studied via some tools usually related to a parameter selection setting. In any case, this approach allows us to develop a general study on this threshold estimator. In particular, we impose a simple assumption on the threshold that is related to the variance of the η_j in each block. This treatment may certainly be applied to other existing adaptive approaches. For instance, one may be interested in wavelet threshold estimators in a wavelet-vaguelette decomposition framework (Donoho 1995). The generalisation of our work in this setting is not straightforward since the 'blocks' are of size one. Nevertheless, this approach may provide some new interesting perspectives.

In order to conclude this article, it seems necessary to discuss the role played by the constant α in the penalty $(\overline{\text{pen}}_j)_{j=1,\dots,J}$. Inequality (11) does not hold for $\alpha = 0$. On the other hand, the proof of Corollary 4.3 indicates that large values for α will not lead to an accurate recovering. The choice of α has already been discussed and illustrated via some numerical simulations in a slightly different setting, see Cavalier and Golubev (2006) or Marteau (in press) for more details. Remark that we do not require α to be greater than one in this article. This is a small difference compared with the constraints expressed in a regularisation parameter choice scheme. This can be explained by the blockwise structure of the variables $(\eta_j)_{j=1,\dots,J}$.

5. Proofs and technical lemmas

5.1. *Ordered processes*

Ordered processes were introduced in Kneip (1994). In Cao and Golubev (2006), these processes are studied in detail and very interesting tools are provided. These stochastic objects may play an important role in the adaptive estimation, see in particular Golubev (2007) or Marteau (in press) for more details.

The aim of this section is not to provide an exhaustive presentation of this theory but rather to introduce some definitions and useful properties.

DEFINITION 5.1 *Let $\zeta(t)$, $t \geq 0$, be a separable random process with $\mathbb{E}\zeta(t) = 0$ and finite variance $\Sigma^2(t)$. It is called ordered if, for all $t_2 \geq t_1 \geq 0$,*

$$\Sigma^2(t_2) \geq \Sigma^2(t_1) \quad \text{and} \quad \mathbb{E}[\zeta(t_2) - \zeta(t_1)]^2 \leq \Sigma^2(t_2) - \Sigma^2(t_1).$$

Let ζ be a standard Gaussian random variable. The process $t \mapsto \zeta t$ is the most simple example of ordered process. Wiener processes are also covered by Definition 5.1. The family of ordered processes is in fact quite large.

ASSUMPTION C1 *There exists $\kappa > 0$ such that*

$$\varphi(\kappa) = \sup_{t_1, t_2} \mathbb{E} \exp\left\{\kappa \frac{\zeta(t_1) - \zeta(t_2)}{\sqrt{\mathbb{E}[\zeta(t_1) - \zeta(t_2)]^2}}\right\} < +\infty.$$

This assumption is not very restrictive. Several processes encountered in linear estimation satisfy it.

The proof of the following result can be found in Cao and Golubev (2006).

LEMMA 5.2 *Let $\zeta(t)$, $t \geq 0$, be an ordered process satisfying $\zeta(0) = 0$ and Assumption C1. There exists a constant $C = C(\kappa)$ such that, for all $\gamma > 0$,*

$$\mathbb{E} \sup_{t \geq 0} [\zeta(t) - \gamma \Sigma^2(t)]_+ \leq \frac{C}{\gamma}.$$

This lemma is rather important in the theory of ordered processes and leads to several interesting results. In particular, the following corollary will be often used in the proofs.

COROLLARY 5.3 *Let $\zeta(t)$, $t \geq 0$, be an ordered process satisfying $\zeta(0) = 0$ and Assumption* C1. *Consider $\hat{t}$ measurable with respect to ζ. Then, there exists $C = C(\kappa) > 0$ such that*

$$\mathbb{E}\zeta(\hat{t}) \leq C\sqrt{\mathbb{E}\Sigma^2(\hat{t})}.$$

Proof Let $\gamma > 0$ be fixed. Using Lemma 5.2,

$$\begin{aligned} \mathbb{E}\zeta(\hat{t}) &= \mathbb{E}\zeta(\hat{t}) - \gamma \mathbb{E}\Sigma^2(\hat{t}) + \gamma \mathbb{E}\Sigma^2(\hat{t}), \\ &\leq \mathbb{E} \sup_{t \geq 0} [\zeta(t) - \gamma \Sigma^2(t)]_+ + \gamma \mathbb{E}\Sigma^2(\hat{t}), \\ &\leq \frac{C}{\gamma} + \gamma \mathbb{E}\Sigma^2(\hat{t}). \end{aligned}$$

Choose $\gamma = (\mathbb{E}\Sigma^2(\hat{t}))^{-1/2}$ in order to conclude the proof. ■

5.2. *Proofs*

Proof of Theorem 2.1 First, remark that

$$\begin{aligned}
&\mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\{\|\hat{\theta}_\lambda - \theta\|^2 - V(\theta,\lambda)\} \\
&\quad = \mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\left\{\sum_{k=1}^{+\infty}(1-\lambda_k)^2\theta_k^2 + \epsilon^2\sum_{k=1}^{+\infty}\lambda_k^2 b_k^{-2}\xi_k^2 - 2\epsilon\sum_{k=1}^{+\infty}\lambda_k(1-\lambda_k)\theta_k b_k^{-1}\xi_k - V(\theta,\lambda)\right\}, \\
&\quad = \mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\left\{\sum_{j=1}^{J}\left[(1-\lambda_j)^2\|\theta\|_{(j)}^2 + \lambda_j^2\sum_{k\in I_j}\epsilon^2 b_k^{-2}\xi_k^2 - 2\lambda_j(1-\lambda_j)X_j\right] + \sum_{k>N}\theta_k^2 - V(\theta,\lambda)\right\}, \\
&\quad = \mathbb{E}_\theta \sum_{j=1}^{J}\left[(1-\hat{\lambda}_j)^2\|\theta\|_{(j)}^2 + \hat{\lambda}_j^2\sum_{k\in I_j}\epsilon^2 b_k^{-2}\xi_k^2 + 2\hat{\lambda}_j(\hat{\lambda}_j-1)X_j\right] + \sum_{k>N}\theta_k^2 - \mathbb{E}_\theta V(\theta,\hat{\lambda}),
\end{aligned}$$

with

$$\hat{\lambda} = \arg\sup_{\lambda\in\Lambda^\star}\{\|\hat{\theta}_\lambda - \theta\|^2 - V(\theta,\lambda)\}$$

and

$$X_j = \epsilon\sum_{k\in I_j}\theta_k b_k^{-1}\xi_k, \quad \forall j\in\{1,\ldots,J\}. \tag{24}$$

Let $j\in\{1,\ldots,J\}$ be fixed. Use the decomposition

$$\begin{aligned}
\mathbb{E}_\theta 2\hat{\lambda}_j(\hat{\lambda}_j-1)X_j &= \mathbb{E}_\theta\hat{\lambda}_j^2 X_j + \mathbb{E}_\theta(\hat{\lambda}_j^2 - 2\hat{\lambda}_j)X_j, \\
&= \mathbb{E}_\theta\hat{\lambda}_j^2 X_j + \mathbb{E}_\theta(1-\hat{\lambda}_j)^2 X_j = A_j^1 + A_j^2,
\end{aligned} \tag{25}$$

since $\mathbb{E}_\theta X_j = 0$. First, consider A_j^1. Let λ_j^0 denote the blockwise constant oracle on the block j. Using Corollary 5.3 in Section 5.1,

$$A_j^1 = \mathbb{E}_\theta\hat{\lambda}_j^2 X_j = \mathbb{E}_\theta[\hat{\lambda}_j^2 - (\lambda_j^0)^2]X_j \le C\sqrt{\mathbb{E}_\theta[\hat{\lambda}_j^2 - (\lambda_j^0)^2]^2\sum_{k\in I_j}\epsilon^2 b_k^{-2}\theta_k^2},$$

where $C>0$ denotes a positive constant. Indeed, both processes $\zeta : t\mapsto (t^2 - (\lambda_j^0)^2)X_j$, $t\in[(\lambda_j^0)^2, 1]$ and $\bar{\zeta} : t\mapsto (t^{-2} - (\lambda_j^0)^2)X_j$, $t\in[(\lambda_j^0)^{-1}; +\infty[$ are ordered and satisfy Assumption C1. For all $\gamma>0$, use

$$[\hat{\lambda}_j^2 - (\lambda_j^0)^2]^2 \le 4[(1-\hat{\lambda}_j)^2 + (1-\lambda_j^0)^2](\hat{\lambda}_j^2 + (\lambda_j^0)^2),$$

and the Cauchy–Schwartz and Young inequalities

$$\begin{aligned}
A_j^1 &\le C\sqrt{\mathbb{E}_\theta[(1-\hat{\lambda}_j)^2 + (1-\lambda_j^0)^2](\hat{\lambda}_j^2 + (\lambda_j^0)^2)\max_{k\in I_j}\epsilon^2 b_k^{-2}\|\theta\|_{(j)}^2}, \\
&\le C\mathbb{E}_\theta[\gamma(1-\hat{\lambda}_j)^2\|\theta\|_{(j)}^2 + \gamma^{-1}\Delta_j\hat{\lambda}_j^2\sigma_j^2] + C\gamma(1-\lambda_j^0)^2\|\theta\|_{(j)}^2 \\
&\quad + C\gamma^{-1}\Delta_j(\lambda_j^0)^2\sigma_j^2 + C\sqrt{\mathbb{E}_\theta(1-\hat{\lambda}_j)^2\hat{\lambda}_j^2\max_{k\in I_j}\epsilon^2 b_k^{-2}\|\theta\|_{(j)}^2},
\end{aligned} \tag{26}$$

for some positive constant C. The bound of the last term in the r.h.s. of Equation (26) requires some work. First, suppose that

$$\|\theta\|_{(j)}^2 \leq \sigma_j^2. \tag{27}$$

In such a situation, for all $\gamma > 0$,

$$\sqrt{\mathbb{E}_\theta (1-\hat{\lambda}_j)^2 \hat{\lambda}_j^2 \max_{k \in I_j} \epsilon^2 b_k^{-2} \|\theta\|_{(j)}^2} \leq \sqrt{\|\theta\|_{(j)}^2 \mathbb{E}_\theta \hat{\lambda}_j^2 \max_{k \in I_j} \epsilon^2 b_k^{-2}},$$
$$\leq \gamma \|\theta\|_{(j)}^2 + \gamma^{-1} \Delta_j \mathbb{E}_\theta \hat{\lambda}_j^2 \sigma_j^2.$$

If Equation (27) holds, then

$$\|\theta\|_{(j)}^2 = \frac{\sigma_j^2 \|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} \left(1 + \frac{\|\theta\|_{(j)}^2}{\sigma_j^2}\right) \leq 2\{(1-\lambda_j^0)^2 \|\theta\|_{(j)}^2 + (\lambda_j^0)^2 \sigma_j^2\},$$

where λ^0 is the oracle defined in Equation (8). Indeed,

$$\lambda_j^0 = \frac{\|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2}, \quad \forall j \in \{1, \dots, J\}.$$

Now, suppose

$$\|\theta\|_{(j)}^2 > \sigma_j^2. \tag{28}$$

Then, for all $\gamma > 0$,

$$\sqrt{\mathbb{E}_\theta (1-\hat{\lambda}_j)^2 \hat{\lambda}_j^2 \max_{k \in I_j} \epsilon^2 b_k^{-2} \|\theta\|_{(j)}^2} \leq \sqrt{\max_{k \in I_j} \epsilon^2 b_k^{-2} \mathbb{E}_\theta (1-\hat{\lambda}_j)^2 \|\theta\|_{(j)}^2},$$
$$\leq \gamma \mathbb{E}_\theta (1-\hat{\lambda}_j)^2 \|\theta\|_{(j)}^2 + \gamma^{-1} \Delta_j \sigma_j^2.$$

Using Equation (28),

$$\sigma_j^2 = \frac{\sigma_j^2 \|\theta\|_{(j)}^2}{\sigma_j^2 + \|\theta\|_{(j)}^2} \left(1 + \frac{\sigma_j^2}{\|\theta\|_{(j)}^2}\right) \leq 2\{(1-\lambda_j^0)^2 \|\theta\|_{(j)}^2 + (\lambda_j^0)^2 \sigma_j^2\}.$$

Setting $\gamma = \sqrt{\Delta_j}$, we eventually obtain

$$A_j^1 \leq C\sqrt{\Delta_j} \mathbb{E}_\theta [(1-\hat{\lambda}_j)^2 \|\theta\|_{(j)}^2 + \hat{\lambda}_j^2 \sigma_j^2] + C\sqrt{\Delta_j} [(1-\lambda_j^0)^2 \|\theta\|_{(j)}^2 + (\lambda_j^0)^2 \sigma_j^2], \tag{29}$$

for some constant $C > 0$ independent of ϵ. The same bound occurs for the term A_j^2 in Equation (25). Hence, there exists $B > 0$ independent of ϵ such that

$$\mathbb{E}_\theta \sup_{\lambda \in \Lambda^\star} \{\|\hat{\theta}_\lambda - \theta\|^2 - V(\theta, \lambda)\}$$
$$\leq \mathbb{E}_\theta \sum_{j=1}^J \left[(1 + B\rho_\epsilon)(1-\hat{\lambda}_j)^2 \|\theta\|_{(j)}^2 + \hat{\lambda}_j^2 \sum_{k \in I_j} \epsilon^2 b_k^{-2} \xi_k^2 + B\rho_\epsilon \hat{\lambda}_j^2 \sigma_j^2 \right]$$
$$+ \sum_{k > N} \theta_k^2 + B\rho_\epsilon R(\theta, \lambda^0) - \mathbb{E}_\theta V(\theta, \hat{\lambda}),$$

$$\leq \mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star} \left\{ \sum_{j=1}^{J} \left[(1+B\rho_\epsilon)(1-\lambda_j)^2\|\theta\|^2_{(j)} + \lambda_j^2 \sum_{k\in I_j} \epsilon^2 b_k^{-2}\xi_k^2 + B\rho_\epsilon\lambda_j^2\sigma_j^2 \right] \right.$$
$$\left. + \sum_{k>N}\theta_k^2 + B\rho_\epsilon R(\theta,\lambda^0) - V(\theta,\lambda) \right\},$$

where ρ_ϵ is defined in Equation (10). Now, using Equations (9) and (12),

$$\mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\{\|\hat{\theta}_\lambda - \theta\|^2 - V(\theta,\lambda)\} \leq \mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\left\{ \sum_{j=1}^{J}[\lambda_j^2\eta_j - 2\lambda_j \mathrm{pen}_j] - C_1\epsilon^2 \right\},$$
$$= \sum_{j=1}^{J}\mathbb{E}_\theta \sup_{\lambda_j\in[0,1]}[\lambda_j^2\eta_j - 2\lambda_j\mathrm{pen}_j] - C_1\epsilon^2.$$

Let $j \in \{1,\dots,J\}$ be fixed. We are looking for $\lambda_j \in [0,1]$ that maximises the quantity $\lambda_j^2\eta_j - 2\lambda_j\mathrm{pen}_j$. If $\eta_j < 0$, the function $\lambda \mapsto \lambda^2\eta_j - 2\lambda\mathrm{pen}_j$ is concave and the maximum on $[0,1]$ is attained for $\lambda = 0$. Now, if $\eta_j > 0$, the function $\lambda \mapsto \lambda^2\eta_j - 2\lambda\mathrm{pen}_j$ is convex and the maximum on $[0,1]$ is attained in 0 or 1. Therefore,

$$\sup_{\lambda_j\in[0,1]}(\lambda_j^2\eta_j - 2\lambda_j\mathrm{pen}_j) = [\eta_j - 2\mathrm{pen}_j]_+, \tag{30}$$

Using inequality (11), we eventually obtain

$$\mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\{\|\hat{\theta}_\lambda - \theta\|^2 - V(\theta,\lambda)\} \leq \sum_{j=1}^{J}\mathbb{E}_\theta[\eta_j - 2\mathrm{pen}_j]_+ - C_1\epsilon^2,$$
$$\leq \sum_{j=1}^{J}\mathbb{E}_\theta[\eta_j - \mathrm{pen}_j]_+ - C_1\epsilon^2 \leq 0.$$

This concludes the proof of Theorem 2.1. ■

Remark Using the same algebra in the proof of Theorem 2.1, it is possible to prove that

$$\mathbb{E}_\theta \sup_{\lambda\in\Lambda^\star}\{\|\hat{\theta}_\lambda - \theta\|^2 - W(\theta,\lambda)\} \leq 0, \tag{31}$$

where

$$W(\theta,\lambda) = (1+B\rho_\epsilon)\left\{ \sum_{j=1}^{J}[(1-\lambda_{K_j})^2\|\theta\|^2_{(j)} + \lambda_{K_j}^2\sigma_j^2 + \lambda_{K_j}^2\mathrm{pen}_j] + \sum_{k>N}\theta_k^2 \right\}$$
$$+ C_1\epsilon^2 + B\rho_\epsilon R(\theta,\lambda^0).$$

Hence, $W(\theta,\lambda)$ is also a risk hull. For all $j \in \{1,\dots,J\}$, the only difference with $V(\theta,\lambda)$ is contained in the bound of

$$\sup_{\lambda_j\in[0,1]}\{\lambda_j^2\eta_j - \lambda_j^2\mathrm{pen}_j\} \leq [\eta_j - \mathrm{pen}_j]_+.$$

Then, we use inequality (11) in order to obtain Equation (31).

Proof of Theorem 3.1 In the situation where inequality (11) holds, Equation (31) yields

$$\mathbb{E}_\theta\|\theta^\star-\theta\|^2 \le W(\theta,\lambda^\star) = (1+B\rho_\epsilon)\bar{R}_{\text{pen}}(\theta,\lambda^\star) + B\rho_\epsilon R(\theta,\lambda^0) + C_1\epsilon^2, \tag{32}$$

where

$$\bar{R}_{\text{pen}}(\theta,\lambda^\star) = \sum_{j=1}^{J}[(1-\lambda_j^\star)^2\|\theta\|^2_{(j)} + (\lambda_j^\star)^2\sigma_j^2 + (\lambda_j^\star)^2\text{pen}_j] + \sum_{k>N}\theta_k^2,$$

and B denotes a positive constant independent of ϵ. Moreover, from Equation (14),

$$U_{\text{pen}}(y,\lambda^\star) \le U_{\text{pen}}(y,\lambda), \quad \forall\lambda\in\Lambda^\star.$$

The proof of Theorem 3.1 is mainly based on these two equalities. First, remark that

$$\begin{aligned}
U_{\text{pen}}(y,\lambda^\star) - \bar{R}_{\text{pen}}(\theta,\lambda^\star) &= \sum_{j=1}^{J}[((\lambda_j^\star)^2 - 2\lambda_j^\star)(\|\tilde{y}\|^2_{(j)} - \sigma_j^2) + (\lambda_j^\star)^2\sigma_j^2 + 2\lambda_j^\star\text{pen}_j \\
&\quad - (1-\lambda_j^\star)^2\|\theta\|^2_{(j)} - (\lambda_j^\star)^2\sigma_j^2 - (\lambda_j^\star)^2\text{pen}_j] - \sum_{k>N}\theta_k^2, \\
&= \sum_{j=1}^{J}[\{(\lambda_j^\star)^2 - 2\lambda_j^\star\}(\|\tilde{y}\|^2_{(j)} - \sigma_j^2) - (1-\lambda_j^\star)^2\|\theta\|^2_{(j)} \\
&\quad + \{2\lambda_j^\star - (\lambda_j^\star)^2\}\text{pen}_j] - \sum_{k>N}\theta_k^2.
\end{aligned}$$

Hence,

$$\begin{aligned}
U_{\text{pen}}(y,\lambda^\star) - \bar{R}_{\text{pen}}(\theta,\lambda^\star) &= \sum_{j=1}^{J}\Bigg[\{(\lambda_j^\star)^2 - 2\lambda_j^\star\}\sum_{k\in I_j}(\theta_k^2 + \epsilon^2 b_k^{-2}(\xi_k^2-1) + 2\epsilon b_k^{-1}\xi_k\theta_k) \\
&\quad - (1-\lambda_j^\star)^2\|\theta\|^2_{(j)} + \{2\lambda_j^\star - (\lambda_j^\star)^2\}\text{pen}_j\Bigg] - \sum_{k>N}\theta_k^2, \\
&= \sum_{j=1}^{J}\{(\lambda_j^\star)^2 - 2\lambda_j^\star\}(\eta_j + 2X_j - \text{pen}_j) - \|\theta\|^2,
\end{aligned}$$

where η_j and X_j are respectively defined in Equations (9) and (24). Hence, from Equation (14)

$$\begin{aligned}
\bar{R}_{\text{pen}}(\theta,\lambda^\star) &= U_{\text{pen}}(y,\lambda^\star) + \|\theta\|^2 + \sum_{j=1}^{J}\{2\lambda_j^\star - (\lambda_j^\star)^2\}(\eta_j + 2X_j - \text{pen}_j), \\
&\le U_{\text{pen}}(y,\lambda^p) + \|\theta\|^2 + \sum_{j=1}^{J}\{2\lambda_j^\star - (\lambda_j^\star)^2\}(\eta_j + 2X_j - \text{pen}_j),
\end{aligned}$$

where

$$\lambda^p = \arg\inf_{\lambda\in\Lambda^\star} R_{\text{pen}}(\theta,\lambda)$$

and $R_{\text{pen}}(\theta, \lambda)$ is defined in Equation (13). Then, with simple algebra

$$\begin{aligned}\mathbb{E}_\theta U_{\text{pen}}(y, \lambda^p) &= \mathbb{E}_\theta \sum_{j=1}^{J} [\{(\lambda_j^p)^2 - 2\lambda_j^p\}(\|\tilde{y}\|_{(j)}^2 - \sigma_j^2) + (\lambda_j^p)^2\sigma_j^2 + 2\lambda_j^p \text{pen}_j],\\ &= R_{\text{pen}}(\theta, \lambda^p) - \|\theta\|^2.\end{aligned}$$

This leads to

$$\mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) \leq R_{\text{pen}}(\theta, \lambda^p) + \mathbb{E}_\theta \sum_{j=1}^{J} \{2\lambda_j - (\lambda_j^\star)^2\}(\eta_j + 2X_j - \text{pen}_j). \tag{33}$$

We are now interested in the behaviour of the right-hand side of Equation (33). First, using Equations (25)–(29) in the proof of Theorem 2.1,

$$\mathbb{E}_\theta\{2\lambda_j - (\lambda_j^\star)^2\}X_j \leq C\rho_\epsilon\{(1-\lambda_j^p)^2\|\theta\|_{(j)}^2 + (\lambda_j^p)^2\sigma_j^2\} + \bar{C}\rho_\epsilon \mathbb{E}_\theta\{(1-\lambda_j^\star)^2\|\theta\|_{(j)}^2 + (\lambda_j^\star)^2\sigma_j^2\},$$

for all $j \in \{1, \dots, J\}$. Here, C and $\bar{C}$ denote positive constants independent of ϵ. In particular, it is always possible to obtain $\bar{C}$ verifying $\bar{C}\rho_\epsilon < 1$ (see the proof of Theorem 2.1 for more details). Hence,

$$\mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) \leq (1 + C\rho_\epsilon) R_{\text{pen}}(\theta, \lambda^p) + \bar{C}\rho_\epsilon \mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) + \mathbb{E}_\theta \sum_{j=1}^{J} \{2\lambda_j^\star - (\lambda_j^\star)^2\}(\eta_j - \text{pen}_j).$$

Then, from inequalities (11) and (30),

$$\mathbb{E}_\theta \sum_{j=1}^{J} \{2\lambda_j^\star - (\lambda_j^\star)^2\}(\eta_j - \text{pen}_j) = \mathbb{E}_\theta \sum_{j=1}^{J} [\eta_j - \text{pen}_j]_+ \leq C_1\epsilon^2.$$

This leads to

$$\begin{aligned}\mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) &\leq (1 + C\rho_\epsilon) R_{\text{pen}}(\theta, \lambda^p) + \bar{C}\rho_\epsilon \mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) + C_1\epsilon^2,\\ &\Rightarrow (1 - \bar{C}\rho_\epsilon)\mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) \leq (1 + C\rho_\epsilon) R_{\text{pen}}(\theta, \lambda^p) + C_1\epsilon^2,\\ &\Rightarrow \mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) \leq \frac{(1 + C\rho_\epsilon)}{(1 - \bar{C}\rho_\epsilon)} R_{\text{pen}}(\theta, \lambda^p) + C\epsilon^2.\end{aligned} \tag{34}$$

Using Equations (32) and (34),

$$\begin{aligned}\mathbb{E}_\theta\|\theta^\star - \theta\|^2 &\leq (1 + B\rho_\epsilon)\mathbb{E}_\theta \bar{R}_{\text{pen}}(\theta, \lambda^\star) + C_1\epsilon^2 + B\rho_\epsilon R(\theta, \lambda^0),\\ &\leq (1 + \mu_\epsilon) R_{\text{pen}}(\theta, \lambda^p) + C\epsilon^2 + B\rho_\epsilon R(\theta, \lambda^0),\end{aligned}$$

where $\mu_\epsilon = \mu_\epsilon(\rho_\epsilon)$ such that $\mu_\epsilon \to 0$ as $\rho_\epsilon \to 0$ and C is a positive constant independent of ϵ. In order to conclude the proof, we just have to compare $R(\theta, \lambda^0)$ with $R_{\text{pen}}(\theta, \lambda_p)$. For all $j \in \{1, \dots, J\}$, introduce

$$R_{\text{pen}}^j(\theta, \lambda) = (1-\lambda_j)^2\|\theta\|_{(j)}^2 + \lambda_j^2\sigma_j^2 + 2\lambda_j \text{pen}_j, \quad \text{and} \quad R^j(\theta, \lambda) = (1-\lambda_j)^2\|\theta\|_{(j)}^2 + \lambda_j^2\sigma_j^2.$$

Then,

$$R^j_{\text{pen}}(\theta, \lambda^p) \leq \frac{\sigma_j^4 \|\theta\|^2_{(j)}}{(\sigma_j^2 + \|\theta\|^2_{(j)})^2} + \frac{\sigma_j^2 \|\theta\|^4_{(j)}}{(\sigma_j^2 + \|\theta\|^2_{(j)})^2} + 2\frac{\text{pen}_\text{j}}{\sigma_j^2} \frac{\sigma_j^2 \|\theta\|^2_{(j)}}{\sigma_j^2 + \|\theta\|^2_{(j)}},$$
$$= \left(1 + 2\frac{\text{pen}_\text{j}}{\sigma_j^2}\right) R^j(\theta, \lambda^0),$$

since $R^j_{\text{pen}}(\theta, \lambda^p) \leq R^j_{\text{pen}}(\theta, \lambda^0)$ from the definition of λ^p. This concludes the proof of Theorem 3.1. ■

Acknowledgements

The author would like to thank both an associate editor and two anonymous referees whose constructive comments and remarks considerably improved the paper.

References

Barron, A., Birgé, L., and Massart, P. (1999), 'Risk Bounds for Model Selection via Penalization', *Probability Theory and Related Fields*, 113, 301–413.

Bissantz, N., Hohage, T., Munk, A., and Ryumgaart, F. (2007), 'Convergence Rates of General Regularization Methods for Statistical Inverse Problems and Applications', *SIAM Journal on Numerical Analysis*, 45, 2610–2636.

Candès, E. (2006), 'Modern Statistical Estimation via Oracle Inequalities', *Acta Numerica*, 15, 257–325.

Cao, Y., and Golubev, Y. (2006), 'On Adaptive Regression by Splines', *Mathematical Methods of Statistics*, 15, 398–414.

Cavalier, L., and Golubev, Y. (2006), 'Risk Hull Method and Regularization by Projections of Ill-Posed Inverse Problems', *The Annals of Statistics*, 34, 1653–1677.

Cavalier, L., and Tsybakov, A.B. (2001), 'Penalized Blockwise Stein's Method, Monotone Oracles and Sharp Adaptive Estimation', *Mathematical Methods of Statistics*, 3, 247–282.

Cavalier, L., and Tsybakov, A.B. (2002), 'Sharp Adaptation for Inverse Problems With Random Noise', *Probability Theory and Related Fields*, 123, 323–354.

Cavalier, L., Golubev, Y., Picard, D., and Tsybakov, A.B. (2002), 'Oracle Inequalities for Inverse Problems', *The Annals of Statistics*, 30, 843–874.

Donoho, D.L. (1995), 'Nonlinear Solutions of Linear Inverse Problems by Wavelet-Vaguelette Decomposition', *Journal of Applied and Computational Harmonic Analysis*, 2, 101–126.

Efromovich, S. (2007), 'A Lower Bound Oracle Inequality for a Blockwise-Shrinkage Estimate', *Journal of Statistical Planning and Inference*, 137, 176–183.

Engl, H.W., Hank, M., and Neubauer, A. (1996), *Regularization of Inverse Problems*, Dordrecht: Kluwer Academic Publishers Group.

Ermakov, M.S. (1989), 'Minimax Estimation of the Solution of an Ill-Posed Convolution Type Problem', *Problems of Information Transmission*, 25, 191–200.

Fan, J. (1991), 'On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems', *The Annals of Statistics*, 19, 1257–1272.

Golubev, Y. (2007), 'On Oracle Inequalities Related to High Dimensional Linear Models', in *Topics in Stochastic Analysis and Nonparametric Estimation*, IMA Proceedings, eds. P.L. Chow, B. Mordukhovich, and G. Yin, New York: Springer-Verlag, pp. 105–122.

Hida, T. (1980), *Brownian Motion*, New York: Springer-Verlag.

Johnstone, I.M., and Silverman, B.W. (1990), 'Speed of Estimation in Positron Emission Tomography and Related Inverse Problems', *The Annals of Statistics*, 18, 251–280.

Kneip, A. (1994), 'Ordered Linear Smoothers', *The Annals of Statistics*, 22, 835–866.

Marteau, C. (in press), 'Risk Hull Method for Spectral Regularization in Linear Statistical Inverse Problems', *ESAIM Probability and Statistics*, DOI: 10.1051/PS/2009011.