Adaptive spectral regularizations
of high dimensional linear models[*]

Yuri Golubev [†]

Centre de Mathématiques et Informatique
39 rue F. Joliot-Curie
13453 Marseille
France
e-mail: golubev.yuri@gmail.com

Abstract: This paper focuses on recovering an unknown vector β from the noisy data $Y = X\beta + \sigma\xi$, where X is a known $n \times p$-matrix, ξ is a standard white Gaussian noise, and σ is an unknown noise level. In order to estimate β, a spectral regularization method is used, and our goal is to choose its regularization parameter with the help of the data Y. In this paper, we deal solely with regularization methods based on the so-called ordered smoothers (see [13]) and extend the oracle inequality from [11] to the case, where the noise level is unknown.

1. Introduction and main results

This paper deals with recovering an unknown vector $\beta \in \mathbb{R}^n$ from the noisy data

$$Y = X\beta + \sigma\xi,$$

where X is a known $n \times p$-matrix with $n \geq p$, $\xi = \left(\xi(1), \ldots, \xi(n)\right)^\top$ is a standard white Gaussian noise ($\mathbf{E}\xi(k) = 0$, $\mathbf{E}\xi^2(k) = 1$, $k = 1, \ldots, n$), and σ is an unknown noise level.

In spite of its simplicity, this mathematical model plays an important role in solving practical inverse problems like gravity problems (see, e.g. [3]), tomography inverse problems [12], and many others. As a rule, in inverse problems n and p are very large and therefore the main goal in this paper is to propose an approach suitable for $n = \infty$, $p = \infty$, severely ill-posed matrices $X^\top X$, and the unknown noise level.

We begin with the standard maximum likelihood estimate

$$\hat{\beta}_0 = \arg\min_{\beta \in \mathbb{R}^p} \|Y - X\beta\|^2 = (X^\top X)^{-1} X^\top Y,$$

where $\|z\|^2 = \sum_{k=1}^n z^2(k)$. It is well known and easy to check that

$$\mathbf{E}(\hat{\beta}_0 - \beta)(\hat{\beta}_0 - \beta)^\top = \sigma^2 (X^\top X)^{-1}$$

and thus, the mean square risk of $\hat{\beta}_0$ is computed as follows:

$$\mathbf{E}\|\hat{\beta}_0 - \beta\|^2 = \sigma^2 \text{trace}\Big[(X^\top X)^{-1}\Big] = \sigma^2 \sum_{k=1}^p \lambda^{-1}(k), \tag{1}$$

where $\lambda(k)$ and $\psi_k \in \mathbb{R}^n$ here and below are the eigenvalues and the eigenvectors of $X^\top X$

$$X^\top X \psi_k = \lambda(k)\psi_k.$$

In this paper, it is assumed solely that $\lambda(1) \geq \lambda(2) \geq \cdots \geq \lambda(p)$. This assumption together with (1) reveals the main difficulty in $\hat{\beta}_0$: *its risk may be very large when p is large or when $X^\top X$ has a large condition number.*

The natural idea to improve $\hat{\beta}_0$ is to suppress large $\lambda^{-1}(k)$ in (1) with the help of a linear smoother. Therefore we make use of the following family of linear estimates

$$\hat{\beta}_\alpha = H_\alpha \hat{\beta}_0, \ \alpha \in (0, \alpha^\circ], \tag{2}$$

where H_α, $\alpha \in (0, \alpha^\circ]$ is a family of $p \times p$-smoothing matrices.

In what follows, we deal with the smoothing matrices admitting the following representation

$$H_\alpha = \sum_{k=1}^p \mathcal{H}_\alpha[\lambda(k)]\psi_k \psi_k^\top,$$

where $\mathcal{H}_\alpha(\lambda): \ \mathbb{R}^+ \to [0, 1]$ is such that

$$\lim_{\alpha \to 0} \mathcal{H}_\alpha(\lambda) = 1, \quad \lim_{\lambda \to 0} \mathcal{H}_\alpha(\lambda) = 0.$$

In the literature (see, e.g., [9]), this method is called *spectral regularization*. It covers widely used regularizations methods such as the Tikhonov-Phillips regularization [20] known in the statistical literature as *ridge regression*, Landweber's iterations [14], the μ-method (see e.g. [9]), and many others.

Summarizing, β is estimated with the help of the family of linear estimates $\hat{\beta}_\alpha$, $\alpha \in (0, \alpha^\circ]$ defined by (2) and our goal is to find based on the data at hand the best estimator within this family. Notice that for given α, the mean square risk of $\hat{\beta}_\alpha$ is computed as follows:

$$L_\alpha(\beta) \overset{\text{def}}{=} \mathbf{E}\|\hat{\beta}_\alpha - \beta\|^2 = \sum_{k=1}^p \big[1 - h_\alpha(k)\big]^2 \langle \beta, \psi_k \rangle^2 + \sigma^2 \sum_{k=1}^p \lambda^{-1}(k) h_\alpha^2(k), \tag{3}$$

where

$$h_\alpha(k) \stackrel{\text{def}}{=} \mathcal{H}_\alpha[\lambda(k)] \quad \text{and} \quad \langle \beta, \psi_k \rangle \stackrel{\text{def}}{=} \sum_{l=1}^{p} \beta(l)\psi_k(l).$$

It is easily seen from (3) that the variance of $\hat{\beta}_\alpha$ is always smaller than that one of $\hat{\beta}_0$, but $\hat{\beta}_\alpha$ has a non-zero bias and therefore adjusting α we may improve the risk of $\hat{\beta}_0$. This improvement may be very significant when $\langle \beta, \psi_k \rangle^2$ are small for large k.

In practice, a good choice of the regularizing matrix family H_α is a delicate problem related to the computational complexity of $\hat{\beta}_\alpha$. For details, we refer interested readers to [9].

As a rule, practical spectral regularization methods (the spectral cut–off, the Tikhonov-Phillips regularization, Landweber's iterations) represent the so-called *ordered smoothers* [13]. This means that the family of functions $\{\mathcal{H}_\alpha(\lambda), \alpha \in (0, \alpha^\circ]\}$ is ordered in the following sense:

Definition 1. The family of functions $\{F_\alpha(\lambda), \alpha \in A, \lambda \in \Lambda \subseteq \mathbb{R}^+\}$ is ordered if:

1. For any given $\alpha \in A$, $F_\alpha(\lambda) : \Lambda \to [0, 1]$ is a monotone function of λ.
2. If for some $\alpha_1, \alpha_2 \in A$ and some $\lambda' \in \Lambda$, $F_{\alpha_1}(\lambda') < F_{\alpha_2}(\lambda')$, then for all $\lambda \in \Lambda$, $F_{\alpha_1}(\lambda) \leq F_{\alpha_2}(\lambda)$.

The next important question usually arising in practice is related to the data-driven choice of the regularization parameter α. In statistical literature, one can find several general approaches to this problem. We cite here, for instance, the Lepski method which has been adopted to inverse problems in [17], [2], [5], and the model selection technique which was used in [15].

The approach proposed in this paper is a slight modification of the unbiased risk estimation. To make the presentation simpler, we begin with the case, where the noise level σ^2 is known. Intuitively, a good data-driven regularization parameter should minimize in some sense the risk $L_\alpha(\beta)$ (see (3)). Obviously, the best regularization parameter minimizing $L_\alpha(\beta)$ cannot be used since it depends on the unknown parameter of interest β. However, the idea of minimization of $L_\alpha(\beta)$ may be put into practice with the help of the empirical risk minimization principle defining the regularization parameter as follows:

$$\hat{\alpha} = \arg\min_\alpha R_\alpha^\sigma[Y, Pen], \tag{4}$$

where

$$R_\alpha^\sigma[Y, Pen] \stackrel{\text{def}}{=} \|\hat{\beta}_0 - \hat{\beta}_\alpha\|^2 + \sigma^2 Pen(\alpha),$$

and $Pen(\alpha) : (0, \alpha^\circ] \to \mathbb{R}^+$ is a given function called *penalty*. The main idea in this approach is to link $L_\alpha(\beta)$ and $R_\alpha^\sigma[Y, Pen]$. Heuristically, we want to find a minimal penalty $Pen(\alpha)$ that ensures the following inequality

$$L_\alpha(\beta) \lesssim R_\alpha^\sigma[Y, Pen] + \mathcal{C}, \tag{5}$$

where $\mathcal{C}$ is a random variable that doesn't depend on α. It is convenient to define this constant as follows:

$$\mathcal{C} = -\|\beta - \hat{\beta}_0\|^2 = -\sigma^2 \sum_{k=1}^{p} \lambda^{-1}(k)\xi^2(k).$$

The traditional approach to solving (5) is based on the minimization of the unbiased risk estimate. In this method, the penalty is computed as a root of the equation

$$L_\alpha(\beta) = \mathbf{E}\Big\{R_\alpha^\sigma[Y, Pen_u] + \mathcal{C}\Big\}. \tag{6}$$

One can check with a simple algebra that

$$Pen_u(\alpha) = 2\sum_{k=1}^{p} \lambda^{-1}(k)h_\alpha(k).$$

The idea of this penalty goes back to [1] and [7] provides some oracle inequalities related to this approach.

Another well-known and widely used approach to the data-driven choice of α is related to the cross validation technique [8]. In the framework of our statistical model, this method prompts a data-driven regularization parameter which is close to

$$\hat{\alpha}_{CV} = \arg\min_\alpha\Big\{\|Y - X\hat{\beta}_\alpha\|^2 + \sigma^2 Pen_{CV}(\alpha)\Big\},$$

with

$$Pen_{CV}(\alpha) = 2\sum_{k=1}^{p} h_\alpha(k).$$

It is well-known (see e.g. [13]) that if the risk of $\hat{\beta}$ is measured by $\mathbf{E}\|X\hat{\beta} - X\beta\|^2$, then this penalty is nearly optimal and it works always well.

However, the question *Does $\hat{\alpha}_{CV}$ works well when the risk is measured by* $\mathbf{E}\|\hat{\beta} - \beta\|^2$ *?* has a delicate answer depending on the spectrum of $X^\top X$. To the best of our knowledge there are no oracle inequalities controlling the risk of $\hat{\beta}\hat{\alpha}_{CV}$ uniformly in β. Notice, however, that one can show with the help of the method for computing minimal penalties in [4], that if $\lambda(k) \leq \exp(-\kappa k)$, then the risk of this method blows up starting from some $\kappa > 0$.

The similar effect takes place in the unbiased risk estimation. This happens because the standard deviation of $R_\alpha^\sigma[Y, Pen_u] + \mathcal{C}$ may be very large with respect to the mean $\mathbf{E}\big\{R_\alpha^\sigma[Y, Pen_u] + \mathcal{C}\big\}$ and therefore (5) may fail with a high probability.

To improve the above mentioned drawbacks of the unbiased risk estimation, we define, following [11], the penalty as a minimal root of the equation

$$\mathbf{E} \sup_{\alpha \leq \alpha^\circ} \Big[L_\alpha(\beta) - R_\alpha^\sigma[Y, Pen] - \mathcal{C}\Big]_+ \leq C_1 \mathbf{E}\Big[L_{\alpha^\circ}(\beta) - R_{\alpha^\circ}^\sigma[Y, Pen] - \mathcal{C}\Big]_+, \tag{7}$$

where $[x]_+ = \max\{0, x\}$ and $C_1 > 1$ is a constant. Heuristic motivation behind this approach is rather transparent. We are looking for the minimal penalty that

balances the excess risks corresponding to all possible $\alpha \in (0, \alpha^\circ]$. Recall that the excess risk is defined by the difference between the risk of the estimate and its penalized empirical risk. Note that in view of (5), we can deal solely with the positive part of the excess risk.

In order to explain heuristically how Equation (7) may be solved, we begin with the spectral representation of the underlying statistical problem. One can check easily that

$$y(k) \overset{\text{def}}{=} \langle X^\top Y, \psi_k \rangle / \lambda(k) = \langle \beta, \psi_k \rangle + \sigma \xi'(k) / \sqrt{\lambda(k)},$$

where $\xi'(k)$ are i.i.d. $\mathcal{N}(0,1)$. With these notations, $\hat{\beta}_\alpha$ admits the following representation

$$\langle \hat{\beta}_\alpha, \psi_k \rangle = h_\alpha(k) y(k) = h_\alpha(k) \beta(k) + \sigma h_\alpha(k) \xi'(k) / \sqrt{\lambda(k)},$$

where $\beta(k) = \langle \beta, \psi_k \rangle$, and therefore

$$\begin{aligned}
\|\hat{\beta}_0 - \hat{\beta}_\alpha\|^2 &= \sum_{k=1}^{p} \big[1 - h_\alpha(k)\big]^2 y^2(k), \\
\|\beta - \hat{\beta}_\alpha\|^2 &= \sum_{k=1}^{p} \big[\beta(k) - h_\alpha(k) y(k)\big]^2.
\end{aligned} \tag{8}$$

In what follows, it is assumed that the penalty has the following structure

$$Pen(\alpha) = Pen_u(\alpha) + (1 + \gamma) Q(\alpha),$$

where γ is a small positive number and $Q(\alpha)$, $\alpha > 0$ is a positive function of α to be defined later on. Recall that the first term at the right-hand side is obtained from the unbiased risk estimation (see Equation (6)). With $Pen(\alpha)$ we can rewrite the excess risk as follows:

$$\begin{aligned}
&L_\alpha(\beta) - R_\alpha^\sigma[Y, Pen] - \mathcal{C} \\
&= \sigma^2 \sum_{k=1}^{p} \lambda^{-1}(k) \big[2h_\alpha(k) - h_\alpha^2(k)\big] (\xi'^2(k) - 1) - (1 + \gamma) \sigma^2 Q(\alpha) \\
&\quad + 2\sigma \sum_{k=1}^{p} \lambda^{-1/2}(k) \big[2h_\alpha(k) - h_\alpha^2(k)\big] \xi'(k) \beta(k).
\end{aligned} \tag{9}$$

The first idea in solving (7) is based on the the fact that the cross term

$$2\sigma \sum_{k=1}^{p} \lambda^{-1/2}(k) \big[2h_\alpha(k) - h_\alpha^2(k)\big] \xi'(k) \beta(k)$$

is typically small with respect to $\mathbf{E}\big\{R_\alpha^\sigma[Y, Pen] + \mathcal{C}\big\}$ (see for more details Lemma 9 in [11]). With this in mind, omitting the cross term, Equation (7) can be rewritten in the following nearly equivalent form

$$\mathbf{E} \sup_{\alpha \le \alpha^\circ} [\eta_\alpha - (1 + \gamma) Q(\alpha)]_+ \lesssim C_1 \mathbf{E} [\eta_{\alpha^\circ} - (1 + \gamma) Q(\alpha^\circ)]_+ \asymp D(\alpha^\circ), \tag{10}$$

where

$$\eta_\alpha \overset{\text{def}}{=} \sum_{k=1}^{p} \lambda^{-1}(k)\big[2h_\alpha(k) - h_\alpha^2(k)\big]\big[\xi'^2(k) - 1\big]$$

and

$$D(\alpha) \overset{\text{def}}{=} \sqrt{\mathbf{E}\eta_\alpha^2} = \left\{2\sum_{k=1}^{p} \lambda^{-2}(k)\big[2h_\alpha(k) - h_\alpha^2(k)\big]^2\right\}^{1/2}.$$

Now we are in a position to compute an approximation of the minimal root for (10). It is clear that $Q(\alpha) \geq Q^+(\alpha)$, where $Q^+(\alpha)$ is a root of

$$\mathbf{E}[\eta_\alpha - Q^+(\alpha)]_+ = D(\alpha^\circ). \tag{11}$$

To find a feasible solution to (11), we make use of the exponential Chebychev inequality resulting in

$$\mathbf{E}[\eta - x]_+^p \leq \Gamma(p+1)\lambda^{-p}\exp(-\lambda x)\mathbf{E}\exp(\lambda\eta), \tag{12}$$

where η is a random variable, $\Gamma(\cdot)$ is the gamma function, and $\lambda > 0$.

Therefore we define $Q^+(\alpha)$ as a root of equation

$$\inf_\lambda \exp[-\lambda Q^+(\alpha)]\mathbf{E}\exp(\lambda\eta_\alpha) = D(\alpha^\circ).$$

It is easy to check with a simple algebra that

$$Q^+(\alpha) = 2D(\alpha)\mu_\alpha \sum_{k=1}^{p} \frac{\rho_\alpha^2(k)}{1 - 2\mu_\alpha\rho_\alpha(k)}, \tag{13}$$

where μ_α is a root of the equation

$$\sum_{k=1}^{p} F[\mu_\alpha\rho_\alpha(k)] = \log\frac{D(\alpha)}{D(\alpha^\circ)}, \tag{14}$$

and

$$\begin{aligned}
F(x) &= \frac{1}{2}\log(1 - 2x) + x + \frac{2x^2}{1 - 2x}, \\
\rho_\alpha(k) &= \sqrt{2}D^{-1}(\alpha)\lambda^{-1}(k)\big[2h_\alpha(k) - h_\alpha^2(k)\big].
\end{aligned} \tag{15}$$

The next result (see also Theorem 1 in [11]) shows that $Q^+(\alpha)$ is a nearly optimal solution to (10).

Proposition 1. *For any $\gamma > q \geq 0$*

$$\mathbf{E}\sup_{\alpha \leq \alpha^\circ}\Big[\eta_\alpha - (1+\gamma)Q^+(\alpha)\Big]_+^{1+q} \leq \frac{CD^{1+q}(\alpha^\circ)}{(\gamma - q)^3},$$

where here and throughout the paper C denotes a generic constant.

Let us now turn to the case, where σ is unknown. To compute the data-driven regularization parameter in this situation, we replace σ^2 in $R_\alpha^\sigma[Y, Pen]$ by the standard variance estimator

$$\hat{\sigma}_\alpha^2 = \frac{\|Y - X\hat{\beta}_\alpha\|^2}{\|1 - H_\alpha\|^2}.$$

Thus we arrive at the following approximation of the empirical risk

$$R_\alpha[Y, Pen] \stackrel{\text{def}}{=} \|\hat{\beta}_0 - \hat{\beta}_\alpha\|^2 + \frac{\|Y - X\hat{\beta}_\alpha\|^2}{\|1 - H_\alpha\|^2} Pen(\alpha)$$

and the data-driven regularization parameter is computed now as follows:

$$\hat{\alpha} = \underset{\alpha_\circ \leq \alpha \leq \alpha^\circ}{\arg\min}\ R_\alpha[Y, Pen].$$

Notice that in contrast to the case of known σ, it assumed that α is bounded from below by $\alpha_\circ$. This constraint ensures that with a hight probability $[\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2]_+ < \sigma^2/2$ uniformly in $\beta \in \mathbb{R}^p$. Unfortunately, when this inequality fails we cannot control correctly the risk of $\hat{\beta}_{\hat{\alpha}}$ since it may blow up (see [4] for similar phenomenon in the model selection). So, to avoid the blowup, we need a relatively good estimate of σ, or equivalently, large $\|1 - H_\alpha\|^2$.

Stress also that since $\alpha_\circ$ cannot depend on σ, we would like to have $\alpha_\circ$ as small as possible to be sure that the methods works for small noise levels. From a mathematical viewpoint, this means that we need a relatively good upper bound for $\mathbf{E}|\sigma - \hat{\sigma}_{\hat{\alpha}}^2|Pen(\hat{\alpha})$. Roughly speaking, we have to check that with a hight probability

$$|\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2|Pen(\hat{\alpha}) \ll \sigma^2 Pen(\hat{\alpha}).$$

The main difficulty in proving this equation is related to the fact that the random variables $\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2$ and $Pen(\hat{\alpha})$ are dependent. To overcome this difficulty we make use of the law of the iterated logarithm for $\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2$ combined with a generalization of the Hölder inequality (see Lemmas 4 and 3 below). To carry out this approach, we need the following additional condition: *there exists a positive constant C_2 such that for all $\alpha \in (0, \alpha^\circ]$*

$$\|h_\alpha\|_\lambda^2 \geq C_2 \sum_{k=1}^p \lambda^{-1}(k)h_\alpha(k), \tag{16}$$

$$\frac{\|h_\alpha\|_\lambda^2}{\log[D(\alpha)/D(\alpha^\circ)]} + \max_k \frac{h_\alpha(k)}{\lambda(k)} \geq C_2 D(\alpha), \tag{17}$$

where

$$\|h_\alpha\|_\lambda^2 = \sum_{k=1}^p \lambda^{-1}(k)h_\alpha^2(k).$$

Denote for brevity

$$\mathcal{O}(\alpha_\circ, \alpha^\circ) \stackrel{\text{def}}{=} \frac{1}{\|1 - h_{\alpha_\circ}\|} \Bigg\{ \left[\log\log\left(1 + \frac{\|1 - h_{\alpha_\circ}\|^2}{\|1 - h_{\alpha^\circ}\|^2} \right) \right]^{1/2} \\ + \log\left(1 + \frac{Pen(\alpha_\circ)}{Pen(\alpha^\circ)} \right) \Bigg\}.$$

The following theorem controls the risk of $\hat{\beta}_{\hat{\alpha}}$ via the penalized oracle risk defined by

$$r(\beta) \stackrel{\text{def}}{=} \inf_{\alpha_\circ \leq \alpha \leq \alpha^\circ} \bar{R}_\alpha(\beta),$$

where

$$\bar{R}_\alpha(\beta) \stackrel{\text{def}}{=} \mathbf{E}_\beta\{ R_\alpha[Y, Pen] + \mathcal{C} \} = L_\alpha(\beta) + (1+\gamma)\sigma^2 Q^+(\alpha) \\ + \frac{Pen(\alpha)}{\|1 - h_\alpha\|^2} \sum_{k=1}^{p} \left[1 - h_\alpha(k) \right]^2 \lambda(k)\beta^2(k).$$

Theorem 1. *Let $Pen(\alpha) = 2\sum_{k=1}^{p} \lambda^{-1}(k)h_\alpha(k) + (1+\gamma)Q^+(\alpha)$ with $Q^+(\alpha)$ defined by (13–15) and suppose (16–17) hold. Then, uniformly in $\beta \in \mathbb{R}^p$,*

$$\mathbf{E}_\beta\|\beta - \hat{\beta}_{\hat{\alpha}}\|^2 \leq \left[1 + C\mathcal{O}(\alpha_\circ, \alpha^\circ) + C\log^{-1/2}\frac{r(\beta)}{\sigma^2 D(\alpha^\circ)} \right] r(\beta) \\ + \frac{C\sigma^2 D(\alpha^\circ)}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+ \sqrt{\gamma}} \mathcal{Q}\left[\frac{r(\beta)}{\sigma^2 \gamma D(\alpha^\circ)} + \frac{1}{\gamma^4} \right], \tag{18}$$

where $\mathcal{Q}(x) = x/\log(x)$.

Notice that Equation 18 can be rewritten in the following concise form

$$\mathbf{E}_\beta\|\beta - \hat{\beta}_{\hat{\alpha}}\|^2 \leq \left[1 + C\mathcal{O}(\alpha_\circ, \alpha^\circ) + \Psi_{\alpha_\circ, \gamma}\left(\frac{r(\beta)}{\sigma^2 D(\alpha^\circ)} \right) \right] r(\beta), \tag{19}$$

where $\Psi_{\alpha_\circ, \gamma}(\cdot)$ is a bounded function such that

$$\lim_{x \to \infty} \Psi_{\alpha_\circ, \gamma}(x) = 0. \tag{20}$$

The statistical sense of (19) is rather transparent: this equation shows that in typical nonparametric situations the method works like the ideal penalized oracle with the risk $r(\beta)$. *The typical nonparametric situation* means that

- p is large, so, for properly chosen $\alpha_\circ$, $\mathcal{O}(\alpha_\circ, \alpha^\circ)$ is small,
- the vector $(\langle \beta, \psi_1 \rangle, \ldots \langle \beta, \psi_p \rangle)^\top$ contains many significant components, and thus $r(\beta) \gg \sigma^2 D(\alpha^\circ)$.

These assumptions are typical in the minimax estimation, where it is assumed that β belongs to an ellipsoid. Notice that with the help of (19–20) one can check relatively easily that for a proper chosen spectral regularization, $\hat{\beta}_{\hat{\alpha}}$ is the asymptotically minimax estimate up to a constant (see for details [11] and [18]).

We finish this section with a short discussion of Conditions (16–17). Equation (16) means that $h_\alpha(k)$ vanishes rather rapidly for large k. This is always true for the spectral cut-off method $(h_\alpha(k) = \mathbf{1}\{\alpha\lambda(k) \geq 1\})$. Indeed, if $\lambda(k) \asymp k^{-p}$ with some $p \geq 0$, then

$$\|h_\alpha\|_\lambda^2 \asymp \alpha^{-p-1}, \quad D(\alpha) \asymp \alpha^{-p-1/2}$$

and it is seen easily that (17) is fulfilled. Assume now that $X^\top X$ is severely ill-posed, i.e., $\lambda(k) \asymp \exp(-\kappa k)$ with $\kappa > 0$. Then

$$\max_k \lambda(k) h_\alpha(k) \asymp \exp(\kappa/\alpha) \quad \text{and} \quad D(\alpha) \asymp \kappa^{-1/2} \exp(\kappa/\alpha).$$

Therefore (17) holds with $C_2 = \kappa^{-1/2}$.

2. Proofs

2.1. Ordered processes and their basic properties

The main results in this paper are based on a general fact which is similar to Dudley's entropy bound (see, e.g., [21]). Let ζ_t be a separable zero mean random process on $\mathbb{R}^+$. Denote for brevity

$$\Delta^\zeta(t_1, t_2) = \zeta_{t_1} - \zeta_{t_2}.$$

The following fact (see Lemma 1 in [11]) plays a cornerstone role in the proof of Proposition 1 and Theorem 1.

Proposition 2. *Let v_u^2, $u \in \mathbb{R}^+$, be a continuous strictly increasing function with $v_0^2 = 0$. Then for any $\lambda > 0$,*

$$\log \mathbf{E} \exp\left\{\lambda \max_{0 \leq s \leq t} \frac{\Delta^\zeta(s,t)}{\sigma_t}\right\} \leq \frac{\log(2)\sqrt{2}}{\sqrt{2}-1}$$
$$+ \max_{0 < s' < s \leq t} \max_{|z| \leq \sqrt{2}/(\sqrt{2}-1)} \log \mathbf{E} \exp\left\{z\lambda \frac{\Delta^\zeta(s',s)}{\bar{\Delta}^v(s',s)}\right\},$$

where $\bar{\Delta}^v(s',s) = \sqrt{|v_s^2 - v_{s'}^2|}$.

Definition 2. A zero mean process ζ_t, $t \in \mathbb{R}^+$ is called ordered if there exists a continuous strictly monotone function v_t^2, $t \in \mathbb{R}^+$ and some $\Lambda > 0$ such that

$$\sup_{s',s \in \mathbb{R}^+ : s' \neq s} \mathbf{E} \exp\left[\Lambda \frac{\Delta^\zeta(s',s)}{\bar{\Delta}^v(s',s)}\right] < \infty.$$

The next two propositions (see Lemmas 2 and 3 in [11]) show that the ordered process ζ_t can be controlled by the deterministic function v_t.

Proposition 3. *Let ζ_t be an ordered process with $\zeta_0 = 0$. Then there exists a constant $C(q', q)$ such that for all $1 < q', q \leq 2$, uniformly in $z > 0$*

$$\mathbf{E}\sup_{t \geq 0}\left[\zeta_t - zv_t^q\right]_+^{q'} \leq \frac{C(q', q)}{z^{q'/(q-1)}},$$

where $[x]_+ = \max(0, x)$.

Proposition 4. *Assume that there exists a monotone function $v_t, t \geq 0$ such that a random process ζ_t, $t \in \mathbb{R}^+$, satisfies*

$$\mathbf{E}\sup_{t \geq 0}\left[\zeta_t - zv_t^q\right]_+^{q'} \leq \frac{C}{z^{q'/(q-1)}},$$

for any $z > 0$ and some $q' \geq 1$, $q > 1$. Then there exists a constant C' such that for any random variable $\tau \in \mathbb{R}^+$ the following inequality holds

$$\left[\mathbf{E}|\zeta_\tau|^{q'}\right]^{1/q'} \leq C'\left[\mathbf{E}v_\tau^{qq'}\right]^{1/(qq')}.$$

In what follows, we focus on typical ordered processes related to the empirical risk. The following two propositions (see Lemmas 4 and 5 in [11]) are essential in controlling the cross term

$$\sigma \sum_{k=1}^{p} \lambda^{-1/2}(k)[2h_\alpha(k) - h_\alpha^2(k)]\xi'(k)\beta(k)$$

in the case, where α is a random variable depending on $\xi'(k)$, $k = 1, \ldots, p$.

Proposition 5. *For any given $\bar{\alpha} > 0$ and any $z > 0$,*

$$\mathbf{E}\sup_{0 \leq \alpha \leq \alpha^\circ}\left\{\sum_{k=1}^{p}\left[h_{\bar{\alpha}}(k) - h_\alpha(k)\right]b(k)\xi'(k)\right.$$
$$\left. - z\left[\sum_{k=1}^{p}\left[h_{\bar{\alpha}}(k) - h_\alpha(k)\right]^2b^2(k)\right]^{q'}\right\}_+^{q} \leq \frac{C}{z^{q/(2q'-1)}}, \quad q' > 1/2.$$

Proposition 6. *Let $\bar{\alpha}$ be a given smoothing parameter. Then for any $p \in [1, 2)$, there exists a constant $C(p)$ so that for any data-driven smoothing parameter $\hat{\alpha}$,*

$$\mathbf{E}\left|\sum_{k=1}^{p}\left[h_{\hat{\alpha}}(k) - h_{\bar{\alpha}}(k)\right]\lambda^{-1/2}(k)\beta(k)\xi'(k)\right|^p$$
$$\leq C(p)\left\{\mathbf{E}\max_k \lambda^{-1}(k)h_{\hat{\alpha}}^2(k)\right\}^{p/2}\left[\sum_{k=1}^{p}\left[1 - h_{\bar{\alpha}}(k)\right]^2\beta^2(k)\right]^{p/2}$$
$$+ C(p)\left\{\max_k \lambda^{-1}(k)h_{\bar{\alpha}}^2(k)\right\}^{p/2}\left[\mathbf{E}\sum_{k=1}^{p}\left[1 - h_{\hat{\alpha}}(k)\right]^2\beta^2(k)\right]^{p/2}.$$

In order to obtain oracle inequalities in the case, where the noise variance is unknown, we will need the following lemma generalizing Proposition 3.

Lemma 1. *Let*

$$\zeta_\alpha(b) = \sum_{k=1}^{p}[1 - h_\alpha(k)]\xi'(k)b(k), \ v_\alpha^2(b) = \sum_{k=1}^{p}[1 - h_\alpha(k)]^2 b^2(k), \ K = \frac{2}{(\sqrt{2}-1)^2}.$$

Then uniformly in $b \in \mathbb{R}^p$

$$\mathbf{E}\exp\Big\{\sup_{\alpha\in\mathbb{R}^+}\big[\zeta_\alpha(b) - Kv_\alpha^2(b)\big]\Big\} \le C.$$

Proof. Since $h_\alpha(\cdot), \alpha \ge 0$, is the family of ordered functions, it is not difficult to check that

$$\mathbf{E}[\zeta_{\alpha'}(b) - \zeta_\alpha(b)]^2 \le |v_{\alpha'}^2(b) - v_\alpha^2(b)|. \tag{21}$$

Indeed, we can rewrite (21) in the following equivalent form

$$\mathbf{E}\zeta_{\alpha'}(b)\zeta_\alpha(b) \ge \min\Big\{v_{\alpha'}^2(b), v_\alpha^2(b)\Big\}.$$

Assume for definiteness that $h_\alpha(k) \ge h_{\alpha'}(k)$, $k = 1, 2, \ldots, p$. Then $1 - h_\alpha(k) \le 1 - h_{\alpha'}(k)$, $k = 1, 2, \ldots, p$, and we get

$$\begin{aligned}
\mathbf{E}\zeta_{\alpha'}(b)\zeta_\alpha(b) &= \sum_{k=1}^{p}[1 - h_\alpha(k)][1 - h_{\alpha'}(k)]b^2(k) \\
&\ge \sum_{k=1}^{p}[1 - h_\alpha(k)]^2 b^2(k) = v_\alpha^2(b),
\end{aligned}$$

thus proving (21).

Since $\zeta_\alpha(b)$ is a Gaussian process, we obtain by (21)

$$\log\mathbf{E}\exp\left\{\lambda\frac{\zeta_{\alpha'}(b) - \zeta_\alpha(b)}{\sqrt{|v_{\alpha'}^2(b) - v_\alpha^2(b)|}}\right\} \le \frac{\lambda^2}{2}. \tag{22}$$

We may assume without loss of generality that σ_α is a continuous function in $\alpha \in \mathbb{R}^+$. Then let us fix some $\epsilon > 0$ and define $\alpha_l \in \mathbb{R}^+$ as roots of equations

$$v_{\alpha_l}^2(b) = (1 + \epsilon)^l, \ l \ge 0.$$

Since $v_\alpha^2(b) \le \sum_{k=1}^{p} b^2(k)$, the set of α_l is always finite but it may be empty.

Let α^* be a root of the equation $v_{\alpha^*}^2(b) = 1$. Then by Proposition 2 and (22) we obtain

$$\mathbf{E}\exp\Big\{\max_{\alpha\in\mathbb{R}^+}\big[\zeta_\alpha(b)-Kv_\alpha^2(b)\big]\Big\}$$

$$\leq \mathbf{E}\exp\Big\{\max_{\alpha>\alpha^*}\big[\zeta_\alpha(b)-Kv_\alpha^2(b)\big]\Big\}+\mathbf{E}\exp\Big\{\max_{\alpha\leq\alpha^*}\big[\zeta_\alpha(b)-Kv_\alpha^2(b)\big]\Big\}$$

$$\leq \mathbf{E}\exp\Big\{\max_{\alpha\leq\alpha^*}\zeta_\alpha(b)\Big\}+\sum_{l\geq0}\mathbf{E}\exp\Big\{\max_{\alpha_l<\alpha\leq\alpha_{l+1}}\big[\zeta_\alpha(b)-Kv_{\alpha_{l-1}}^2(b)\big]\Big\}$$

$$\leq C+\sum_{l\geq0}\mathbf{E}\exp\Big\{\max_{0<\alpha\leq\alpha_l}\Big[v_{\alpha_l}(b)\frac{\zeta_\alpha(b)}{v_{\alpha_l}(b)}-Kv_{\alpha_{l-1}}^2(b)\Big]\Big\}$$

$$\leq C+C\sum_{l\geq0}\exp\Big\{\Big[\frac{v_{\alpha_l}^2(b)}{(\sqrt{2}-1)^2}-Kv_{\alpha_{l-1}}^2(b)\Big]\Big\}$$

$$\leq C+C\sum_{l\geq0}\exp\Big\{\Big[(1+\epsilon)^l\Big(\frac{1}{(\sqrt{2}-1)^2}-\frac{K}{1+\epsilon}\Big)\Big]\Big\}$$

$$= C+C\sum_{l\geq0}\exp\Big\{-(1+\epsilon)^{l-1}\frac{1-\epsilon}{(\sqrt{2}-1)^2}\Big\}.$$

This equation with $\epsilon = 0.5$ completes the proof. $\qquad\square$

2.2. Recovering the noise variance

In this section, we focus on basic probabilistic properties of the variance estimator

$$\hat\sigma_{\hat\alpha}^2 = \frac{\|Y-X\hat\beta_{\hat\alpha}\|^2}{\|1-H_{\hat\alpha}\|^2},$$

in the case, where $\hat\alpha$ is a data-driven smoothing parameter. We begin with a simple auxiliary fact.

Lemma 2. *Let η' and η be nonnegative random variables. Then the following inequality*

$$\mathbf{E}\eta'\eta^q \leq \frac{2^{q-1}\lambda^q}{(2-q)^q}\mathbf{E}\eta'\log^q\Big\{1+\frac{\eta'}{\mathbf{E}\eta'}\Big\}$$
$$+\frac{2^{q-1}\lambda^q}{(2-q)^q}\mathbf{E}\eta'\log^q\Big\{1+\mathbf{E}\Big[\exp\Big(\frac{\eta}{\lambda}\Big)-\frac{\eta}{\lambda}-1\Big]\Big\}+q\lambda^q\mathbf{E}\eta'$$

(23)

holds for any $\lambda > 0$ and $q \in (1,2)$.

Proof. Consider the following function

$$F(z,y) \overset{\text{def}}{=} \max_{x\geq0}\Big\{x^q y - z[\exp(x)-1-x]\Big\}.$$

Differentiating $x^q y - z[\exp(x)-1-x]$ in x, it is easy to check that

$$F(z,y) = x_*^q y - z[\exp(x_*)-1-x_*] \leq x_*^q y,$$

where x_* is a root of the equation

$$x_* = \log\left(1 + \frac{qyx_*^{q-1}}{z}\right).$$

Since $\log(x)$ is convex, it is clear

$$\log\left(1 + \frac{qyx_*^{q-1}}{z}\right) \le \log\left(1 + \frac{qy}{z}\right) + \left(1 + \frac{qy}{z}\right)^{-1}\frac{q(q-1)y}{z}(x^* - 1).$$

Therefore $x_* \le x^*$, where x^* is a root of the following linear equation

$$x^* = \log\left(1 + \frac{qy}{z}\right) + \left(1 + \frac{qy}{z}\right)^{-1}\frac{q(q-1)y}{z}(x^* - 1).$$

Since $q > 1$, with a little algebra we get

$$x^* \le \left(1 + \frac{qy}{z}\right)\left[1 + \frac{q(2-q)y}{z}\right]^{-1}\log\left(1 + \frac{qy}{z}\right) \le \frac{1}{2-q}\log\left(1 + \frac{qy}{z}\right),$$

thus arriving at the following upper bound

$$F(z, y) \le \frac{y}{(2-q)^q}\log^q\left(1 + \frac{qy}{z}\right).$$

Now we are in a position to finish the proof. Notice that for any $\lambda > 0$

$$\eta'\left(\frac{\eta}{\lambda}\right)^q - z\left[\exp\left(\frac{\eta}{\lambda}\right) - 1 - \frac{\eta}{\lambda}\right]$$

$$\le \max_{x \ge 0}\left\{\eta'\left(\frac{x}{\lambda}\right)^q - z\left[\exp\left(\frac{x}{\lambda}\right) - 1 - \frac{x}{\lambda}\right]\right\} = F(z, \eta'),$$

and therefore

$$\mathbf{E}\eta'\eta^q \le \lambda^q\left\{\mathbf{E}F(z, \eta') + z\mathbf{E}\left[\exp\left(\frac{\eta}{\lambda}\right) - 1 - \frac{\eta}{\lambda}\right]\right\}$$

$$\le \lambda^q\left\{\frac{1}{(2-q)^q}\mathbf{E}\eta'\log^q\left(1 + \frac{q\eta'}{z}\right) + z\mathbf{E}\left[\exp\left(\frac{\eta}{\lambda}\right) - 1 - \frac{\eta}{\lambda}\right]\right\}.$$

Next, substituting in the above equation

$$z = q\mathbf{E}\eta'\left\{\mathbf{E}\left[\exp\left(\frac{\eta}{\lambda}\right) - 1 - \frac{\eta}{\lambda}\right]\right\}^{-1},$$

we obtain

$$\mathbf{E}\eta'\eta^q \le \lambda^q\left\{\frac{1}{(2-q)^q}\mathbf{E}\eta'\log^q\left(1 + \frac{\eta'}{\mathbf{E}\eta'}\mathbf{E}\left[\exp\left(\frac{\eta}{\lambda}\right) - 1 - \frac{\eta}{\lambda}\right]\right) + q\mathbf{E}\eta'\right\}. \quad (24)$$

Finally, applying the following inequality

$$\log^q(1 + xy) \leq [\log(1 + y) + \log(1 + x)]^q \leq 2^{q-1} \log^q(1 + x)$$
$$+ 2^{q-1} \log^q(1 + y), \ x, y > 0,$$

we get

$$\mathbf{E}\eta' \log^q \left\{ 1 + \frac{\eta'}{\mathbf{E}\eta'} \mathbf{E}\left[\exp\left(\frac{\eta}{\lambda}\right) - 1 - \frac{\eta}{\lambda}\right] \right\} \leq 2^{q-1} \mathbf{E}\eta' \log^q\left(1 + \frac{\eta'}{\mathbf{E}\eta'}\right)$$
$$+ 2^{q-1} \log^q \left\{ 1 + \mathbf{E}\left[\exp\left(\frac{\eta}{\lambda}\right) - 1 - \frac{\eta}{\lambda}\right] \right\},$$

and combining this equation with (24), we finish the proof of (23). $\quad\square$

Lemma 3. *Let η be a nonnegative sub-Gaussian random variable, i.e., such that for all $\lambda > 0$ and some $S > 0$*

$$\mathbf{E}\exp(\eta/\lambda) \leq C \exp(S^2/\lambda^2). \tag{25}$$

Then for any $q \in [1, 2)$

$$\left[\mathbf{E}\eta'^q \eta^q\right]^{1/q} \leq \frac{CS}{2 - q} \left[\mathbf{E}\eta'^q \log^q\left(1 + \frac{\eta'^q}{\mathbf{E}\eta'^q}\right)\right]^{1/q}. \tag{26}$$

Proof. Replacing η' in (23) by η'^q and substituting (25) in (23), we get with $\lambda = S$

$$\mathbf{E}\eta'^q \eta^q \leq \frac{2^{q-1} S^q}{(2 - q)^q} \mathbf{E}\eta'^q \log^q\left[1 + \frac{\eta'^q}{\mathbf{E}\eta'^q}\right] + \frac{2^{q-1} S^q}{(2 - q)^q} \mathbf{E}\eta'^q + q S^q \mathbf{E}\eta'^q.$$

Let $F(x) = x \log^q(1 + x)$. It is clear that

$$F'(x) = \log^q(1 + x) + \frac{qx \log^{q-1}(1 + x)}{1 + x}$$

is increasing in x and therefore $F(x)$ is convex. Therefore by Jensen's inequality

$$\mathbf{E}\eta'^q \log^q\left[1 + \frac{\eta'^q}{\mathbf{E}\eta'^q}\right] \geq \log(2)\mathbf{E}\eta'^q,$$

and thus, we arrive at (26). $\quad\square$

Lemma 4. *Let*

$$\zeta_\alpha = \sum_{k=1}^p \left[1 - h_\alpha(k)\right]^2 [1 - \xi'^2(k)]$$

and

$$\Sigma_\alpha = 2\|(1 - h_\alpha)^2\| \sqrt{\log\log \frac{\|(1 - h_{\alpha^\circ})^2\|^2 \exp(2)}{\|(1 - h_\alpha)^2\|^2}}.$$

Then for any $s \in (1, 2]$,

$$\mathbf{P}\left\{ \sup_{\alpha \leq \alpha^\circ} \frac{\zeta_\alpha - s\Sigma_\alpha}{\|(1 - h_\alpha)^2\|} \geq x \right\} \leq \frac{C}{(s - 1)^3} \exp\left\{ -\frac{(3 - s)^2 x^2}{16} \right\}.$$

Proof. For some $\epsilon > 0$ define α_k, $k \geq 0$, as roots of equations

$$\|(1 - h_{\alpha_k})^2\|^2 = (1 + \epsilon)^{-k}\|(1 - h_{\alpha^\circ})^2\|^2.$$

Then, denoting for brevity

$$G_{k+1}(x) = s\Sigma_{\alpha_{k+1}} + x\|(1 - h_{\alpha_{k+1}})^2\|,$$

we obtain

$$
\begin{aligned}
\mathbf{P}\left\{\sup_{\alpha \leq \alpha^\circ} \frac{\zeta_\alpha - s\Sigma_\alpha}{\|(1 - h_\alpha)^2\|} \geq x\right\} &\leq \sum_{k=0}^{\infty} \mathbf{P}\left\{\sup_{\alpha \in [\alpha_{k+1}, \alpha_k]} \frac{\zeta_\alpha - s\Sigma_\alpha}{\|(1 - h_\alpha)^2\|} \geq x\right\} \\
&\leq \sum_{k=0}^{\infty} \mathbf{P}\left\{\sup_{\alpha \in [\alpha_{k+1}, \alpha_k]} \zeta_\alpha \geq G_{k+1}(x)\right\} \\
&\leq \sum_{k=0}^{\infty} \mathbf{P}\left\{\zeta_{\alpha_{k+1}} \geq [1 - f(\epsilon)]G_{k+1}(x)\right\} \\
&+ \sum_{k=0}^{\infty} \mathbf{P}\left\{\sup_{\alpha \in [\alpha_{k+1}, \alpha_k]} [\zeta_\alpha - \zeta_{\alpha_{k+1}}] \geq f(\epsilon)G_{k+1}(x)\right\},
\end{aligned}
\tag{27}
$$

where $f(\epsilon)$ will be chosen later on.

Since $\log(1 + x) \geq x - x^2/2$, $x \geq 0$, then for any $\lambda > 0$

$$\mathbf{E}\exp(\lambda\zeta_\alpha) \leq \exp\left[\lambda^2\|(1 - h_\alpha)^2\|^2\right], \tag{28}$$

and by the exponential Tchebychev inequality we get

$$
\begin{aligned}
\mathbf{P}\left\{\zeta_{\alpha_{k+1}} \geq [1 - f(\epsilon)]G_{k+1}(x)\right\} &\leq \exp\left\{-\frac{[1 - f(\epsilon)]^2 G_{k+1}^2(x)}{4\|(1 - h_{\alpha_{k+1}})^2\|^2}\right\} \\
&\leq \exp\left\{-s^2[1 - f(\epsilon)]^2 \log[(k + 1)\log(1 + \epsilon)] - \frac{[1 - f(\epsilon)]^2 x^2}{4}\right\}.
\end{aligned}
\tag{29}
$$

To bound from above the last term in Equation (27), we make use of that $2h_\alpha(k) - h_\alpha^2(k)$ is a family of ordered functions, and thus (see (21))

$$\|(1 - h_{\alpha_k})^2 - (1 - h_{\alpha_{k+1}})^2\|^2 \leq \|(1 - h_{\alpha_k})^2\|^2 - \|(1 - h_{\alpha_{k+1}})^2\|^2.$$

Similarly to (28)

$$\mathbf{E}\exp\left\{\lambda[\zeta_{\alpha_k} - \zeta_{\alpha_{k+1}}]\right\} \leq \exp\left[\lambda^2\|(1 - h_{\alpha_k})^2 - (1 - h_{\alpha_{k+1}})^2\|^2\right].$$

Therefore with the help of Proposition 2 and the exponential Tchebychev inequality we obtain

$$\mathbf{P}\left\{\sup_{\alpha_{k+1}<\alpha\le\alpha_k}[\zeta_\alpha-\zeta_{\alpha_{k+1}}]\ge f(\epsilon)G_{k+1}(x)\right\}\le\min_{\lambda>0}\exp\left\{-\lambda f(\epsilon)G_{k+1}(x)\right.$$

$$\left.+\frac{\lambda^2(\sqrt{2}-1)^2\|(1-h_{\alpha_k})^2-(1-h_{\alpha_{k+1}})^2\|^2}{4[\|(1-h_{\alpha_k})^2\|^2-\|(1-h_{\alpha_{k+1}})^2\|^2]}\right\}$$

$$\le C\exp\left\{-\frac{(\sqrt{2}-1)^2f^2(\epsilon)G_{k+1}^2(x)}{8[\|(1-h_{\alpha_k})^2\|^2-\|(1-h_{\alpha_{k+1}})^2\|^2]}\right\}\quad(30)$$

$$=C\exp\left\{-\frac{(\sqrt{2}-1)^2s^2f^2(\epsilon)}{4\epsilon}\log[(k+1)\log(1+\epsilon)]\right.$$

$$\left.-\frac{(\sqrt{2}-1)^2x^2f^2(\epsilon)}{8\epsilon}\right\}.$$

Now we chose $f(\epsilon)$ to balance the exponents at the right-hand sides in (29) and (30), thus arriving at following equation for this function

$$\frac{(\sqrt{2}-1)^2f^2(\epsilon)}{2\epsilon}=[1-f(\epsilon)]^2.$$

This yields

$$f(\epsilon)=\frac{\sqrt{2\epsilon}}{\sqrt{2}-1+\sqrt{2\epsilon}}.$$

With this $f(\epsilon)$ and with (27–30) we get

$$\mathbf{P}\left\{\sup_{\alpha\le\alpha^\circ}\frac{\zeta_\alpha-s\Sigma_\alpha}{\|(1-h_\alpha)^2\|}\ge x\right\}\le\frac{C\exp\{-[1-f(\epsilon)]^2x^2/4\}}{\epsilon^{s^2[1-f(\epsilon)]^2}\{s^2[1-f(\epsilon)]^2-1\}_+}.$$

Finally, choosing ϵ as a root of $f(\epsilon)=(s-1)/2$, we finish the proof. $\square$

We summarize the main properties of the variance estimator in the following lemma.

Lemma 5. *For any* $q\in(1,2)$

$$\mathbf{E}\left\{[\sigma^2-\hat{\sigma}_{\hat{\alpha}}^2]_++\sigma^{-2}Pen(\hat{\alpha})\right\}^q\le C(2-q)^{-q}\mathcal{O}^q(\alpha_\circ,\alpha^\circ)\mathbf{E}\big[Pen(\hat{\alpha})\big]^q.$$

Proof. By (8) we obtain

$$\sigma^2-\hat{\sigma}_{\hat{\alpha}}^2=\frac{\sigma^2}{\|1-h_{\hat{\alpha}}\|^2}\sum_{k=1}^p\big[1-h_{\hat{\alpha}}(k)\big]^2[1-\xi'^2(k)]$$

$$-\frac{2\sigma}{\|1-h_{\hat{\alpha}}\|^2}\sum_{k=1}^p\big[1-h_{\hat{\alpha}}(k)\big]^2\xi'(k)\beta(k)\sqrt{\lambda(k)}\quad(31)$$

$$-\frac{1}{\|1-h_{\hat{\alpha}}\|^2}\sum_{k=1}^p\big[1-h_{\hat{\alpha}}(k)\big]^2\beta^2(k)\lambda(k).$$

The first term at the right-hand side in (31) is controlled with the help of Lemmas 3 and 4 (with $s = 2$) as follows:

$$
\left\{ \mathbf{E} \left| \frac{Pen(\hat{\alpha})}{\|1 - h_{\hat{\alpha}}\|^2} \sum_{k=1}^{p} [1 - h_{\hat{\alpha}}(k)]^2 [1 - \xi'^2(k)] \right|^q \right\}^{1/q}
$$

$$
= \left\{ \mathbf{E} \frac{Pen^q(\hat{\alpha})}{\|1 - h_{\hat{\alpha}}\|^q} \left| \frac{\zeta_{\hat{\alpha}} - 2\Sigma_{\hat{\alpha}} + 2\Sigma_{\hat{\alpha}}}{\|1 - h_{\hat{\alpha}}\|} \right|^q \right\}^{1/q} \tag{32}
$$

$$
\leq \frac{C}{\|1 - h_{\alpha_\circ}\|} \left\{ \log\log\left(1 + \frac{\|1 - h_{\alpha_\circ}\|^2}{\|1 - h_{\alpha^\circ}\|^2}\right) \right\}^{1/2} \left[\mathbf{E} Pen^q(\hat{\alpha}) \right]^{1/q}
$$

$$
+ \frac{C}{(2 - p)\|1 - h_{\alpha_\circ}\|} \log\left[1 + \frac{Pen(\alpha_\circ)}{Pen(\alpha^\circ)}\right] \left[\mathbf{E} Pen^q(\hat{\alpha}) \right]^{1/q}.
$$

To control the last two terms in (31), notice that $\tilde{h}_\alpha(k) = 2h_\alpha(k) - h_\alpha^2(k)$, $\alpha > 0$, is a family of ordered functions. Hence, applying Lemma 1 with

$$
b(k) = \frac{2\beta(k)\sqrt{\lambda(k)}}{K\sigma},
$$

we have

$$
\mathbf{E} \exp\left\{ \frac{2}{K\sigma^2} \left[2\sigma \sum_{k=1}^{p} [1 - \tilde{h}_{\hat{\alpha}}(k)] \xi'(k)\beta(k)\sqrt{\lambda(k)} \right. \right.
$$

$$
\left. \left. - \sum_{k=1}^{p} [1 - \tilde{h}_{\hat{\alpha}}(k)]^2 \beta^2(k)\lambda(k) \right] \right\} \leq C. \tag{33}
$$

This inequality and Lemma 2 with

$$
\eta' = Pen^q(\alpha)
$$

$$
\eta = \frac{2}{K\sigma^2} \left[2\sigma \sum_{k=1}^{p} [1 - \tilde{h}_{\hat{\alpha}}(k)] \xi'(k)\beta(k)\sqrt{\lambda(k)} - \sum_{k=1}^{p} [1 - \tilde{h}_{\hat{\alpha}}(k)]^2 \beta^2(k)\lambda(k) \right],
$$

yield

$$
\mathbf{E} \left[\frac{2\sigma}{\|1 - h_{\hat{\alpha}}\|^2} \sum_{k=1}^{p} [1 - h_{\hat{\alpha}}(k)]^2 \xi'(k)\beta(k)\sqrt{\lambda(k)} \right.
$$

$$
\left. - \frac{1}{\|1 - h_{\hat{\alpha}}\|^2} \sum_{k=1}^{p} [1 - h_{\hat{\alpha}}(k)]^2 \beta^2(k)\lambda(k) \right]^q Pen^q(\hat{\alpha}) \tag{34}
$$

$$
\leq \frac{C\sigma^{2q}}{(2 - q)^q \|1 - h_{\alpha_\circ}\|^{2q}} \log^q\left[1 + \frac{Pen(\alpha_\circ)}{Pen(\alpha^\circ)}\right] \mathbf{E} Pen^q(\hat{\alpha}).
$$

Finally, combining (31), (32), and (34) and using Jensen's inequality, we finish the proof. $\qquad \square$

2.3. Proof of Theorem 1

The following proposition (see Lemma 7 in [11]) summarizes some basic properties of the penalty defined by (13–15).

Proposition 7.

$$Q^+(\alpha) \geq D(\alpha) \max\left\{ \sqrt{\log \frac{D(\alpha)}{D(\alpha^\circ)}}, \frac{1}{\mu_\alpha} \log \frac{D(\alpha)}{D(\alpha^\circ)} \right\},$$

$$\mu_\alpha \geq \min\left\{ \frac{1}{2} \sqrt{\log \frac{D(\alpha)}{D(\alpha^\circ)}}, \frac{1}{4} \right\},$$

If $D(\alpha) \geq \exp(2)D(\alpha^\circ)$, then

$$D(\alpha) \geq \mu_\alpha Q^+(\alpha) \left[\log \frac{\mu_\alpha Q^+(\alpha)}{D(\alpha^\circ)} \right]^{-1}.$$

For any $\alpha_1 \leq \alpha_2$

$$\frac{D(\alpha_1)}{D(\alpha_2)} \leq \frac{Q^+(\alpha_1)}{Q^+(\alpha_2)}.$$

We begin the proof of Theorem 1 with a simple generalization of Proposition 3. Consider the following random process

$$\eta_\alpha^\epsilon = \sum_{k=1}^p \lambda^{-1}(k) h_\alpha^\epsilon(k) [\xi'^2(k) - 1],$$

where $h_\alpha^\epsilon(k) = [2(1+\epsilon)h_\alpha(k) - \epsilon h_\alpha^2(k)]/(2+\epsilon)$.

Lemma 6. *Let $q \in (1, 2]$. Then for any random variable $\hat{\alpha} \leq \alpha^\circ$*

$$\mathbf{E}\eta_{\hat{\alpha}}^\epsilon \leq \frac{C\sigma_{\alpha^\circ}^\epsilon}{\sqrt{q-1}} \left[\mathbf{E}\left(\frac{\sigma_{\hat{\alpha}}^\epsilon}{\sigma_{\alpha^\circ}^\epsilon} \right)^q \right]^{1/q},$$

where

$$\sigma_\alpha^\epsilon = \left\{ 2 \sum_{k=1}^p \lambda^{-2}(k) [h_\alpha^\epsilon(k)]^2 \right\}^{1/2}.$$

Proof. It is based on the following fact. Let $S(x) = x^{1/(q-1)}$, $x \in \mathbb{R}^+$. Then

$$\rho(z) \stackrel{\text{def}}{=} \mathbf{E} \sup_{\alpha \leq \alpha^\circ} \left\{ \eta_\alpha^\epsilon - z\sigma_\alpha^\epsilon S^{-1}\left(\frac{\sigma_\alpha^\epsilon}{\sigma_{\alpha^\circ}^\epsilon} \right) \right\}_+$$

$$\leq C\sigma_{\alpha^\circ}^\epsilon \int_0^\infty xS\left(\frac{x}{z} \right) e^{-Cx^2} \, dx,$$

where $S^{-1}(x) = x^{q-1}$ denotes the inverse function to $S(x)$.

To prove this inequality, define $\alpha_k,\; k = 0, 1, 2, \ldots$ as roots of the following equations

$$\sigma^\epsilon_{\alpha_k} = \sigma^\epsilon_{\alpha^\circ} S\big(1/z\big)\mathrm{e}^k.$$

Then, noticing that $\eta^\epsilon_\alpha - \eta^\epsilon_{\alpha_k}$ is an ordered process, we obtain by (12) and Proposition 2

$$\rho(z) \le \mathbf{E} \sup_{\alpha_0 \le \alpha \le \alpha^\circ} |\eta^\epsilon_\alpha| + \sum_{k=2}^\infty \mathbf{E} \sup_{\alpha_k \le \alpha < \alpha_{k-1}} \left\{ \eta^\epsilon_\alpha - z\sigma^\epsilon_{\alpha_{k-1}} S^{-1}\left(\frac{\sigma^\epsilon_{\alpha_{k-1}}}{\sigma^\epsilon_{\alpha^\circ}} \right) \right\} \Big]\Big\}_+$$

$$\le C\sigma^\epsilon_{\alpha^\circ} S\left(\frac{1}{z} \right) \sum_{k=0}^\infty \mathrm{e}^k \exp\left\{ -C\big[zS^{-1}\big(S(z^{-1})\mathrm{e}^k \big) \big]^2 \right\}$$

$$\le C\sigma^\epsilon_{\alpha^\circ} \int_0^\infty \exp\left\{ -C\big[zS^{-1}(u) \big]^2 \right\} du = C\sigma^\epsilon_{\alpha^\circ} \int_0^\infty \mathrm{e}^{-Cz^2 v^2}\, dS(v)$$

$$\le C\sigma^\epsilon_{\alpha^\circ} z^2 \int_0^\infty S(v) v \mathrm{e}^{-Cz^2 v^2}\, dv = C\sigma^\epsilon_{\alpha^\circ} \int_0^\infty x S\left(\frac{x}{z} \right) \mathrm{e}^{-Cx^2}\, dx.$$

Next we get by the Laplace method

$$\rho(z) = C^{q/(q-1)} S\left(\frac{1}{z} \right) \int_0^\infty x^{q/(q-1)} \mathrm{e}^{-x^2/2}\, dx$$

$$\le C^{q/(q-1)} \left(\frac{1}{z} \right)^{1/(q-1)} \exp\left[\frac{q}{2(q-1)} \log \frac{q}{q-1} \right]. \tag{35}$$

To finish the proof, denote for brevity

$$E = \mathbf{E}\left(\frac{\sigma^\epsilon_{\hat\alpha}}{\sigma^\epsilon_{\alpha^\circ}} \right)^q.$$

Then by (35) we obtain with a simple algebra

$$\mathbf{E}\eta^\epsilon_\alpha \le \min_z \left\{ z\mathbf{E}\sigma^\epsilon_{\hat\alpha} \sigma^\epsilon_{\alpha^\circ} S^{-1}\left(\frac{\sigma^\epsilon_{\hat\alpha}}{\sigma^\epsilon_{\alpha^\circ}} \right) + S\left(\frac{C}{z} \right) \exp\left[\frac{q}{2(q-1)} \log \frac{q}{q-1} \right] \right\}$$

$$\le C\sigma^\epsilon_{\alpha^\circ} \min_z \left\{ zE + \left(\frac{C}{z} \right)^{1/(q-1)} \exp\left[\frac{q}{2(q-1)} \log \frac{q}{q-1} \right] \right\}$$

$$\le \frac{C}{\sqrt{q-1}} \sigma^\epsilon_{\alpha^\circ} E^{1/q}. \qquad \square$$

The following important lemma provides an upper bound for $L_{\hat\alpha}(\beta) + (1 + \gamma)Q^+(\hat\alpha)$.

Lemma 7. *For any data-driven $\hat\alpha$ and any given $\bar\alpha \in [\alpha_\circ, \alpha^\circ]$, the following inequality*

$$\left\{ \mathbf{E}\big[\sigma^{-2} L_{\hat\alpha}(\beta) + (1 + \gamma)Q^+(\hat\alpha) \big]^{1+\gamma/4} \right\}^{1/(1+\gamma/4)}$$

$$\le \frac{C}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+} \left[\frac{\bar R_{\bar\alpha}(\beta)}{\gamma\sigma^2} + \frac{D(\alpha^\circ)}{\gamma^4} \right]$$

holds uniformly in $\beta \in \mathbb{R}^p$ and $\gamma \in (0, 1/4)$.

Proof. In view of the definition of $\hat{\alpha}$, for any given smoothing parameter $\bar{\alpha}$, $R_{\hat{\alpha}}[Y, Pen] \leq R_{\bar{\alpha}}[Y, Pen]$. It is easy to check with the help of (8) that this inequality is equivalent to the following one

$$
L_{\hat{\alpha}}(\beta) + (1 + \gamma)\sigma^2 Q^+(\hat{\alpha}) - \sigma^2 \sum_{k=1}^{p} \lambda^{-1}(k)\tilde{h}_{\hat{\alpha}}(k)[\xi'^2(k) - 1]
$$

$$
+ 2\sigma \sum_{k=1}^{p} \lambda^{-1/2}(k)[1 - h_{\hat{\alpha}}(k)]^2 \xi'(k)\beta(k) + [\hat{\sigma}_{\hat{\alpha}}^2 - \sigma^2]Pen(\hat{\alpha})
$$

$$
\tag{36}
$$

$$
\leq L_{\bar{\alpha}}(\beta) + (1 + \gamma)\sigma^2 Q^+(\bar{\alpha}) - \sigma^2 \sum_{k=1}^{p} \lambda^{-1}(k)\tilde{h}_{\bar{\alpha}}(k)[\xi'^2(k) - 1]
$$

$$
+ 2\sigma \sum_{k=1}^{p} \lambda^{-1/2}(k)[1 - h_{\bar{\alpha}}(k)]^2 \xi'(k)\beta(k) + [\hat{\sigma}_{\bar{\alpha}}^2 - \sigma^2]Pen(\bar{\alpha}),
$$

where $\tilde{h}_{\alpha}(k) = 2h_{\alpha}(k) - h_{\alpha}^2(k)$. We can rewrite (36) as follows:

$$
\frac{\gamma}{2}\big[L_{\hat{\alpha}}(\beta) + (1 + \gamma)\sigma^2 Q^+(\hat{\alpha})\big] \leq L_{\bar{\alpha}}(\beta) + (1 + \gamma)\sigma^2 Q^+(\bar{\alpha})
$$

$$
- \sigma^2 \sum_{k=1}^{p} \lambda^{-1}(k)\tilde{h}_{\bar{\alpha}}(k)[\xi'^2(k) - 1]
$$

$$
+ \sigma^2 \sum_{k=1}^{p} \lambda^{-1}(k)\tilde{h}_{\hat{\alpha}}(k)[\xi'^2(k) - 1] - \left(1 + \frac{\gamma}{2} - \frac{\gamma^2}{2}\right)\sigma^2 Q^+(\hat{\alpha}) \tag{37}
$$

$$
+ 2\sigma \sum_{k=1}^{p} \lambda^{-1/2}(k)[\tilde{h}_{\hat{\alpha}}(k) - \tilde{h}_{\bar{\alpha}}(k)]\xi'(k)\beta(k) - \left(1 - \frac{\gamma}{2}\right)L_{\hat{\alpha}}(\beta)
$$

$$
+ [\hat{\sigma}_{\bar{\alpha}}^2 - \sigma^2]_+ Pen(\bar{\alpha}) + [\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2]_+ Pen(\hat{\alpha}).
$$

Since $\bar{\alpha}$ is given, we get by Jensen's inequality

$$
\mathbf{E}\left|\sum_{k=1}^{p} \lambda^{-1}(k)\tilde{h}_{\bar{\alpha}}(k)[\xi'^2(k) - 1]\right|^{1+\gamma/4} \leq C\left\{\sum_{k=1}^{p} \lambda^{-2}(k)\tilde{h}_{\bar{\alpha}}^2(k)\right\}^{1/2+\gamma/8}
$$

$$
\tag{38}
$$

$$
\leq C\big[D(\bar{\alpha})\big]^{1+\gamma/4} \leq C\big[\sigma^{-2}\bar{R}_{\bar{\alpha}}(\beta)\big]^{1+\gamma/4}.
$$

The third line in (37) is bounded by Proposition 1 as follows:

$$
\mathbf{E}\left[\sum_{k=1}^{p} \lambda^{-1}(k)\tilde{h}_{\hat{\alpha}}(k)[\xi'^2(k) - 1] - \left(1 + \frac{\gamma}{2} - \frac{\gamma^2}{2}\right)Q^+(\hat{\alpha})\right]_+^{1+\gamma/4}
$$

$$
\tag{39}
$$

$$
\leq \frac{CD^{1+\gamma/4}(\alpha^\circ)}{\gamma^3},
$$

where $\gamma \leq 1/\sqrt{2}$.

The upper bound for the fourth line in (37) is a little bit more tricky. Since $\big\{\tilde{h}_\alpha(\cdot), \alpha \in (0, \alpha^\circ]\big\}$ is a family of ordered functions, we obtain by Proposition 5 that for any $\epsilon > 0$ and given $q' > 1/2$,

$$
\mathbf{E}\left|2\sigma \sum_{k=1}^{p} \lambda^{-1/2}(k)\big[\tilde{h}_{\hat{\alpha}}(k) - \tilde{h}_{\bar{\alpha}}(k)\big]\xi'(k)\beta(k) \right.
$$
$$
\left. - \epsilon\Big\{\sigma^2 \sum_{k=1}^{p} \lambda^{-1}(k)\big[\tilde{h}_{\hat{\alpha}}(k) - \tilde{h}_{\bar{\alpha}}(k)\big]^2\beta^2(k)\Big\}^{q'}\right|^q \leq \frac{1}{(C\epsilon)^{q/(2q'-1)}}. \tag{40}
$$

To continue this inequality, notice that if $\hat{\alpha} \geq \bar{\alpha}$, then

$$
\frac{\tilde{h}_{\hat{\alpha}}(k)}{\tilde{h}_{\bar{\alpha}}(k)} \leq 1, \quad \frac{\tilde{h}_{\hat{\alpha}}(k)}{\tilde{h}_{\bar{\alpha}}(k)} \geq \tilde{h}_{\hat{\alpha}}(k)
$$

and therefore

$$
\sum_{k=1}^{\infty}\big[\tilde{h}_{\bar{\alpha}}(k) - \tilde{h}_{\hat{\alpha}}(k)\big]^2\frac{\beta^2(k)}{\lambda(k)} = \sum_{k=1}^{\infty}\tilde{h}_{\bar{\alpha}}^2(k)\bigg[1 - \frac{\tilde{h}_{\hat{\alpha}}(k)}{\tilde{h}_{\bar{\alpha}}(k)}\bigg]^2\frac{\beta^2(k)}{\lambda(k)}
$$
$$
\leq \max_s \frac{\tilde{h}_{\bar{\alpha}}^2(s)}{\lambda(s)}\sum_{k=1}^{\infty}\big[1 - \tilde{h}_{\hat{\alpha}}(k)\big]^2\beta^2(k) \tag{41}
$$
$$
\leq \max_s \frac{4h_{\bar{\alpha}}^2(s)}{\lambda(s)}\sum_{k=1}^{\infty}\big[1 - h_{\hat{\alpha}}(k)\big]^2\beta^2(k).
$$

Similarly, if $\hat{\alpha} < \bar{\alpha}$, then

$$
\sum_{k=1}^{\infty}\big[\tilde{h}_{\bar{\alpha}}(k) - \tilde{h}_{\hat{\alpha}}(k)\big]^2\frac{\beta^2(k)}{\lambda(k)} \leq \max_s \frac{\tilde{h}_{\hat{\alpha}}^2(s)}{\lambda(s)}\sum_{k=1}^{\infty}\big[1 - \tilde{h}_{\bar{\alpha}}(k)\big]^2\beta^2(k)
$$
$$
\leq \max_s \frac{4h_{\hat{\alpha}}^2(s)}{\lambda(s)}\sum_{k=1}^{\infty}\big[1 - h_{\bar{\alpha}}(k)\big]^2\beta^2(k). \tag{42}
$$

So, combining (41–42) with Young's inequality

$$
yx^s - x \leq (1-s)s^{s/(1-s)}y^{1/(1-s)}, \quad x, y \geq 0, \ s < 1,
$$

gives

$$
\epsilon\bigg[\sigma^2 \sum_{k=1}^{\infty}\big[\tilde{h}_{\bar{\alpha}}(k) - \tilde{h}_{\hat{\alpha}}(k)\big]^2\lambda^{-1}(k)\beta^2(k)\bigg]^{q'} - \bigg(1 - \frac{\gamma}{2}\bigg)L_{\hat{\alpha}}(\beta)
$$
$$
\leq \bigg(1 - \frac{\gamma}{2}\bigg)^{-1}\epsilon\bigg[\sigma^2 \sum_{k=1}^{\infty}\big[\tilde{h}_{\bar{\alpha}}(k) - \tilde{h}_{\hat{\alpha}}(k)\big]^2\lambda^{-1}(k)\beta^2(k)\bigg]^{q'} - L_{\hat{\alpha}}(\beta) \tag{43}
$$
$$
\leq C\epsilon^{1/(1-q')}\bigg[\sum_{k=1}^{\infty}\big[1 - h_{\bar{\alpha}}(k)\big]^2\beta^2(k) + \sigma^2 \max_k \frac{h_{\bar{\alpha}}^2(k)}{\lambda(k)}\bigg]^{q'/(1-q')}.
$$

Thus, by (40) and (43) we obtain

$$
\mathbf{E}\left|2\sigma\sum_{k=1}^{p}\lambda^{-1/2}(k)\big[\tilde{h}_{\hat{\alpha}}(k)-\tilde{h}_{\bar{\alpha}}(k)\big]\xi'(k)\beta(k)-\left(1-\frac{\gamma}{2}\right)L_{\hat{\alpha}}(\beta)\right|^{q}
$$

$$
\leq C\mathbf{E}\left|2\sigma\sum_{k=1}^{p}\lambda^{-1/2}(k)\big[\tilde{h}_{\hat{\alpha}}(k)-\tilde{h}_{\bar{\alpha}}(k)\big]\xi'(k)\beta(k)\right.
$$

$$
\left.-\left[\sigma^2\sum_{k=1}^{p}\lambda^{-1}(k)\big[\tilde{h}_{\hat{\alpha}}(k)-\tilde{h}_{\bar{\alpha}}(k)\big]^2\beta^2(k)\right]^{q'}\right|^{q}
$$

$$
+C\mathbf{E}\left|\left[\sigma^2\sum_{k=1}^{p}\lambda^{-1}(k)\big[\tilde{h}_{\hat{\alpha}}(k)-\tilde{h}_{\bar{\alpha}}(k)\big]^2\beta^2(k)\right]^{q'}-\left(1-\frac{\gamma}{2}\right)L_{\hat{\alpha}}(\beta)\right|^{q}
$$

$$
\leq (C\epsilon)^{-\frac{q}{2q'-1}}+C\epsilon^{\frac{q}{1-q'}}\left[\sum_{k=1}^{\infty}\big[1-h_{\bar{\alpha}}(k)\big]^2\beta^2(k)+\sigma^2\max_{k}\frac{h_{\bar{\alpha}}^2(k)}{\lambda(k)}\right]^{\frac{qq'}{1-q'}}.
$$

Therefore, substituting in the above equation $q'=2/3$ and

$$
\epsilon=\left[\sum_{k=1}^{\infty}\big[1-h_{\bar{\alpha}}(k)\big]^2\beta^2(k)+\sigma^2\max_{k}\frac{h_{\bar{\alpha}}^2(k)}{\lambda(k)}\right]^{-3q},
$$

we get

$$
\mathbf{E}\left|2\sigma\sum_{k=1}^{p}\lambda^{-1/2}(k)\big[\tilde{h}_{\hat{\alpha}}(k)-\tilde{h}_{\bar{\alpha}}(k)\big]\xi'(k)\beta(k)-\left(1-\frac{\gamma}{2}\right)L_{\hat{\alpha}}(\beta)\right|^{q}
$$
$$
\leq C\left[\sum_{k=1}^{\infty}\big[1-h_{\bar{\alpha}}(k)\big]^2\beta^2(k)+\sigma^2\max_{k}\frac{h_{\bar{\alpha}}^2(k)}{\lambda(k)}\right]^{q}. \tag{44}
$$

Now we proceed with the last line in Equation (37). Since $\bar{\alpha}$ is given, we have by (31)

$$
\left\{\mathbf{E}[\hat{\sigma}_{\bar{\alpha}}^2-\sigma^2]^2\right\}^{1/2}\leq\frac{\sigma^2}{\|1-h_{\bar{\alpha}}\|^2}\left\{\sum_{k=1}^{p}\big[1-h_{\bar{\alpha}}(k)\big]^4\right\}^{1/2}
$$

$$
+\frac{2\sigma}{\|1-h_{\bar{\alpha}}\|^2}\left\{\sum_{k=1}^{p}\big[1-h_{\bar{\alpha}}(k)\big]^4\beta^2(k)\lambda(k)\right\}^{1/2}
$$

$$
+\frac{1}{\|1-h_{\bar{\alpha}}\|^2}\sum_{k=1}^{p}\{1-h_{\bar{\alpha}}(k)\}^2\beta^2(k)\lambda(k)
$$

$$
\leq\frac{C\sigma^2}{\|1-h_{\bar{\alpha}}\|}+\frac{C}{\|1-h_{\bar{\alpha}}\|^2}\sum_{k=1}^{p}\big[1-h_{\bar{\alpha}}(k)\big]^2\beta^2(k)\lambda(k)
$$

and therefore

$$
\mathbf{E}\big\{[\hat{\sigma}_{\bar{\alpha}}^2-\sigma^2]_{+}Pen(\bar{\alpha})\big\}^{1+\gamma/4}\leq C\big[\sigma^{-2}\mathcal{O}(\alpha_{\circ},\alpha^{\circ})\bar{R}_{\bar{\alpha}}(\beta)\big]^{1+\gamma/4}. \tag{45}
$$

The last term in (37) can be bounded by Lemma 5 and (16) as follows:

$$
\mathbf{E}\big\{[\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2]_+ Pen(\hat{\alpha})\big\}^{1+\gamma/4}
$$
$$
\leq C\mathcal{O}^{1+\gamma/4}(\alpha_\circ)\mathbf{E}\big\{\sigma^{-2}L_{\hat{\alpha}}(\beta) + (1+\gamma)Q^+(\hat{\alpha})\big\}^{1+\gamma/4}. \tag{46}
$$

Finally combining Equations (37), (38), (39), (44), (45), (46), we finish the proof. □

The next idea in the proof of Theorem 1 is that the data-driven parameter $\hat{\alpha}$ defined by (4) cannot be very small, or equivalently, that the ratio $D(\hat{\alpha})/D(\alpha^\circ)$ cannot be very large.

Lemma 8. *For any data-driven $\hat{\alpha}$ and any given $\bar{\alpha} \in [\alpha_\circ, \alpha^\circ]$, the following upper bound holds*

$$
\left\{\mathbf{E}\left[\frac{D(\hat{\alpha})}{D(\alpha^\circ)}\right]^{1+\gamma/4}\right\}^{1/(1+\gamma/4)}
$$
$$
\leq \frac{C}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+}\mathcal{Q}\left[\frac{\bar{R}_{\bar{\alpha}}(\beta)}{\sigma^2\gamma D(\alpha^\circ)} + \frac{1}{\gamma^4}\right], \tag{47}
$$

for any $\gamma \in (0, 1/4)$.

Proof. Representing

$$
(1+\gamma)Q^+(\hat{\alpha}) = \left(1 + \frac{\gamma}{2}\right)Q^+(\hat{\alpha}) + \frac{\gamma}{2}Q^+(\hat{\alpha}),
$$

we obtain with a simple algebra from (37)

$$
\frac{\gamma\sigma^2}{2}\left[\sum_{k=1}^{p}\frac{h_{\hat{\alpha}}^2(k)}{\lambda(k)} + (1+\gamma)Q^+(\hat{\alpha})\right] \leq \bar{R}_{\bar{\alpha}}(\beta) - \sigma^2\sum_{k=1}^{p}\frac{\tilde{h}_{\bar{\alpha}}(k)}{\lambda(k)}[\xi'^2(k) - 1]
$$
$$
+ \sigma^2\sup_{\alpha \leq \alpha^\circ}\left[\sum_{k=1}^{p}\frac{\tilde{h}_\alpha(k)}{\lambda(k)}[\xi'^2(k) - 1] - \left(1 + \frac{\gamma}{2} - \frac{\gamma^2}{2}\right)Q^+(\alpha)\right]_+
$$
$$
+ 2\sigma\sum_{k=1}^{p}\lambda^{-1/2}(k)\big[\tilde{h}_{\hat{\alpha}}(k) - \tilde{h}_{\bar{\alpha}}(k)\big]\xi'(k)\beta(k) - \left(1 - \frac{\gamma}{2}\right)L_{\hat{\alpha}}(\beta)
$$
$$
+ [\hat{\sigma}_{\bar{\alpha}}^2 - \sigma^2]Pen(\bar{\alpha}) + [\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2]_+ Pen(\hat{\alpha}).
$$

Combining this with Equations (38), (39), (44), (45), (46), we obtain

$$
\left\{\mathbf{E}\left[\frac{\|h_{\hat{\alpha}}\|_\lambda^2 + Q^+(\hat{\alpha})}{D(\alpha^\circ)}\right]^{1+\gamma/4}\right\}^{1/(1+\gamma/4)}
$$
$$
\leq \frac{C}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+}\left[\frac{C\bar{R}_{\bar{\alpha}}(\beta)}{\sigma^2\gamma D(\alpha^\circ)} + \frac{1}{\gamma^4}\right]. \tag{48}
$$

To continue this inequality, we need a lower bound for $\|h_\alpha\|_\lambda^2 + Q^+(\alpha)$. Notice that

$$f(x) \overset{\text{def}}{=} F(x) - \frac{x^2}{1 - 2x} = \frac{1}{2}\log(1 - 2x) + x + \frac{x^2}{1 - 2x}$$

is a non-negative function for $x \geq 0$ since

$$f'(x) = \frac{2x^2}{(1 - 2x)^2} \geq 0 \text{ and } f(0) = 0.$$

Therefore the following inequality holds

$$F(x) \geq \frac{x^2}{1 - 2x}. \tag{49}$$

Let

$$k_\alpha = \arg\max_k \frac{h_\alpha(k)}{\lambda(k)}, \tag{50}$$

then by (14) and (49) we obviously get

$$\log \frac{D(\alpha)}{D(\alpha^\circ)} \geq F[\mu_\alpha \rho_\alpha(k_\alpha)] \geq \frac{[\mu_\alpha \rho_\alpha(k_\alpha)]^2}{1 - 2\mu_\alpha \rho_\alpha(k_\alpha)}.$$

With this inequality we obtain

$$\mu_\alpha \rho_\alpha(k_\alpha) \leq \left\{1 + \left[1 + \log^{-1}\frac{D(\alpha)}{D(\alpha^\circ)}\right]^{1/2}\right\}^{-1},$$

thus arriving at

$$\mu_\alpha^{-1} \geq 2\rho_\alpha(k_\alpha). \tag{51}$$

Now we are in a position to bound from below $\|h_\alpha\|_\lambda^2 + Q^+(\alpha)$. By (13–15), (50–51), and (17) we obtain

$$\|h_\alpha\|_\lambda^2 + Q^+(\alpha) \geq \|h_\alpha\|_\lambda^2 + \frac{2D(\alpha)}{\mu_\alpha}\sum_{k=1}^p \frac{[\mu_\alpha \rho_\alpha(k)]^2}{1 - 2\mu_\alpha \rho_\alpha(k)}$$

$$\geq \|h_\alpha\|_\lambda^2 + \frac{D(\alpha)}{\mu_\alpha}\sum_{k=1}^p F[\mu_\alpha \rho_\alpha(k)] = \|h_\alpha\|_\lambda^2 + \frac{D(\alpha)}{\mu_\alpha}\log\frac{D(\alpha)}{D(\alpha^\circ)}$$

$$\geq \|h_\alpha\|_\lambda^2 + 2\rho_\alpha(k_\alpha)D(\alpha)\log\frac{D(\alpha)}{D(\alpha^\circ)} \tag{52}$$

$$\geq \|h_\alpha\|_\lambda^2 + \frac{h_\alpha(k_\alpha)}{\lambda(k_\alpha)}\log\frac{D(\alpha)}{D(\alpha^\circ)}$$

$$\geq \|h_\alpha\|_\lambda^2 + \log\frac{D(\alpha)}{D(\alpha^\circ)}\max_k \frac{h_\alpha(k)}{\lambda(k)} \geq CD(\alpha)\log\frac{D(\alpha)}{D(\alpha^\circ)}.$$

With the help of (52) we continue (48) as follows:

$$
\left\{ \mathbf{E} \left[\frac{D(\hat{\alpha})}{D(\alpha^\circ)} \log \frac{D(\hat{\alpha})}{D(\alpha^\circ)} \right]^{1+\gamma/4} \right\}^{1/(1+\gamma/4)}
$$

$$
\leq \frac{C}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+} \left[\frac{C\bar{R}_{\bar{\alpha}}(\beta)}{\sigma^2 \gamma D(\alpha^\circ)} + \frac{1}{\gamma^4} \right].
$$

(53)

To control from below the left-hand side in the above equation, notice that

$$
\mathbf{E} \left[\frac{D(\hat{\alpha})}{D(\bar{\alpha})} \log \frac{D(\hat{\alpha})}{D(\bar{\alpha})} \right]^{1+\gamma/4} = \frac{1}{(1+\gamma/4)^{1+\gamma/4}} \mathbf{E} \left[\frac{D(\hat{\alpha})}{D(\bar{\alpha})} \right]^{1+\gamma/4}
$$

$$
\times \left\{ \log \left[\frac{D(\hat{\alpha})}{D(\bar{\alpha})} \right]^{1+\gamma/4} \right\}^{1+\gamma/4}.
$$

(54)

To finish the proof, let us consider the function $f(x) = x \log^{1+\gamma/4}(x)$, $x \geq 1$. Computing its second order derivative, one can easily check that $f(x)$ is convex for all $x \geq \exp(1) = e$. So, $f(x + e - 1)$ is convex for $x \geq 1$. It is easily seen there exists a constant $C > 0$ such that for all $x \geq 1$

$$
f(x) \geq \frac{1}{2} f(x + e - 1) - C.
$$

Therefore by (54) and Jensen's inequality,

$$
\mathbf{E} \left[\frac{D(\hat{\alpha})}{D(\bar{\alpha})} \log \frac{D(\hat{\alpha})}{D(\bar{\alpha})} \right]^{1+\gamma/4} \geq C \left\{ \mathbf{E} \left[\frac{D(\hat{\alpha})}{D(\bar{\alpha})} \right]^{1+\gamma/4} + e - 1 \right\}
$$

$$
\times \log^{1+\gamma/4} \left\{ \mathbf{E} \left[\frac{D(\hat{\alpha})}{D(\bar{\alpha})} \right]^{1+\gamma/4} + e - 1 \right\} - C.
$$

(55)

Let

$$
\psi(x) = (x + e - 1) \log^{1+\gamma/4}(x + e - 1).
$$

It is easy to check that the inverse function $\psi^{-1}(x)$ satisfies the following inequality

$$
\psi^{-1}(x) \leq (x + e - 1) \log^{-1-\gamma/4}(x + e - 1).
$$

Therefore combining this equation and (55) with (53), we arrive at (47). $\square$

Now we are ready to proceed with the proof of Theorem 1. Let $\epsilon > 0$ be a small given number to be defined later on. By (8) and (9), the following equation for the skewed excess risk

$$\mathcal{E}(\epsilon) \overset{\text{def}}{=} \sup_{\beta \in \mathbb{R}^p} \mathbf{E}_\beta \Big\{ \|\beta - \hat{\beta}_{\hat{\alpha}}\|^2 - (1+\epsilon)\{R_{\hat{\alpha}}[Y] + \mathcal{C}\} \Big\}$$

$$= \sup_{\beta \in \mathbb{R}^p} \mathbf{E}_\beta \Big\{ -\epsilon \sum_{k=1}^p \big[1 - h_{\hat{\alpha}}(k)\big]^2 \beta^2(k) - \epsilon \sigma^2 \sum_{k=1}^p \lambda^{-1}(k) h_{\hat{\alpha}}^2(k)$$

$$-(1+\epsilon)(1+\gamma)\sigma^2 Q^+(\hat{\alpha})$$

$$-2\sigma \sum_{k=1}^p \big\{ 1 + \epsilon - [(1+2\epsilon)h_{\hat{\alpha}}(k) - \epsilon h_{\hat{\alpha}}^2(k)] \big\} \beta(k) \lambda^{-1/2}(k) \xi'(k) \qquad (56)$$

$$+\sigma^2 \sum_{k=1}^p \lambda^{-1}(k) \big[2(1+\epsilon)h_{\hat{\alpha}}(k) - \epsilon h_{\hat{\alpha}}^2(k)\big] \big[\xi'^2(k) - 1\big]$$

$$+[\sigma^2 - \hat{\sigma}_{\hat{\alpha}}^2] Pen(\hat{\alpha}) \Big\}$$

holds.

We proceed with the second line from below at the right-hand side of this display. By Lemmas 6 and 8, we obtain

$$\sigma^2 \mathbf{E} \sum_{k=1}^p \lambda^{-1}(k) \big[(1+\epsilon)h_{\hat{\alpha}}(k) - \epsilon h_{\hat{\alpha}}^2(k)\big] \big[\xi'^2(k) - 1\big]$$

$$\leq \frac{C}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+ \sqrt{\gamma}} \mathcal{Q}\bigg[\frac{\bar{R}_{\bar{\alpha}}(\beta)}{\sigma^2 \gamma D(\alpha^\circ)} + \frac{1}{\gamma^4} \bigg]. \qquad (57)$$

The next step is to bound the third line from below at the right-hand side of (56). It suffices to note that $\tilde{h}_\alpha^\epsilon(k) = \big[(1+2\epsilon)h_\alpha(k) - \epsilon h_\alpha^2(k)\big]/(1+\epsilon)$ is the family of ordered functions. Hence, by Proposition 6, we get with $\bar{\alpha} = \arg\min_{\alpha \in [\alpha_\circ, \alpha^\circ]} \bar{R}_\alpha(\beta)$

$$2\sigma \mathbf{E} \sum_{k=1}^p \big\{ 1 + \epsilon - [(1+2\epsilon)h_{\hat{\alpha}}(k) - \epsilon h_{\hat{\alpha}}^2(k)] \big\} \beta(k) \lambda^{-1/2}(k) \xi'(k)$$

$$= 2(1+\epsilon)\sigma \mathbf{E} \sum_{k=1}^p \big[h_{\bar{\alpha}}^\epsilon(k) - h_{\hat{\alpha}}^\epsilon(k)\big] \beta(k) \lambda^{-1/2}(k) \xi'(k)$$

$$\leq C\bigg[\sigma^2 \mathbf{E} \max_k \lambda^{-1}(k) h_{\hat{\alpha}}^2(k) \sum_{k=1}^p \big[1 - h_{\bar{\alpha}}(k)\big]^2 \beta^2(k) \bigg]^{1/2}$$

$$+ C\bigg[\sigma^2 \max_k \lambda^{-1}(k) h_{\bar{\alpha}}^2(k) \mathbf{E} \sum_{k=1}^p \big[1 - h_{\hat{\alpha}}(k)\big]^2 \beta^2(k) \bigg]^{1/2} \qquad (58)$$

$$\leq C\sigma^2 \epsilon^{-1} \mathbf{E} \max_k \lambda^{-1}(k) h_{\hat{\alpha}}^2(k) + \epsilon \sum_{k=1}^p \big[1 - h_{\bar{\alpha}}(k)\big]^2 \beta^2(k)$$

$$+ C\sigma^2 \epsilon^{-1} \max_k \lambda^{-1}(k) h_{\bar{\alpha}}^2(k) + \epsilon \mathbf{E} \sum_{k=1}^p \big[1 - h_{\hat{\alpha}}(k)\big]^2 \beta^2(k).$$

Therefore, substituting (45), (46), (57), (58) in (56), we obtain the following upper bound for the skewed excess risk

$$\mathcal{E}(\epsilon) \leq C\sigma^2\epsilon^{-1}\mathbf{E}\max_k \lambda^{-1}(k)h_{\hat{\alpha}}^2(k) - \sigma^2\mathbf{E}Q^+(\hat{\alpha}) + C\mathcal{O}(\alpha_\circ, \alpha^\circ)\bar{R}_{\bar{\alpha}}(\beta)$$

$$+C\sigma^2\epsilon^{-1}\max_k \lambda^{-1}(k)h_{\bar{\alpha}}^2(k) + \epsilon\sum_{k=1}^p \left[1 - h_{\bar{\alpha}}(k)\right]^2\beta^2(k) \qquad (59)$$

$$+ \frac{C\sigma^2 D(\alpha^\circ)}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+\sqrt{\gamma}}\mathcal{Q}\left[\frac{\bar{R}_{\bar{\alpha}}(\beta)}{\sigma^2\gamma D(\alpha^\circ)} + \frac{1}{\gamma^4}\right].$$

Let us consider the function

$$U(\epsilon) = \max_{\alpha \leq \alpha^\circ}\left\{C\epsilon^{-1}\max_k \lambda^{-1}(k)h_\alpha^2(k) - Q^+(\alpha)\right\}.$$

Since

$$\max_k \frac{h_\alpha^2(k)}{\lambda(k)} \leq \max_k \frac{h_\alpha(k)}{\lambda(k)} \leq \left[\sum_{k=1}^p \frac{h_\alpha^2(k)}{\lambda^2(k)}\right]^{1/2}$$

$$\leq \left\{\sum_{k=1}^p \frac{h_\alpha^2(k)}{\lambda^2(k)}[2 - h_\alpha(k)]^2\right\}^{1/2} \leq \frac{D(\alpha)}{\sqrt{2}}$$

and by Proposition 7

$$Q^+(\alpha) \geq D(\alpha)\sqrt{\log\frac{D(\alpha)}{D(\alpha^\circ)}},$$

we get

$$U(\epsilon) \leq D(\alpha^\circ)\max_{\alpha \leq \alpha^\circ}\left\{\frac{C}{\epsilon}\frac{D(\alpha)}{D(\alpha^\circ)} - \frac{D(\alpha)}{D(\alpha^\circ)}\left[\log\frac{D(\alpha)}{D(\alpha^\circ)}\right]^{1/2}\right\}$$

$$\leq D(\alpha^\circ)\max_{x \geq 1}\left\{\frac{Cx}{\epsilon} - x\sqrt{\log(x)}\right\}.$$

One can easily check with a simple algebra that

$$\max_{x \geq 1}\left\{\frac{Cx}{\epsilon} - x\sqrt{\log(x)}\right\} \leq \frac{\epsilon}{C}\exp\left[\frac{C^2}{\epsilon^2}\right]. \qquad (60)$$

Indeed, let $x^* = \arg\max_x\left\{Cx/\epsilon - x\sqrt{\log(x)}\right\}$. Then, differentiating $Cx/\epsilon - x\sqrt{\log(x)}$ in x, we obtain the following equation for x^*

$$\frac{C}{\epsilon} - \sqrt{\log(x^*)} - \frac{1}{2\sqrt{\log(x^*)}} = 0.$$

Therefore

$$x^* = \exp\left\{\left(\frac{C}{2\epsilon} + \sqrt{\frac{C^2}{4\epsilon^2} - 1}\right)^2\right\} \leq \exp\left(\frac{C^2}{\epsilon^2}\right).$$

This equation proves (60) since

$$\max_{x \geq 1}\left\{\frac{Cx}{\epsilon} - x\sqrt{\log(x)}\right\} \leq \frac{Cx^*}{\epsilon}.$$

With (60) we continue (59) as follows:

$$\mathcal{E}(\epsilon) \leq C\sigma^2 D(\alpha^\circ)\epsilon \exp\frac{C^2}{\epsilon^2} + C\mathcal{O}(\alpha_\circ, \alpha^\circ)\bar{R}_{\bar{\alpha}}(\beta)$$

$$+ C\sigma^2\epsilon \sum_{k=1}^{p}\lambda^{-1}(k)h_{\bar{\alpha}}^2(k) + \epsilon\sum_{k=1}^{p}\left[1 - h_{\bar{\alpha}}(k)\right]^2\beta^2(k)$$

$$+ \frac{C\sigma^2 D(\alpha^\circ)}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+\sqrt{\gamma}}\mathcal{Q}\left[\frac{\bar{R}_{\bar{\alpha}}(\beta)}{\sigma^2\gamma D(\alpha^\circ)} + \frac{1}{\gamma^4}\right]$$

$$\leq C\sigma^2 D(\alpha^\circ)\epsilon \exp\frac{C^2}{\epsilon^2} + C\epsilon\bar{R}_{\bar{\alpha}}(\beta) + C\mathcal{O}(\alpha_\circ, \alpha^\circ)\bar{R}_{\bar{\alpha}}(\beta)$$

$$+ \frac{C}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+\sqrt{\gamma}}\mathcal{Q}\left[\frac{\bar{R}_{\bar{\alpha}}(\beta)}{\sigma^2\gamma D(\alpha^\circ)} + \frac{1}{\gamma^4}\right].$$

Therefore, substituting this equation in

$$\mathbf{E}\|\beta - \hat{\beta}_{\hat{\alpha}}\|^2 \leq (1 + \epsilon)\bar{R}_{\bar{\alpha}}(\beta) + \mathcal{E}(\epsilon),$$

we get

$$\mathbf{E}\|\beta - \hat{\beta}_{\hat{\alpha}}\|^2 \leq [1 + C\mathcal{O}(\alpha_\circ, \alpha^\circ)]r(\beta) + C\sigma^2 D(\alpha^\circ)\times$$

$$\times \inf_{\epsilon}\left[\epsilon \exp\frac{C^2}{\epsilon^2} + \frac{\epsilon r(\beta)}{\sigma^2 D(\alpha^\circ)}\right] \tag{61}$$

$$+ \frac{C\sigma^2 D(\alpha^\circ)}{[1 - C\mathcal{O}(\alpha_\circ, \alpha^\circ)/\gamma]_+\sqrt{\gamma}}\mathcal{Q}\left[\frac{r(\beta)}{\sigma^2\gamma D(\alpha^\circ)} + \frac{1}{\gamma^4}\right].$$

Hence, to finish the proof of the theorem, it remains to check that

$$\inf_{\epsilon} F(\epsilon, u) = \inf_{\epsilon}\left[\epsilon \exp\frac{C^2}{\epsilon^2} + \epsilon u\right] \leq \frac{Cu}{\sqrt{\log(u)}}. \tag{62}$$

Let $\epsilon_* = \arg\min_{\epsilon} F(\epsilon, u)$. Then, differentiating $F(\epsilon, u)$ in ϵ, we arrive at the following equation for ϵ_*

$$\exp\left(\frac{C^2}{\epsilon_*^2}\right) - \frac{C^2}{\epsilon_*^2}\exp\left(\frac{C^2}{\epsilon_*^2}\right) + u = 0.$$

Thus

$$\frac{C^2}{\epsilon_*^2} + \log\left(\frac{C^2}{\epsilon_*^2} - 1\right) = u$$

and it follows immediately from the above equation that

$$\epsilon_* \leq \frac{C}{\sqrt{\log(u)}}$$

and therefore

$$F(\epsilon_*, u) \leq 2u\epsilon_* \leq \frac{2Cu}{\sqrt{\log(u)}},$$

thus proving (62).

Finally, substituting (62) with $u = r(\beta)/[\sigma^2 D(\alpha^\circ)]$ in (61), we complete the proof of the theorem.

References

[1] AKAIKE, H. (1973). Information theory and an extension of the maximum likelihood principle. *Proc. 2nd Intern. Symp. Inf. Theory, Petrov P.N. and Csaki F. eds. Budapest.* 267-281. MR483125

[2] BAUER, F. AND HOHAGE T. (2005). A Lepski-type stopping rule for regularized Newton methods, *Inverse Problems*, **21**, 1975–1991. MR2183662

[3] BAUER, F., MATHÉ, P. AND PEREVERZEV, S. (2007) Local solutions to inverse problems in geodesy. *Journal of Geodesy*, **81**, no. 1, pp. 39-51.

[4] BIRGÉ, L. AND MASSART, P. (2007) Minimal penalties for Gaussian model selection *Probab. Theory Relat. Fields*, **138**, 33–73. MR2288064

[5] BISSANTZ, N., HOHAGE, T., MUNK A. AND RUYMGAART F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.*, **45**, no. 6, 2610–2636. MR2361904

[6] CAVALIER, L. AND GOLUBEV, YU. (2006). Risk hull method and regularization by projections of ill-posed inverse problems, *Ann. of Stat.*, **34**, pp. 1653–1677. MR2283712

[7] CAVALIER L., GOLUBEV G., PICARD, D., TSYBAKOV B. (2002). Oracle inequalities for inverse problems. *Annals of Stat.*, **30** No 3, 843–874. MR1922543

[8] DEY, A.K., RUYMGAART, F. H. AND MAIR, B. A. (1996). Cross-validation for parameter selection in inverse estimation problems. *Scandinavian Journal of Statistics*, vol. 23, pp. 609-620. MR1439715

[9] ENGL, H.W., HANKE M. AND NEUBAUER, A. (1996). *Regularization of Inverse Problems.* Kluwer Academic Publishers. MR1408680

[10] GOLUBEV, YU. (2004). The principle of penalized empirical risk in severely ill-posed problems. *Probab. Theory and Relat. Fields*, **130**, 18–38. MR2092871

[11] GOLUBEV, YU. (2010). On universal oracle inequalities related to high dimensional linear models. *Ann. of Statist.*, 38, Number 5, 2751–2780. MR2722455

[12] HERMAN, G. T. (2009) *Fundamentals of computerized tomography: Image reconstruction from projection.* 2nd edition, Springer.

[13] KNEIP, A. (1994). Ordered linear smoothers, *Ann. Statist.*, **22**, pp. 835–866. MR1292543

[14] LANDWEBER, L. (1951). An iteration, formula for Fredholm integral equations of the first kind, *Amer. J. Math.*, **73**, 615–624. MR0043348

[15] LOUBES, J.-M. AND LUDEÑA, C. (2008). Adaptive complexity regularization for linear inverse problems, *Electronic J. of Statist.*, **2**, 661–677. MR2426106

[16] MAIR B. AND RUYMGAART F.H. (1996). Statistical estimation in Hilbert scale. *SIAM J. Appl. Math.*, **56**, no. 5, 1424–1444. MR1409127

[17] MATHÉ, P. (2006). The Lepskii principle revised, *Inverse Problemes*, **22**, no. 3, L11–L15. MR2235633

[18] PINSKER, M.S. (1980). Optimal filtration of square-integrable signals in Gaussian noise. *Problems Inform. Transmission.* **16** 120–133. MR0624591

[19] SULLIVAN, F.O. (1986). A statistical perspective on ill-posed inverse problems, *Statist. Sci.*, **1**, no. 4, 501–527. MR874480

[20] TIKHONOV, A.N. AND ARSENIN, V.A. (1977). *Solution of Ill-posed Problems. Preface by translation editor Fritz John. Scripta Series in Mathematics.* V. H. Winston & Sons, Washington, D.C.: John Wiley & Sons, New York. MR0455365

[21] VAN DER VAART, A. AND WELLNER, J. (1996). *Weak convergence and empirical processes.* Springer-Verlag, New York. MR1385671