RESTORATION OF BINARY IMAGES FOR BERNOULLIAN NOISE MODELS

F. Götze[1] B.A. Zalesky[1]

December 11 1996

Abstract. In this paper the problem of the restoration of binary images from Bernoulli random noise is investigated. Assuming that an image has regular boundary, we show that linear averaging estimators are nearly optimal up to a logarithmic factor of the number of pixels. Moreover, we provide examples showing that for any estimator there exist images with regular boundary such that an estimation error is of the order of the size of this boundary. Furthermore we introduce a class of *hierarchical threshold* estimators which are wavelet based estimators achieving this lower bound. In terms of Besov norms these estimators have optimal convergence rates. Some simulation examples illustrate the performance of these estimators.

1. Introduction.

The restoration of images from a random noise is an important problem in image processing and non-parametric estimation. Various methods have been proposed for this purpose. One can divide them in two classes. Methods which use specific prior information and methods which use generic prior information like a specification of smoothness, e.g., Besov norms. Below we consider algorithms for the latter. Extensive numerical testing of several estimation procedures which do not use specific prior information and are based on linear averaging, non-linear thresholding and morphology operations, led to the conclusion that all of them have a comparable quality of restoration of binary images. To explain this observation in a model which is comparatively simple to analyze we investigate in Section 2 the problem of recovering binary images sampled with Bernoulli random noise as studied in a number of papers, see for instance, Greig *et al.* (1989). We prove that for a set of images with "simple" boundary (for instance, areas restricted by differentiable curves or polygons) linear averaging estimators are optimal in terms of the L^s-risk, $s \geq 1$, (or equivalently in

[1]Supported by the SFB 343 in Bielefeld

1991 *Mathematics Subject Classification.* Primary 62G05; secondary 62M40.

Key words and phrases. Binary images, random noise, L^s-losses, L^s-risk, linear estimator, threshold estimator.

the number of misclassified pixels) up to a logarithmic factor. See Corollary 2.3 and Theorem 2.4. But the situation may change drastically once specific prior information is available. In Example 2.6 we show that by using such a prior information one can restore a rectangle by a simple linear estimator with exponentially decreasing L^s-risk.

In Section 3 we estimate Besov norms in terms of isoperimetric ratios of images, i.e., by the relative number of pixels on the boundaries of a binary image in order to compare the averaging procedures of Section 2 with the more refined estimators defined later.

In Section 4 we introduce an improved version of the threshold estimator defined by Donoho and Johnstone (1992, 1994a,b, 1995).

The so called *hierarchical threshold* estimator checks the coefficients of Haar basis decomposition of a binary image at every level comparing it to some cutoff value determined by large deviations approximations. These tests may suggest that at *any* multi-resolution level this coefficient should have been zero which means that the pertaining dyadic square (for this coefficient in the Haar basis) should consist of white or black pixels only. This enforces all coefficients of the Haar basis decomposition at the higher levels (describing the image in this square) to be zero. This comparison allows to reduce the probability of misclassification of pixels outside vicinity of the boundary of a binary image. Those square regions of the image, where the tests lead to an undetermined conclusion, (i.e., which are not overwhelmingly colored black or white) are subsequently processed by the standard procedures like non-linear threshold estimators or linear moving average estimators.

Finally, upper and lower estimates of the L^s-risk for these hierarchic wavelet threshold estimators in terms of Besov semi-norms are compared with the risk of the linear estimators of Section 2.

Acknowledgment. The authors are grateful to V. Bentkus and A. Yu. Zaitsev for help in preparing the paper.

2. Linear estimators for images with regular boundaries

Let $\mathbb{R}^2$ be an Euclidean plane and $\mathbb{Z}^2 \subset \mathbb{R}^2$ the standard lattice of points with integer coordinates. A function $f : x \mapsto f_x : \mathbb{R}^2 \to \mathbb{R}$ will be called an image. The image is called binary if the function f takes values 0 and 1 only. There is an obvious one to one correspondence between binary images f and subsets $A \subset \mathbb{R}^2$ via $A = \{x \in \mathbb{R}^2 : f(x) = 1\}$. Speaking about binary images we shall use both notation—f and A, assuming always that $A = \{x \in \mathbb{R}^2 : f(x) = 1\}$.

In the following we shall consider discrete images. For any positive integer N, we consider a resolution $1/N$, a scaled lattice $\mathbf{Z} = N^{-1}\mathbb{Z}^2$ and denote by Q the unit half-open square $[0,1) \subset \mathbb{R}^2$. Write $\mathbf{Q} = Q \cap \mathbf{Z}$. Here and in the following we do not indicate the dependence on the underlying resolution $1/N$. For example, if $N = 1$ then we have $\mathbf{Z} = \mathbb{Z}^2$.

Let us represent the space $\mathbb{R}^2$ as a union of the disjoint pixel squares $z^* \overset{\text{def}}{=} z + N^{-1}Q \subset \mathbb{R}^2$, for $z \in \mathbf{Z}$, that is, $\mathbb{R}^2 = \bigcup_{z\in\mathbf{Z}} z^*$. For simplicity we shall identify the pixel squares z^* with the pixels $z \in \mathbf{Z}$; that will not lead to a confusion since z defines in a unique way the pixel square z^*. An image f is called discrete (of the resolution $1/N$) if the function f is constant on pixel squares, i.e.,

$$f_z = f_y, \qquad \text{for all } z \in \mathbf{Z} \text{ and } y \in z^*. \tag{2.1}$$

In the view of (2.1), for any discrete image $f : \mathbb{R}^2 \to \mathbb{R}$ (resp. binary image $A \subset \mathbb{R}^2$), there is a unique image $f : \mathbf{Z} \to \mathbb{R}$ (resp. $A \subset \mathbf{Z}$). Indeed, knowing f as a function on $\mathbf{Z}$ (resp. $A \subset \mathbf{Z}$), we can recover f as a function on $\mathbb{R}^2$ (resp. $A \subset \mathbb{R}^2$) in a unique way. Finally, we speak about images (resp. binary images) restricted to the window Q if $f_x = 0$, for $x \in \mathbb{R}^2 \setminus Q$ (resp. if $A \subset Q$).

In the case of discrete images we introduce the following notions of neighborhoods of pixels and boundaries of binary images. Write $B_z = B + z$, for $z \in \mathbf{Z}$ and $B \subset \mathbf{Z}$. The square $C = N^{-1}\{-1, 0, 1\}^2$ has the center 0 and consists of the nearest neighbors of 0 in $\mathbf{Z}$. Note that C contains 9 points of the lattice $\mathbf{Z}$. For a resolution $1/N$ and an image $A \subset \mathbf{Z}$, we define the (discrete) *boundary* of A as

$$\eth A = \bigl\{ z \in \mathbf{Z} : \ C_z \cap A^c \neq \varnothing \text{ and } C_z \cap A \neq \varnothing \bigr\},$$

where $A^c = \mathbf{Z} \setminus A$ and C denotes the square described above. Furthermore, B-neighborhoods A_B of $A \subset \mathbf{Z}$ are defined by $A_B = \bigcup_{z\in A} B_z$. The frame of Q will be denoted $\operatorname{Fr} Q = \mathbf{Q} \cap \eth Q$. Due to the correspondence of images in $\mathbb{R}^2$ and $\mathbf{Z}$, the images B_z, C, $\eth A$, A_B and $\operatorname{Fr} Q$ as subsets of $\mathbb{R}^2$ are defined in a natural way as unions of the corresponding pixel squares in $\mathbb{R}^2$. For example, $\operatorname{Fr} Q$ consists of pixel squares in Q nearest to the (topological) boundary of Q in $\mathbb{R}^2$. Alternatively, we can consider $\operatorname{Fr} Q$ as a subset of $\mathbf{Q} \subset \mathbf{Z}$.

The following definition (Götze 1992) characterizes discrete images with regular boundaries. Let $|A|$ denote the cardinality of $A \subset \mathbf{Z}$.

Definition 2.1. The (discrete) isoperimetric ratio of an image $A \subset \mathbf{Z}$ is defined by

$$I(f) = I(A) = \frac{|\eth A|}{\sqrt{|A|}}\,.$$

In case of varying pixel resolutions in which the *same* image A can be expressed, the quantity $I(A)$ may be interpreted as a characteristic of the image. It is well known, that for any binary image $A \subset \mathbb{R}^2$ with Euclidean boundary ∂A, the isoperimetric ratio of Euclidean length $\mathrm{L}(\partial A)$ of ∂A and area $\mathrm{S}(A)$ of A satisfies

$$\mathrm{L}(\partial A)\big/\sqrt{\mathrm{S}(A)} \geq 2\sqrt{\pi}.$$

A simple comparison of $|\eth A|$ and $|A|$ with $\mathrm{L}(\partial A)$ and $\mathrm{S}(A)$ shows that a corresponding inequality $I(A) \geq 2\sqrt{\pi}$ holds for the discrete isoperimetric ratio.

Let us introduce the following statistical model. Let ξ_z, $z \in \mathbf{Z}$, be independent Bernoulli random variables (r.v.) taking values 0 and 1 only such that

$$p_z = \mathbf{P}\left\{\xi_z = 1\right\}, \qquad \mathbf{P}\left\{\xi_z = 0\right\} = 1 - p_z, \qquad 0 < p_z \leq p < 1/2. \tag{2.2}$$

In the case of independent identically distributed (i.i.d.) random variables ξ_z we shall assume that $p_z = p$, for all z.

We shall consider discrete binary images f observed with Bernoullian noise ξ_z, that is, we shall assume that we observe images Y sampled via

$$Y_z = f_z + (1 - 2f_z)\,\xi_z = \begin{cases} 1 - f_z, & \text{if } \xi_z = 1, \\ \quad f_z, & \text{if } \xi_z = 0. \end{cases} \tag{2.3}$$

Note, that the estimation of f_z for $p_z > 1/2$ is equivalent to the estimation in the case $p_z < 1/2$ of $1 - f_z$. The case $p_z = 1/2$ is unsolvable since any estimate of a pixel will have error probability $1/2$.

An estimator of discrete binary images in $\mathbf{Z}$ (respectively on $\mathbf{Q}$) is a map $f \mapsto \widehat{f}$: $\{0,1\}^{\mathbf{Z}} \to \mathbb{R}^{\mathbf{Z}}$ (respectively $f \mapsto \widehat{f} : \{0,1\}^{\mathbf{Q}} \to \mathbb{R}^{\mathbf{Q}}$). As an example of an estimator, let us consider the *linear averaging* estimator

$$\widehat{f}_z = \mathbf{I}\Big\{\, |B|^{-1} \sum_{y \in B_z} Y_y > r \,\Big\}, \tag{2.4}$$

where $\mathbf{I}\{D\}$ denotes the indicator function of the event D. The linear averaging estimator (2.4) depends on two parameters: a subset $B \subset \mathbf{Z}$ and a threshold $0 \leq r \leq 1$. For images observed in Q, the linear averaging estimator we define as

$$\widehat{f}_z = \mathbf{I}\Big\{\, |\mathbf{B}_z|^{-1} \sum_{y \in \mathbf{B}_z} Y_y > r \,\Big\}, \qquad \text{where } \mathbf{B}_z = B_z \cap \mathbf{Q}. \tag{2.5}$$

Note that for $z \in (\eth \operatorname{Fr} Q)_B$, that is, for points z near to the B-neighborhood of the frame, the sets $\mathbf{B}_z$ are clipped.

Let $\|f\|_s$ denote the $L^s(Q)$-norm with respect to the Lebesgue-measure on Q. For (discrete) images of resolution $1/N$ in the window Q, we introduce the L^s-risk of an estimator $\widehat{f}$ of f by

$$R(f, \widehat{f}) = \mathbf{E}\,\big\|f - \widehat{f}\big\|_s^s = N^{-2}\,\mathbf{E} \sum_{z \in \mathbf{Q}} |f_z - \widehat{f}_z|^s, \qquad 0 \leq s < \infty. \tag{2.6}$$

If f and $\widehat{f}$ are binary then we have

$$R(f, \widehat{f}) = N^{-2} \sum_{z \in \mathbf{Q}} \mathbf{P}\left\{f_z \neq \widehat{f}_z\right\}, \tag{2.7}$$

i.e., the risk is *independent* of $0 < s < \infty$ and it is just the expected misclassification probability.

The proofs of the next Proposition and other results will be given in Section 5. Introduce the square $C(m) = \mathbf{Z} \cap N^{-1}\left[-m,\, m\right]^2$.

Proposition 2.2. *Let A be a discrete binary image in Q. Then the risk of the linear estimator* (2.5) *with $r = 1/2$ and $B = C(m)$ satisfies*

$$R(f, \widehat{f}) \le \exp\bigl\{-(2-4p)^2 m^2\bigr\} + N^{-2}\bigl|(\eth A)_{C(m)}\bigr|, \qquad 0 < s < \infty.$$

The following simple conclusion illustrates the near optimality property of linear estimators for a wide class of images. It suggests some suitable asymptotic neighborhood size for averaging as the resolution $1/N \to 0$. Estimating $\bigl|(\eth A)_{C(m)}\bigr| \le |\eth A|\,(2m+1)^2$ and $|A| \le N^2$, Proposition 2.2 yields the following

Corollary 2.3. *For binary images in Q and the averaging estimator* (2.5) *with $r = 1/2$ and $B = C_m$, we have*

$$R(f, \widehat{f}) \le \exp\bigl\{-(2-4p)^2 m^2\bigr\} + I(A)\,(2m+1)^2 N^{-1}. \tag{2.8}$$

In particular, for $m = \bigl\lceil \sqrt{a_p \log N}\,\bigr\rceil$ with $a_p = 1/(2-4p)^2$ we have

$$R(f, \widehat{f}) \le c_p\, I(A)\, N^{-1} \log N. \tag{2.9}$$

Throughout $c, c_1, c_2, \ldots$ denote generic absolute constants. If a constant depends on a parameter, say p, then we shall write c_p or $c(p)$. By $\lceil a \rceil$ we denote the minimal integer such that $\lceil a \rceil \ge a$.

We derived Corollary 2.3 using very rough estimates. For sets A such that $\bigl|(\eth A)_{C(m)}\bigr| \le cmN\,I(A)$ one can improve (2.8) and (2.9). In particular, in that case (2.9) holds with $\sqrt{\log N}$ instead of $\log N$.

The next Theorem shows that linear averaging estimators are optimal up to logarithmic factors. Let $\mathcal{F}_\mu$ denote the set of binary discrete images A in Q such that $I(A) < \mu$.

Theorem 2.4. *Let $\mu > 20$. For sufficiently large $N \ge N_\mu$, we have*

$$\inf_{\widehat{f}} \sup_{f \in \mathcal{F}_\mu} R(f, \widehat{f}) \ge \frac{c\mu p}{N}, \tag{2.10}$$

where the inf *is taken over all binary estimators $\widehat{f}$, and c is a positive absolute constant.*

Remark 2.5. The lower bound in (2.10) is attained already on images of a very simple structure. Let us consider the following class $\mathcal{F}_0$ of images in Q: let $A_0 \subset \mathbf{Q}$ denote the square with side length $\lceil N/4 \rceil$. Let $A_1 \subset \mathbf{Q}$ denote a set such that $|A_1| = \bigl\lceil \mu N/100 \bigr\rceil$ and $A_1 \cap A_0 = \varnothing$. Assume that the sets $\eth\{x\}$ and $\eth\{z\}$ are disjoint and $\eth\{x\} \cap A = \varnothing$, for all $x, z \in A_1$, $x \ne z$. Then $\mathcal{F}_0$ can be defined as

$$\mathcal{F}_0 = \bigl\{ A : A_0 \subset A \subset A_0 \cup A_1 \bigr\}.$$

It is easy to verify that $I(A) \le \mu$, for sufficiently large $N \ge N_\mu$ and $A \in \mathcal{F}_0$.

Thus for images with regular boundaries the risk can not be better than the relative size of the boundary of the image. The following Example 2.6 shows that the situation may drastically change if some prior information about the image is available.

We shall use the following parametrization of square pixels in Q: the set Q is a union of N^2 of square pixels, and for each such square, say z, we write $z \sim (i,j) \in \mathbb{Z}^2$ if the square is in i-th row and j-th column, counting rows from above and columns from left to right, $1 \le i, j \le N$. For example, $\operatorname{Fr} Q$ is the union of $4N-2$ square pixels z in Q such that $z \sim (i,j)$ with $i = 1, N$ or $j = 1, N$.

Let $\mathcal{F}_h$ denote the set of images f which are half-squares in Q, i.e.,

$$f = \mathbf{I}\big\{ z \in \mathbf{Q} : \ z \sim (i,j), \ i \le k \big\}, \qquad\qquad 1 \le k \le N.$$

In the case of $f \in \mathcal{F}_h$, the estimation of f is equivalent to the estimation of the parameter k. Introduce the estimator

$$\widehat{k} = \min\Big\{ l : \ 1 \le l \le N, \sum_{z \in \mathbf{Q}, \ z \sim (l,j)} Y_z \le N/2 \Big\}.$$

Example 2.6. For images $f \in \mathcal{F}_h$, the estimator $\widehat{f} = \mathbf{I}\big\{ z \sim (i,j) : i \le \widehat{k} \big\}$ satisfies

$$R(f, \widehat{f}) \le \exp\big\{ -(1-2p)^2 N/4 \big\}.$$

3. Besov spaces and isoperimetric ratios

In Section 4 we shall introduce a hierarchical type of (nonlinear) threshold estimator. We shall investigate the risk of these estimators in terms of Besov norms and the isoperimetric ratios. The results are applicable to gray level images and general nonparametric problems. We shall describe connections between Besov semi-norms and isoperimetric ratios of binary images.

Recall that we can consider binary images f as function $f : \mathbb{R}^2 \to \mathbb{R}$ constant on square pixels. For $f \in L^s(Q)$, define the modulus $w(t; f, s)$ of smoothness by

$$w^s(t; f, s) = \sup_{|h| \le t} \int_Q \big| f(x+h) - f(x) \big|^s \mathbf{I}\big\{ x + h \in Q \big\}\, dx.$$

For $1 < q < \infty$ and $0 < \alpha \le 1$, Besov semi-norm $|f|_{\alpha,q,s}$ and norm $\|f\|_{\alpha,q,s}$ define by

$$|f|^q_{\alpha,q,s} = \int_0^1 \Big(\frac{w(t; f, s)}{t^\alpha} \Big)^q \frac{dt}{t}, \qquad\qquad \|f\|_{\alpha,q,s} = \|f\|_s + |f|_{\alpha,q,s}. \tag{3.1}$$

In the case of binary images, Besov norms are closely related to the isoperimetric ratio.

Proposition 3.1. *Assume that an image f is binary and that $1 < s < \infty$, $0 < \alpha < 1/s$. Then we have*

$$\|f\|_{\alpha,q,s}^{s} \le cN^{-1}\,|\eth A| \le 2\,c\,I(A), \qquad\qquad c = 4^{1/s}/(-\alpha + 1/s). \tag{3.2}$$

The estimate (3.2) *is precise up to a constant, that is, there exist a binary image f and absolute positive constants c_1 and c_2 such that $\|f\|_{\alpha,q,s}^{s} \ge c_1 N^{-1}\,|\eth A\,| > c_2\,I(A)$.*

The proof of Proposition 3.1 is given in Section 5.

The following example shows that in general the lower bound of Proposition 3.1 does not hold.

Example 3.2. Let A denote a binary image in Q which consists of $N/2$ identical rectangles with length 1 and width N^{-1} arranged with distances N^{-1} in Q. For this image $|f|_{\alpha,q,s}^{s}$ is equivalent to $N^{\alpha s}$ and the ratio $N^{-1}\big|\,\eth A_f\,\big|$ is proportional to N.

4. Hierarchical threshold estimators

Threshold estimators have been studied intensively in non-parametric estimation problems for functions sampled with random noise because they promise some advantages over linear estimators in special function classes of Besov, Tribel, and Hölder type. Wavelet based threshold estimator have been proposed and studied by Donoho (1992), Donoho and Johnstone (1992, 1994a,b), Donoho *et al.* (1995). Different aspects of the wavelet thresholding estimators have been studied by Coifman, Meyer and Wickerhauser (1992), DeVore and Lucier (1991), Kerkyacharian and Picard (1992) and others.

Below we describe a class of *hierarchical threshold* estimators which seem especially appropriate for estimation problems of binary images sampled with random noise.

In order to formulate our noise model for functions in Besov spaces we introduce the following notation: For a dyadic number $N = 2^l$ and $0 \le m \le l$, let D_m denote the set of squares of a dyadic decomposition of $Q = [0,1)^2$ into $2^{2(l-m)}$ squares of side length 2^{-m} each. The square pixels in D_m are denoted by J_m and its indicator function by φ_{J_m}. The images of resolution $1/N$ are piecewise constant functions $f_x = f_{J_l}$, for $x \in J_l$, and we have

$$f = \sum_{J_l \in D_l} f_{J_l}\,\varphi_{J_l},$$

where the coefficients f_{J_l} assume values 0 and 1 only. We consider the i.i.d. Bernoullian noise model (2.3), that is, ξ_{J_l}, $J_l \in D_l$, denotes a system of i.i.d. random variables taking values 0 and 1 only such that

$$\mathbf{P}\left\{\,\xi_{J_l} = 1\,\right\} = p \quad \text{and} \quad \mathbf{P}\left\{\,\xi_{J_l} = 0\,\right\} = 1 - p, \qquad 0 < p < 1/2.$$

Here $\xi_{J_l} = 1$ or $\xi_{J_l} = 0$ means that the corresponding square pixel will be altered by noise changing 1 into 0 or 0 into 1. Hence the coefficients Y_{J_l} of the observed random image Y with are given by

$$Y_{J_l} = f_{J_l} + (2\, f_{J_l} - 1)\, \xi_{J_l} \quad \text{and} \quad Y = \sum_{J_l \in D_l} Y_l\, \varphi_{J_l}.$$

For a square J, let N_J denote the number of pixel squares J_l contained in J. Define the projection P_m of f (by averaging over N_J pixel values) to the space V_m spanned by the functions φ_{J_m}, $J_m \in D_m$ via

$$f_J \stackrel{\text{def}}{=} N_J^{-1} \sum_{J_l \subset J} f_l \quad \text{and} \quad P_m f = \sum_{J_m \in D_m} f_{J_m}\, \varphi_{J_m}.$$

In order to define the class of hierarchical estimators fix integers k and K such that $0 \le k < K \le l$. Assume that we have estimated the overall percentage of errors p via an estimate $\widehat{p}$. We consider the relative frequencies Y_{J_m} of "black" pixels on the dyadic squares J_m in the level m such that $k \le m \le K$. For the level k we test any dyadic square J whether it is dominantly black or white, accepting this hypothesis if

$$Y_J \ge 1 - \widehat{p} - r_m \quad \text{or} \quad Y_J \le \widehat{p} + r_m, \tag{4.1}$$

for some fixed sequence of positive numbers $0 < r_1 < \cdots < r_l < \frac{1}{2} - \widehat{p}$ which will be defined later. Denote by $\mathcal{E}_k$ the (random) set and by $\mathcal{F}_k$ the (random) union of all dyadic squares J which passed this test on the level k. We now proceed recursively. On the following level ($K \ge m > k$) of smaller squares we will test only those dyadic squares J which are *not* contained in squares which have been determined so far to be dominantly black or white. Let $\mathcal{E}_m$ again denote the set of squares $J = J_m$ not contained in $\mathcal{F}_{m-1}$ which pass the test (4.1) at the m-th level and define the union $\mathcal{F}_m$ of dominant black or white regions (up to the m-th level) as the union of $\mathcal{F}_{m-1}$ and the squares in $\mathcal{E}_m$.

In the remaining region $Q \setminus \mathcal{F}_K$ consisting of the "undecided" squares we use another arbitrary estimator $\widehat{f}$. In the following we shall discuss linear estimators and (nonlinear) threshold estimators as examples. More precisely we introduce

Definition 4.1. Define the class of hierarchical estimators via

$$\widehat{g} = \sum_{m=k}^{K} \sum_{J_m \in \mathcal{E}_m} \Big((1 - \widehat{p})\, \mathbf{I}\big\{ Y_{J_m} \ge 1 - \widehat{p} - r_m \big\} + \widehat{p}\, \mathbf{I}\big\{ Y_{J_m} \le \widehat{p} + r_m \big\} \Big)\, \varphi_{J_m} + \widehat{f}\, \mathbf{I}\big\{ Q \setminus \mathcal{F}_K \big\}, \tag{4.2}$$

for some asymptotically unbiased estimator $\widehat{f}$ acting on the subset of "critical" squares $Q \setminus \mathcal{F}_K$.

For $\widehat{f}$ in (4.2) chosen as the linear asymptotically unbiased averaging estimator

$$\widehat{f} = (P_K Y - \widehat{p})/(1 - \widehat{p})$$

we obtain a hierarchical estimator estimator called $\widehat{g}_L$.

Another example is the nonlinear *threshold* estimator $\widehat{f}_\varepsilon$ which is connected with the wavelet representation of an image. Let

$$\Delta_{J_m} Y \overset{\text{def}}{=} Y_{J_m} - Y_{J_{m-1}}, \qquad J_m \subset J_{m-1}, \quad \Delta_{J_1} Y \overset{\text{def}}{=} Y_{J_1}$$
$$Y = \sum_{m=0}^{l} \sum_{J_m \in D_m} \Delta_{J_m} Y \varphi_{J_m}$$

(see, for instance, DeVore and Popov (1987)). The *wavelet threshold* estimator removes the small coefficients from the wavelet decomposition of Y, i.e.,

$$\widehat{f}_\varepsilon \overset{\text{def}}{=} \sum_{m=0}^{l} \sum_{J_m \in D_m} \Delta_{J_m} Y \, \mathbf{I}\big\{ |\Delta_{J_m} Y| > 2^m \varepsilon \big\} \, \varphi_{J_m},$$

for a real number $\varepsilon = c_0 \, 2^{-l} l^{-1/2}$ and a thresholding constant $c_0 > 0$. Donoho and Johnstone (1992) proved that for the $\mathcal{N}(0,1)$ distributed Gaussian random noise the thresholding constant c_0 can be chosen as $c_0 = \sqrt{2}$. Using

$$\widehat{f} = (\widehat{f}_\varepsilon - \widehat{p})/(1 - \widehat{p})$$

as an asymptotically unbiased estimator in (4.2), we obtain another example of an hierarchical estimator called $\widehat{g}_{NL}$.

Before we formulate the results we add some general remarks.

Remark 4.2. (i) Note that for $k \le K$ the condition $r_m < \dfrac{1}{2} - \widehat{p}$ must be satisfied since otherwise $Y_{J_m} \ge 1 - \widehat{p} - r_m$ and $Y_{J_m} \le \widehat{p} + r_m$ might occur together and our estimators would not be correctly defined.

(ii) Furthermore, the Besov spaces $B_{\alpha,q,s}$ contain discontinuous functions only if $\alpha < 1/s$ since the subspace $\text{span}\big\{ \varphi_{J_l} \,:\, J_l \in D_l \big\}$ of $B_{\alpha,q,s}$, $1/s \le \alpha \le 1$ consist of constants only.

In order to describe the performance of these estimators we investigate the L^2-risk in terms of Besov norms and isoperimetric ratios. Consider Besov spaces $B_{\alpha,\tau,\tau}$

with $0 < \alpha \le 1/2$ and $\tau = 2/(\alpha+1)$, and denote by $|f|_{\alpha,\tau} = |f|_{\alpha,\tau,\tau}$ the Besov semi-norm as introduced in (3.1). Notice that $4/3 \le \tau < 2$.

As an estimate $\widehat{p}$ of the unknown parameter $p < 1/2$ we define the root $\widehat{p} < 1/2$ of

$$\widehat{p}\,(1-\widehat{p}) = \frac{4}{3\,|D_l|} \sum_{J_l \in D_l} (\Delta_{J_l} Y)^2 \tag{4.3}$$

if the right hand side in (4.3) is smaller than 1/4. We define $\widehat{p} = 1/2$ if the right hand side in (4.3) is larger than 1/4. The error of this estimator can be bounded in terms of the Besov semi-norm $|f|_{\alpha,\tau}$ as follows.

Proposition 4.3. *We have*

$$\mathbf{E}\,(\widehat{p}-p)^2 \le 2^{-\tau\alpha l}\left(1 + |f|_{\alpha,\tau}\right), \qquad \tau = 2/(\alpha+1).$$

Furthermore, we choose $k = \lceil \tau\, l/2 \rceil$ and assume that the natural numbers l, K, and the real numbers

$$r_m = 6\,(\log 2)^{1/2}\, l^{1/2}\, 2^{-l+m} \quad \text{and} \quad \gamma_l \overset{\text{def}}{=} 3\, c_2\, 2^{\alpha\tau} 2^{-l\alpha\tau}\, |f|^{\tau}_{\alpha,\tau} + 2^{-l+2}\, l^{1/2}$$

and $k \le m \le K$ satisfy the inequality

$$16\sqrt{\gamma_l/3} \le r_m < 1 - 2\widehat{p}, \tag{4.4}$$

where c_2 is an absolute constant defined in (5.9).

Theorem 4.4. *Define* $\overline{\varkappa} = K/l$ *and* $\underline{\varkappa} = k/l$. *For the numbers* l, k, K *and* r_m *satisfying condition* (4.4), *the mean square risk of* $\widehat{g}_L$ *is bounded by*

$$\mathbf{E}\left\|f - \widehat{g}_L\right\|_2^2 \le c_p\, N^{-2\overline{\varkappa}\alpha}\left(1 + \|f\|^2_{2,\alpha}\right) + N^{-\alpha\,\tau(\underline{\varkappa}+1)/2}\, l^{\alpha\,\tau/4}\, \|f\|^{\tau}_{\alpha,\tau}.$$

Note that this bound can be expressed in terms of $B_{\alpha,\tau,s}$ *via* $||f||^2_{2,\alpha} \le ||f||^{\tau}_{\alpha,\tau}$.

In order to formulate the result for $\widehat{g}_{NL}$ we shall need an estimate of $\mathbf{E}\left\|f - \widehat{f}_\varepsilon\right\|_2^2$. The proof of the following proposition repeats the proof of an analogous result of Donoho and Johnstone (1992) for the threshold estimator (see also Götze and Zalesky (1995)).

Corollary 4.5. *Assume that the conditions of Theorem 4.4 are satisfied. Then there exists a thresholding constant c_0 such that for $f \in B_{\alpha,\tau,s}$ we have*

$$\mathbf{E}\,\big\|f - \widehat{f}_\varepsilon\big\|_2^2 \le c_p N^{-\alpha\tau}\, l^{\alpha\tau/2}\big(1 + \|f\|_{\alpha,\tau}^\tau\big).$$

Now we can formulate the result for the estimator g_{NL}.

Theorem 4.6. *Under conditions of Theorem 4.4 the L^2-risk of g_{NL} can be bounded by*

$$\mathbf{E}\,\big\|f - \widehat{g}_{NL}\big\|_2^2 \le c N^{-\alpha\tau(1+\underline{\varkappa})/2}\, l^{\alpha\tau/4}\,\big(1 + \|f\|_{\alpha,\tau}^\tau\big).$$

The following proposition shows that the estimates of Theorem 4.4 are optimal up to logarithmic factors.

Corollary 4.7. *There exists an image f such that*

$$\mathbf{E}\,\big\|f - \widehat{g}_L\big\|_2^2 \ge N^{-\alpha\tau(1+\underline{\varkappa})/2}\,\|f\|_{\alpha,\tau}^\tau, \qquad \mathbf{E}\,\big\|f - \widehat{g}_{NL}\big\|_2^2 \ge N^{-\alpha\tau(1+\underline{\varkappa})/2}\,\|f\|_{\alpha,\tau}^\tau.$$

In Section 5 below we displayed in Figures 3, 4 and 5 simulation results for g_{NL}. Although the theoretical bounds in terms of Besov norms are somewhat weaker than those for the wavelet threshold estimators, the small scale artifacts typical for all kind of wavelet restorations (Haar as well as regular) are much reduced due to the hierarchical testing strategy. For comparison we included in Figures 6 and 7 the results of linear average estimates.

5. Proofs.

The following large deviation bound will be frequently used.

Lemma 5.1 (Hoeffding 1963). *Let Z_m denote independent random variables such that $\mathbf{P}\big\{a_m \le\ Z_m \le\ b_m\big\} = 1$, for some non-random a_m and b_m and $1 \le m \le n$. Then we have*

$$\mathbf{P}\Big\{\sum_{m=1}^n (Z_m - \mathbf{E}\,Z_m) \ge\ nx\Big\} \le \exp\Big\{-2n^2x^2\Big/\sum_{m=1}^n (b_m - a_m)^2\Big\}, \qquad x \ge 0.$$

Proof of Proposition 2.2. Recall that $B = C(m)$ and introduce the sets

$$D_1 = \mathbf{Q} \cap (\eth A)_B, \qquad D_2 = A \setminus D_1, \qquad D_3 = A^c \setminus D_1,$$

where $A^c = \mathbf{Q} \setminus A$. Then $\mathbf{Q} = D_1 \cup D_2 \cup D_3$. Both f and $\widehat{f}$ are binary. Therefore, using (2.7), we have

$$R(f, \widehat{f}) \leq N^{-2}(R_1 + R_2 + R_3),$$

where

$$R_1 = \sum_{z \in D_1} \mathbf{P}\left\{ f_z \neq \widehat{f}_z \right\}, \qquad R_2 = \sum_{z \in D_2} \mathbf{P}\left\{ f_z \neq \widehat{f}_z \right\}, \qquad R_3 = \sum_{z \in D_3} \mathbf{P}\left\{ f_z \neq \widehat{f}_z \right\}.$$

In order to prove the Proposition, it suffices to show that

$$R_1 \leq \left| (\eth A)_B \right|, \quad R_2 \leq |A| \exp\left\{ -(2-4p)^2 m^2 \right\}, \quad R_3 \leq |A^c| \exp\left\{ -(2-4p)^2 m^2 \right\} \tag{5.1}$$

since $|A| + |A^c| = N^2$. The first inequality in (5.1) is obvious since $\mathbf{P}\left\{ f_z \neq \widehat{f}_z \right\} \leq 1$.

Let us estimate R_2. Write $S_z = \sum_{x \in \mathbf{B}_z} \xi_x$. Notice that $\mathbf{B}_z \subset A$, for all $z \in D_2$. Therefore $f_x = 1$, for all $x \in \mathbf{B}_z$, and using the definition (2.5) of $\widehat{f}_z$ and (2.3), we obtain

$$\begin{aligned} \mathbf{P}\left\{ f_z \neq \widehat{f}_z \right\} &= \mathbf{P}\left\{ \widehat{f}_z \neq 1 \right\} = \mathbf{P}\left\{ \widehat{f}_z = 0 \right\} \\ &= \mathbf{P}\Big\{ \sum_{y \in \mathbf{B}_z} (1 - \xi_y) \leq |\mathbf{B}_z|/2 \Big\} = \mathbf{P}\Big\{ S_z \geq |\mathbf{B}_z|/2 \Big\}. \end{aligned} \tag{5.2}$$

Notice that the set $\mathbf{B}_z = \big(C(m) + z \big) \cap \mathbf{Q}$ contains at least m^2 points since the shifted square $C(m) + z$ contains $(2m+1)^2$ points and has center $z \in \mathbf{Q}$. Therefore, by an application of Lemma 5.1, we obtain

$$\mathbf{P}\Big\{ S_z \geq |\mathbf{B}_z|/2 \Big\} \leq \exp\left\{ -(2-4p)^2 |\mathbf{B}_z| \right\} \leq \exp\left\{ -(2-4p)^2 m^2 \right\}. \tag{5.3}$$

Combining (5.2) and (5.3), summing over $z \in D_2$ and estimating $|D_2| \leq |A|$, we derive the desired bound for R_2.

The estimation of R_3 is similar to that of R_2. In this case $\mathbf{B}_z \subset A^c$, for all $z \in D_3$. Therefore $f_x = 0$, for all $x \in \mathbf{B}_z \cap \mathbf{Q}$, and similarly to (5.2) and (5.3) we have

$$\mathbf{P}\left\{ f_z \neq \widehat{f}_z \right\} = \mathbf{P}\left\{ S_z > |\mathbf{B}_z|/2 \right\} \leq \exp\left\{ -(2-4p)^2 m^2 \right\}. \tag{5.4}$$

Summing (5.4) over $z \in D_1$ and estimating $|D_3| \leq |A^c|$, we obtain (5.1) for R_3. $\square$

Proof of Theorem 2.4. Let $\mathcal{F}_0$ be the class of images described in Remark 2.5. Fix a binary estimator, say H, that is, fix a mapping $H : g \mapsto H(g) : \{0,1\}^{\mathbf{Q}} \to \{0,1\}^{\mathbf{Q}}$. According to the noise model (2.3) we observe random images $Y(g)$ such that

$$Y_z(g) = g_z(1 - \xi_z) + (1 - g_z)\xi_z, \qquad z \in \mathbf{Q}.$$

To prove the Theorem, it suffices to show that $\sup_{g\in\mathcal{F}_0} R\big(g, H\big(Y(g)\big)\big) \geq cp\mu/N$. Introducing the average risk

$$R_0 = N_0^{-1} \sum_{g\in\mathcal{F}_0} R\big(g, H\big(Y(g)\big)\big), \qquad N_0 = \operatorname{card}\mathcal{F}_0,$$

it suffices to prove that $R_0 \geq cp\mu/N$. Thus, we have to estimate R_0. We have

$$R_0 \geq \frac{1}{N_0\,N^2} \sum_{g\in\mathcal{F}_0} \sum_{x\in A_1} \mathbf{E}\left|g_x - H_x\big(Y(g)\big)\right| = \frac{1}{N_0\,N^2} \sum_{x\in A_1} \sum_{g\in\mathcal{F}_0} \mathbf{E}\left|g_x - H_x\big(Y(g)\big)\right|. \tag{5.5}$$

For $x \in A_1$, introduce the mapping $g \mapsto g^{(x)} : \mathcal{F}_0 \to \mathcal{F}_0$ such that $g_z^{(x)} = g_z$, for $z \neq x$, and $g_x^{(x)} = 1 - g_x$. The mapping is a bijection of $\mathcal{F}_0$ onto $\mathcal{F}_0$, and therefore

$$\sum_{g\in\mathcal{F}_0} \left|g_x - H_x\big(Y(g)\big)\right| = \sum_{g\in\mathcal{F}_0} \left|g_x^{(x)} - H_x\big(Y(g^{(x)})\big)\right|.$$

Thus, writing

$$\theta = \left|g_x - H_x\big(Y(g)\big)\right| + \left|g_x^{(x)} - H_x\big(Y(g^{(x)})\big)\right|$$

and using (5.5), we have

$$R_0 \geq 2^{-1}\,N_0^{-1}\,N^{-2} \sum_{g\in\mathcal{F}_0} \sum_{x\in A_1} \mathbf{E}\;\mathbf{E}_x\,\theta, \tag{5.6}$$

where $\mathbf{E}_x$ denotes the partial summation with respect to the distribution of ξ_x, that is, the conditional expectation given ξ_z, $z \neq x$.

Define $X_z = Y_z(g)\,\mathbf{I}\{z \neq x\}$ and $Z_z = \mathbf{I}\{z = x\}$. Then

$$Y(g) = X + \big(g_x\,(1-\xi_x) + (1-g_x)\,\xi_x\big)\,Z, \quad Y(g^{(x)}) = X + \big((1-g_x)(1-\xi_x) + g_x\,\xi_x\big)\,Z. \tag{5.7}$$

Notice that θ is independent of g_x. Indeed, using (5.7) with $g_x = 0$ and $g_x = 1$, we obtain

$$\theta = 1 - H_x\big(X + (1-\xi_x)\,Z\big) + H_x\big(X + \xi_x\,Z\big).$$

Since $0 < p < 1/2$ and $0 \leq H_x \leq 1$, we have

$$\mathbf{E}_x\,\theta = 1 + (1-2p)\,H_x(X) + (2p-1)\,H_x\big(X+Z\big) \geq 1 + (2p-1)\,H_x\big(X+Z\big) \geq 2p.$$

Consequently, (5.6) implies $R_0 \geq pN^{-2}|A_1| \geq cp\mu/N$ since $|A_1| = \lceil \mu N/100 \rceil$. $\square$

Proof of Proposition 3.1. Expressing the difference of the indicator functions $f(x)$ and $f(x+h)$ via the symmetric difference of sets we get

$$\begin{aligned}\big|\, f(x+h) - f(x)\,\big| &= \mathbf{I}\big\{\, x \in (A+h)\,\Delta\, A\,\big\}\\ &= \sum_{y\in A}\ \sum_{z\notin A} \mathbf{I}\big\{\, x\in (y+h)\cap z\,\big\} + \mathbf{I}\big\{\, x\in y\cap (z+h)\,\big\}\end{aligned}$$

and therefore by simple geometric estimates

$$w^s(t; f, s) \le \sup_{|h|\le t} \int_Q \mathbf{I}\big\{\, x\in (A+h)\,\Delta\, A\,\big\}\, dx \le 4tN^{-1}\,|\eth A|$$

On the other hand there exists a binary image A such that for $t \le \frac{1}{8}$ the inequality

$$w^s(t; f, s) \ge \frac{1}{2}\, tN^{-1}\,|\eth A|$$

holds. For example, one can choose $f = \mathbf{I}\big\{ x \in \big[0,\, 1/2\big]^2 \big\}$. These inequalities together prove the Proposition. □

Proof of Proposition 4.3. Introduce the process Z taking values 1 and -1 such that $Y = f + Z$ and $Z_{J_l} = (1-2f_{J_l})\,\xi_{J_l}$. Recall that $Z_{J_{l-1}} = \frac{1}{4}\sum_{J_l\subset J_{l-1}} Z_{J_l}$ with corresponding definitions for $f_{J_{j-1}}$ and $Y_{J_{j-1}}$. Split the part

$$\widehat{I} = \frac{1}{|D_l|} \sum_{J_l\in D_l} (\Delta_{J_l} Y)^2$$

of the Besov semi-norm introduced in (4.3) as follows

$$\begin{aligned}\widehat{I} &= \sum_{J_l\in\mathcal{D}_l} 2^{-2l}(Y_{J_l} - Y_{J_{l-1}})^2\\ &= \sum_{J_{l-1}\in\mathcal{D}_{l-1}}\ \sum_{J_l\in J_{l-1}} 2^{-2l}\big((\Delta_{J_l} f)^2 + 2\,(\Delta_{J_l} f)\,(\Delta_{J_l} Z) + (\Delta_{J_l} Z)^2\big)\\ &= I_1 + \widehat{I}_2 + \widehat{I}_3. \end{aligned} \tag{5.8}$$

Up to absolute constants c_1 and c_2 there is an equivalence of the Besov semi-norm $|f|_{\alpha,\tau}$ such that

$$|f|^\tau_{\alpha,\tau} \le c_1\, 2^{-\alpha\tau} \sum_{m=0}^{l}\ \sum_{J_m\in D_m} 2^{-\tau m}\,|\Delta_{J_m} f|^\tau \le c_2\, |f|^\tau_{\alpha,\tau}, \tag{5.9}$$

(see deVore and Popov (1987), p. 243). We shall estimate $\mathbf{E}\,\widehat{I}$ by means of this semi-norm. First note that

$$|f_{J_l} - f_{J_{l-1}}| \geq \frac{1}{4}\,\mathbf{I}\{\,f_{J_l} \neq f_{J_{l-1}}\,\}.$$

Thus we have

$$2^{-\alpha\tau}\,2^{(2-\tau)l}\,4^{2-\tau}\,I_1 \leq 2^{-\alpha\tau}\sum_{J_l\in D_l} 2^{-\tau l}\,|\Delta_{J_l}f|^{\tau}\big(4\,|\Delta_{J_l}f|\big)^{-(2-\tau)} \leq c_2\,|f|^{\tau}_{\alpha,\tau}.$$

Hence

$$I_1 \leq R_l \overset{\text{def}}{=} 2^{\alpha\tau+2}\,2^{-2\alpha l/(\alpha+1)}c_2\,|f|^{\tau}_{\alpha,\tau}.$$

Furthermore, since $\mathbf{E}\,(Z_{J_l} - Z_{J_{l-1}}) = -2p\,\Delta_{J_l}f$ we obtain $|\,\mathbf{E}\,\widehat{I}_2\,| \leq 4pR_l$. Similarly we get after some elementary computations

$$\frac{1}{4}\sum_{J_l\subset J_{l-1}} \mathbf{E}\,(Z_{J_l} - Z_{j_{l-1}})^2 = \frac{3}{4}\,p(1-p) + \frac{p^2}{4}\sum_{J_l\subset J_{l-1}} (\Delta_{J_l}f)^2.$$

Thus we conclude

$$\Big|\,\mathbf{E}\,\widehat{I} - \frac{3}{4}\,p(1-p)\,\Big| \leq (1+4p+p^2)\,2^{\alpha\tau+2}\,2^{-\alpha l\tau}\,c_2\,|f|^{\tau}_{\alpha,\tau}. \tag{5.10}$$

By (5.8), $\widehat{I}$ is a sum of $2^{2(l-1)}$ bounded independent random variables

$$H_{J_{l-1}} \overset{\text{def}}{=} \sum_{J_l\in J_{l-1}} 2^{-2l}\big(2\,(\Delta_{J_l}f)\,(\Delta_{J_l}Z) + (\Delta_{J_l}Z)^2\big).$$

Since $|H_{J_{l-1}}| \leq 2^{-2l}\cdot 2\cdot 3\cdot \left(\frac{3}{4}\right)^2 \leq 7\cdot 2^{-2l}$ we obtain from Lemma 5.1 with ($nx = 2^{-l+2}l^{1/2}$ and $n = 2^{2(l-1)}$)

$$\mathbf{P}\left\{\,|\,\mathbf{E}\,\widehat{I} - \widehat{I}\,| \geq 2^{-l+2}l^{1/2}\right\} \leq \exp\{-2\,l\}.$$

Thus we conclude with probability at least $q_l \overset{\text{def}}{=} 1 - \exp\{-2\,l\}$ in view of (5.10)

$$\Big|\,\widehat{I} - \frac{3}{4}\,p(1-p)\,\Big| \leq \gamma_l = 3\,c_2\,2^{\alpha\tau}2^{-l\alpha\tau}\,|f|^{\tau}_{\alpha,\tau} + 2^{-l+2}\,l^{1/2}.$$

Hence by definition of $\widehat{p}$ we obtain at least with probability q_l

$$|p-\widehat{p}| \leq \min\left\{\,\sqrt{4\gamma_l/3},\; 4\gamma_l\Big/\Big(9\,\big[1 - 16\,(\widehat{I}+\gamma_l)/3\big]_+\Big)^{1/2}\right\}, \tag{5.11}$$

where we used the function $[s]_+ = \max\{0,\, s\}$. This stochastic bound immediately implies for sufficiently large l

$$\mathbf{E}\,(\widehat{p}-p)^2 \le 2^{-l\alpha\tau}\left(1+|f|^{\tau}_{\alpha,\tau}\right),$$

thus proving the result. $\square$.

Proof of Theorem 4.4. We have

$$\mathbf{E}\left\|f-\widehat{g}_L\right\|_2^2 \le 2\,\mathbf{E}\left\|f-P_Kf\right\|_2^2 + 2\,\mathbf{E}\left\|P_Kf-\widehat{g}_L\right\|_2^2,$$

where

$$f-P_Kf = \sum_{J_K\in D_K}\ \sum_{J_l\in J_K}\left(f_{J_l}-f_{J_K}\right)\varphi_{J_l}$$

Since $|f_{J_l}-f_{J_K}|\le 1$ the inequality

$$\begin{aligned}\left\|f-P_Kf\right\|_2^2 &= \sum_{J_K\in D_K}\ \sum_{J_l\in J_K} 2^{-2l}\left(f_{J_l}-f_{J_K}\right)^2\\ &\le \sum_{J_K\in D_K}\ \sum_{J_l\in J_K} 2^{-2l}\left|f_{J_l}-f_{J_K}\right|^{\tau} \le \left\|f-P_Kf\right\|_{\tau}^{\tau}\end{aligned}$$

holds. It is known that for any $1<\tau\le 2$ the following Sobolev type inequality holds

$$\left\|f-P_Kf\right\|_{\tau}^{\tau} \le c\,w^{\tau}(2^{-K};f,\tau),$$

where c does not depend on K (see, for instance, DeVore and Popov (1987)). Therefore,

$$\left\|f-P_Kf\right\|_2^2 \le\ c\min\left\{2^{-2\alpha K}\|f\|^2_{2,\alpha}\,,\ 2^{-\alpha\tau K}\|f\|^{\tau}_{\alpha,\tau}\right\}.$$

In order to estimate $\mathbf{E}\left\|P_Kf-\widehat{g}_L\right\|_2^2$ we use the decomposition $Q=\mathcal{F}_K\cup I\setminus\mathcal{F}_K$ and obtain

$$\mathbf{E}\left\|P_Kf-\widehat{g}_L\right\|_2^2 = \mathbf{E}\left\|P_Kf-\widehat{g}_L\right\|^2_{L^2(\mathcal{F}_K)} + \mathbf{E}\left\|P_Kf-\widehat{g}_L\right\|^2_{L^2(Q\setminus\mathcal{F}_K)}.$$

Estimating the second term, we get

$$\begin{aligned}I_0 &\overset{\text{def}}{=} \mathbf{E}\left\|P_Kf-\widehat{g}_L\right\|^2_{L^2(Q\setminus\mathcal{F}_K)} \le\ \mathbf{E}\left\|P_Kf-\widehat{h}_L\right\|_2^2\\ &\le \mathbf{E}\sum_{J_K\in D_K} 2^{-2K}\left(f_{J_K}-\frac{Y_{J_K}-\widehat{p}}{1-2\widehat{p}}\right)^2 = \sum_{J_K\in D_K} 2^{-2K}\,\mathbf{E}\,V^2_{J_K},\end{aligned}$$

where we write

$$V_{J_K} = \frac{(1-2\widehat{p})f_{J_K} - Y_{J_K} + \widehat{p}}{1-2\widehat{p}} = \frac{\widehat{p}(1-2f_{J_K}) - X_{J_K}}{1-2\widehat{p}} = \frac{p(1-2f_{J_K}) - X_{J_K}}{1-2p} + R_p,$$

and where $|R_p| \le c_1(p)\,|\widehat{p} - p|$ with probability $1 - \exp\{-2l\}$ and the constant depends on p only (see (5.11)). Here we have used the representation $Y = f + X$. Using Lemma 5.1 and computing the 2nd moment of $p\,(1 - f_{J_K}) - X_{J_K}$, we obtain the bound

$$I_0 \le c_2(p)\, 2^{-2(l-K)}$$

with a constant $c_2(p)$ increasing to infinity when p tends to $1/2$. Hence,

$$\mathbf{E}\,\big\| f - \widehat{g}_L \big\|^2_{L^2(Q\setminus\mathcal{F}_K)} \le c_p\, \min\Big\{ 2^{-2\alpha K}\, \|f\|^2_{\alpha,\tau}\,,\; 2^{-2\alpha\tau K}\, \|f\|_{\alpha,\tau} \Big\} + c_p\, 2^{-2(l-K)} \tag{5.12}$$

Thus it suffices to estimate the expectation $\mathbf{E}\,\big\| P_K f - \widehat{g}_L \big\|^2_{L^2(\mathcal{F}_K)}$. Here we need some additional notation. Let n_{J_m} denote the minimum of the numbers of black and white pixels in J_m, i.e., $n_{J_m} = 2^{2(l-m)} \min\big\{ f_{J_m},\, 1 - f_{J_m} \big\}$ and introduce the threshold number $k_{J_m} = 2^{2(l-m)+1} r_m / (1-2p)$. Furthermore denote by J_m^+ those dyadic cubes with numbers n_{J_m} exceeding the threshold k_{J_m} (i.e., squares with comparatively large number of different values 1 and 0). Let D_m^+ denote the union of all $J_m^+ \in D_m$ and D_m^- denote the remaining squares. Define the random set E_+ of squares which pass the test (4.1) on the level m, i.e., let E_+ denote the set of $J_m \subset D_m^+$ such that $J_m \not\subset \mathcal{F}_{m-1}$ and

$$Y_{J_m} \ge 1 - \widehat{p} - r_m \quad \text{or} \quad Y_{J_m} \le \widehat{p} + r_m \tag{5.13}$$

holds. Let E_- denote correspondingly the set of $J_m \subset D_m^-$ such that $J_m \not\subset \mathcal{F}_{m-1}$ and (5.13) fails. We have

$$\mathbf{E}\,\big\| P_K f - \widehat{g}_L \big\|^2_{L^2(\mathcal{F}_K)} = \mathbf{E} \sum_{m=k}^{K} \Big(\sum_{J_m \in E_m^+} + \sum_{J_m \in E_m^-} \Big) \big\| P_K f - \widehat{g}_L \big\|^2_{L^2(J_m)} = F^+ + F^-. \tag{5.14}$$

Moreover, let χ_m denote the indicator of the test (5.13) which is 1 if (5.13) holds and 0 otherwise. Then

$$F^+ \le \sum_{m=k}^{K} \sum_{J_m \in D_m^+} \mathbf{E}\,\big\| P_K f - \widehat{g}_L \big\|^2_{L^2(J_m)}\, \chi_m.$$

Every expectation $\mathbf{E}\,\big\| P_K f - \widehat{g}_L \big\|^2_{L^2(J_m)}\, \chi_m$ in the last inequality is bounded from above by

$$2^{-2m} \Big(\mathbf{P}\,\big\{ Y_{J_m} > 1 - \widehat{p} - r_m \big\} + \mathbf{P}\,\big\{ Y_{J_m} < \widehat{p} + r_m \big\} \Big).$$

But for $J_m \in D_m^+$

$$p + 2r_m < \mathbf{E}\, Y_{J_m} < 1 - p - 2r_m.$$

Hence

$$\mathbf{P}\left\{ Y_{J_m} > 1 - \widehat{p} - r_m \right\} \le \mathbf{P}\left\{ \eta > p - \widehat{p} + r_m \right\}, \qquad \eta \overset{\text{def}}{=} Y_{J_m} - \mathbf{E}\, Y_{J_m},$$

and

$$\mathbf{P}\left\{ Y_{J_m} < \widehat{p} + r_m \right\} \le \mathbf{P}\left\{ \eta < \widehat{p} - p - r_m \right\}.$$

Thus

$$\begin{aligned}\mathbf{P}\left\{ \eta > p - \widehat{p} + r_m \right\} &\le \mathbf{P}\left\{ \eta > p - \widehat{p} + r_m,\ |\widehat{p} - p| \le \sqrt{42\gamma_l/3} \right\} \\ &\qquad + \mathbf{P}\left\{ |\widehat{p} - p| \ge \sqrt{4\gamma_l/3} \right\} \\ &\le \mathbf{P}\left\{ \eta > r_m/2 \right\} + \exp\{-2l\} \le c\, l^{-1} 2^{-2l}.\end{aligned}$$

Similarly $\mathbf{P}\left\{ \eta > \widehat{p} - r_m \right\} \le c\, l^{-1} 2^{-2l}$. Hence,

$$F^+ \le 2^{-2l}. \tag{5.15}$$

We now consider the set F^-. Introduce the indicator variables

$$\chi'_m = \mathbf{I}\left\{ Y_{J_m} \ge 1 - \widehat{p} - r_m \right\} \qquad \text{and} \qquad \chi''_m = \mathbf{I}\left\{ Y_{J_m} \le \widehat{p} + r_m \right\}.$$

If a square $J_m \in E_m^-$ contains more black pixels (that is, $f_{J_m} > 1/2$) then $\mathbf{E}\, Y_{J_m} > 1 - p - 2r_m$ and the inequality

$$\begin{aligned}\mathbf{E}\, \chi''_m = \ &\mathbf{P}\left\{ Y_{J_m} \le \widehat{p} + r_m \right\} \le \mathbf{P}\left\{ Y_{J_m} - \mathbf{E}\, Y_{J_m} \le 3r_m - (1 - 2\widehat{p}) + p - \widehat{p} \right\} \\ \le\ &\mathbf{P}\left\{ Y_{J_m} - \mathbf{E}\, Y_{J_m} \le -r_m/2 \right\} + \exp\{-2l\} \le c\, l^{-l} 2^{-2l}\end{aligned}$$

holds. Here we have used condition (4.4) and the estimates of similar probabilities in F^+. On the other hand if $f_{J_m} < 1/2$ in J_m we can prove similarly

$$\mathbf{E}\, \chi'_m \le c\, l^{-l} 2^{-2l}.$$

Suppose that $J_m \in E_m^-$. In this case

$$\mathbf{E} \left\| P_K f - \widehat{g}_L \right\|^2_{L^2(J_m)} \left(\chi'_m + \chi''_m \right) \le \mathbf{E} \left\| P_K f - \widehat{g}_L \right\|^2_{L^2(J_m)} \chi'_m + c\, l^{-l} 2^{-2(l+m)}, \tag{5.16}$$

if $f_{J_m} > 1/2$ and

$$\mathbf{E} \left\| P_K f - \widehat{g}_L \right\|^2_{L^2(J_m)} \left(\chi'_m + \chi''_m \right) \le \mathbf{E} \left\| P_K f - \widehat{g}_L \right\|^2_{L^2(J_m)} \chi''_m + c\, l^{-l} 2^{-2(l+m)}, \tag{5.17}$$

if $f_{J_m} < 1/2$. Moreover,

$$\| P_K f - \widehat{g}_L \|^2_{L^2(J_m)} \chi'_m = n_{J_m} 2^{-2l} \chi'_m. \tag{5.18}$$

Recall the equivalent representation (5.9) of a Besov semi-norm

$$|f|^\tau_{\alpha,\tau} \approx 2^{-\alpha\tau} \sum_{m=0}^{l} \sum_{J_m \in D_m} 2^{-\tau m} |\Delta_{J_m} Y|^\tau.$$

Expressing the $L^2(J_m)$-norm in terms of wavelet coefficients we get

$$\begin{aligned} \| f - P_{m-1} f \|^2_{L^2(J_{m-1})} &= \sum_{h=m}^{l} \| (P_h - P_{h-1}) f \|^2_{L^2(J_{m-1})} \\ &= \sum_{h=m}^{l} \sum_{J_h \in D_h} 2^{-2h} (\Delta_{J_h} Y)^2. \end{aligned}$$

Hence we may estimate the conditional minimum of the Besov semi-norm given the $L^2(J_m)$-norm in terms of coordinates as follows (using $|\beta_{J_h}| \le 1$ and $n_{J_m} \le k_m$)

$$\begin{aligned} \delta_{J_m} &= \min_{\beta_{J_h}} \left\{ \sum_{h=m}^{l} \sum\nolimits^* 2^{-h\tau} |\beta_{J_h}|^\tau \;\middle|\; \sum_{h=m+1}^{l} \sum\nolimits^* 2^{-2h} \beta^2_{J_h} \ge n_{J_m} 2^{-2l} \right\} \\ &\ge c \big(n_{J_m} 2^{-2l} \big)^{\tau/2}, \end{aligned} \tag{5.19}$$

where $\sum^*$ denotes the sum over all h such that $J_i \in D_i,\ J_i \subset J_m$ and $c > 0$ denotes an absolute constant. Hence

$$n_{J_m} 2^{-2l} \chi'_m \le n_{J_m} 2^{-2l} \delta^{-1}_{J_m} \| P_K f - \widehat{g}_L \|^\tau_{B^{\alpha,1}_\tau (L^\tau(J_m))} \chi'_m.$$

This inequality together with (5.16)–(5.19) implies

$$\begin{aligned} F^- &\le \mathbf{E} \sum_{m=k}^{K} \sum_{J_m \in E^-_m} \| P_K f - \widehat{g}_L \|^2_{L^2(J_m)} \big(\chi'_m + \chi''_m \big) \\ &\le \sum_{m=k}^{K} \sum_{J_m \in E^-_m} \big(n_{J_m} 2^{-2l} \big)^{1-\tau/2} \| P_K f - \widehat{g}_L \|^\tau_{B^{\alpha,1}_\tau (L^\tau(J_m))} \chi'_m + 2^{-2l}. \end{aligned}$$

By the definition of the set E_m^- and using $n_j \le k_j$ we finally obtain

$$F^- \le c\, 2^{-(l+k)\alpha/(\alpha+1)}\, l^{\alpha/(2\alpha+2)}\, \|f\|_{\alpha,\tau}^{\tau} + 2^{-2l}.$$

Now the Theorem follows from this combined with (5.12), (5.14) and (5.15). □

Proof of Theorem 4.6. To prove the Theorem we shall use the representation

$$\mathbf{E}\left\|f - \widehat{g}_{NL}\right\|_2^2 = \mathbf{E}\left\|f - \widehat{g}_{NL}\right\|_{L^2(\mathcal{F}_K)}^2 + \mathbf{E}\left\|f - \widehat{g}_{NL}\right\|_{L^2(Q\setminus\mathcal{F}_K)}^2$$

and estimate the first expectation by similar arguments as those used in the previous proof. The second expectation is bounded via the obvious inequality

$$\mathbf{E}\left\|f - \widehat{g}_{NL}\right\|_{L^2(Q\setminus\mathcal{F}_K)}^2 = \mathbf{E}\left\|f - \widehat{f}_\varepsilon\right\|_{L^2(Q\setminus\mathcal{F}_K)}^2 \le \mathbf{E}\left\|f - \widehat{f}_\varepsilon\right\|_2^2.$$

An application of Corollary 4.5 concludes the proof. □

Proof of Corollary 4.7. The lower estimate is attained for an image f which contains one small square of side length $2^{1/2}(\log 2)^{1/4}\, 2^{-(l+k)/2}\, l^{1/4}$ in every cube J_K. In this case

$$\mathbf{E}\left\|f - \widehat{g}_{NL}\right\|_2^2 \ge c\, 2^{-(l-k)}\, l^{1/2}$$

and

$$\|f\|_{\alpha,\tau,1}^{\tau} \le c\, 2^{2k-(l+k)(2-\alpha\tau)/2}\, l^{1/2}.$$

Therefore

$$\mathbf{E}\left\|f - \widehat{g}_{NL}\right\|_2^2 \ge c\, 2^{-(l+k)\alpha\tau/2} \|f\|_{\alpha,\tau,1}^{\tau}. \qquad \square$$

References

Coifman, R., Meyer, Y., and Wickerhauser, M. V., *Wavelet analysis and signal processing. Wavelets and their Applications*, Jones and Batlett Publishers, Boston, 1992.

DeVore, R. A., and Popov, V., *Free multivariate splines*, Constr. Approx. **3** (1987), no. 2, 239–248.

DeVore, R. A., and Popov, V., *Interpolation of Besov spaces*, Trans. American Math. Soc. **305** (1988), 397–414.

DeVore, R. A., and Lucier, B. J., *Wavelets*, Acta Numerica **1** (1991), 1–56.

Donoho, D. L., *De–noising by soft–thresholding*, Tech. Report Dep. Statistics, Stanford univ., 1992.

Donoho, D. L., and Johnstone, I. M., *Minimax estimation by wavelet shrinkage*, Tech. Report Dep. Statistics, Stanford univ., 1992.

Donoho, D. L., and Johnstone, I. M., *Minimax risk over l_p-balls for l_q-error*, vol. 99, Probab. Theory Related Fields, 1994a, pp. 277-303.

Donoho, D. L., and Johnstone, I. M., *Ideal denoising in an orthonormal basis chosen from a library of bases*, Comptes Rendus—Academie des Sciences Paris. Serie 1 **319** (1994b), no. 12.

Donoho, D. L., Johnstone, I.M., Kerkyacharian, G., and Picard, D., *Wavelet shrinkage: asymptopia? (with discussion)*, J. R. Statist. Soc. B **57** (1995), no. 2, 301–369.

Götze, F., *Rate of convergence of simulated annealing processes*, Preprint 92-008, SFB 343, University of Bielefeld.

Götze, F., and Zalesky, B. A., *Nonparametric wavelet based minimax estimation in nongaussian models*, Preprint 95-118, SFB 343, University of Bielefeld.

Greig, D.M., Porteous, B.T., and Seheult, A.H., *Exact maximum a posteriori estimation for binary images*, J. R. Statist. Soc. B **51** (1989), no. 2, 271–279.

Hoeffding, W., *Probability inequalities for sums of bounded random variables*, J. Amer. Statist. Assoc. **58** (1963), no. 301, 13–30.

Kerkyacharian, G., and Picard, D., *Density estimation in Besov spaces*, Stat. and Probab. Letters **13** (1992), 15–24.

Petrov, V. V., *Sums of independent random variables*, Springer-Verlag, 1975.

Friedrich Götze
Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld 1
Germany
E-mail address: goetze@mathematik.uni-bielefeld.de

Boris Zalesky
Fakultät für Mathematik, SFB 343
Universität Bielefeld
Postfach 100131
33501 Bielefeld 1
Germany
E-mail address: borisz@mathematik.uni-bielefeld.de