Diplomarbeit

Dichteschätzung mit Wavelets

eingereicht von
Michael Diether

Fachbereich Physik, Mathematik und Informatik
Johannes Gutenberg-Universität Mainz

Mainz, im Dezember 2008

Vorwort

Wavelets sind ein verhältnismäßig junges Konzept in der Mathematik. Die Theorie wurde unter anderen von Yves Meyer und Ingrid Daubechies in den 1980er Jahren entwickelt. Gerade in der Approximation von Funktionen (insbesondere von gestörten Signalen in der Informationsübertragung) zeigt sich die Überlegenheit von Wavelets gegenüber älteren Methoden, wie etwa der Fourierentwicklung. Auch in der nichtparametrischen Statistik haben Wavelets Anwendung gefunden. Dies ist auch der Anwendungsbereich, auf den in dieser Diplomarbeit eingegangen wird.

Diese Diplomarbeit ist eine detaillierte Ausarbeitung des Artikels „Density Estimation by Wavelet Thresholding“ von David Donoho, Iain Johnstone, Gérard Kerkyacharian und Dominique Picard, [4]. Darin wird das Problem der Dichteschätzung mit einer nichtparametrischen Grundgesamtheit (so genannten Besovräumen) untersucht, das auf verschiedene Weisen behandelt werden kann. Das zentrale Ergebnis des Artikels ist die Angabe einer Konvergenzgeschwindigkeit für einen speziellen nichtlinearen Waveletschätzer, die asymptotisch für eine große Zahl von Beobachtungen gezeigt wird. Dies ist auch das Hauptaugenmerk dieser Diplomarbeit und wird in Kapitel 2 ausgeführt.

Es hat sich gezeigt, dass der Beweis, der in [4] für diesen zentralen Satz angegeben wird, einen Fehler enthält. In Kapitel 2 wurde dieser Fehler behoben, indem die Behauptung abgeschwächt und der Beweis aus [4] detailliert ausgearbeitet wurde. Eine andere Variante wird in Abschnitt 2.5 angegeben.

Die weiteren Ergebnisse des Artikels [4], beispielsweise die Angabe anderer Waveletschätzer und deren optimale Konvergenzrate, werden in ihrer Beweisidee skizziert und in Verbindung mit dem oben erwähnten zentralen Satz gestellt. Bei diesen übrigen Sätzen sollen weniger die technischen Beweisdetails, als vielmehr die Interpretation im Vordergrund stehen. Dies ist der Inhalt des Kapitels 3.

Um die Aussagen der Sätze formulieren zu können, sind Ergebnisse aus der Funktionalanalysis, namentlich der Theorie der Wavelets und der Besovräume, notwendig. Die benötigten Sätze sind in Kapitel 1 aufgeführt.

Mainz, im Dezember 2008 Michael Diether

Inhaltsverzeichnis

Symbolverzeichnis

$\mathbb{N}$	Menge der natürlichen Zahlen
$\mathbb{Z}$	Menge der ganzen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{R}^+$	Menge der positiven reellen Zahlen
$\lvert M\rvert$	Lebesguemaß der Menge $M \subset \mathbb{R}$ oder Mächtigkeit der Menge M
$\lvert a\rvert$	Absolutbetrag der reellen oder komplexen Zahl a
$\lVert f\rVert_p$	$\left(\int_0^1 \lvert f(x)\rvert^p \, dx\right)^{\frac{1}{p}}$
$\lVert f\rVert_{L^p}$	$\left(\int_{\mathbb{R}} \lvert f(x)\rvert^p \, dx\right)^{\frac{1}{p}}$
$\lVert f\rVert_\infty$	$\sup_{x\in\mathbb{R}} \lvert f(x)\rvert$
$\lVert f\rVert_{\sigma pq}$	Norm im Besovraum $B_{\sigma pq}$, $\lVert\alpha_{0\bullet}\rVert_{\ell^p} + \left(\sum_{j\geqslant 0}\left(2^{j(\sigma+\frac{1}{2}-\frac{1}{p})}\lVert\beta_{j\bullet}\rVert_{\ell^p}\right)^q\right)^{\frac{1}{q}}$
$\lVert a_\bullet\rVert_{\ell^p}$	$\left(\sum_{n=0}^{\infty} \lvert a_n\rvert^p\right)^{\frac{1}{p}}$
$\lVert a_\bullet\rVert_{\ell^\infty}$	$\sup_{n\in\mathbb{N}} \lvert a_n\rvert$
$\mathbb{1}_M(\cdot)$	Indikatorfunktion der Menge M
a^+	Positivteil der reellen Zahl a; $a^+ := \max(a, 0)$
$a \gg 0$	a ist „genügend groß“ und nicht-negativ
$a_n \asymp b_n$	„a_n wächst wie b_n“, d. h. $\exists c, C > 0 \,:\, c \leqslant \frac{a_n}{b_n} \leqslant C$ für $n \in \mathbb{N}$, $n \gg 0$
$a \wedge b$	$\min\{a, b\}$
$a \vee b$	$\max\{a, b\}$
α_{jk}, β_{jk}	Koeffizienten einer Wavelet-Entwicklung
$\hat{\alpha}_{jk}$, $\hat{\beta}_{jk}$	empirische Koeffizienten einer Wavelet-Entwicklung

$\tilde{\beta}_{jk}$	empirische Koeffizienten, auf die Thresholding angewendet wurde
$B_{\sigma pq}$	Besovraum, Menge der Funktionen mit $\|f\|_{\sigma pq} < \infty$
C	Konstante aus $\mathbb{R}^+$, die bei jedem Auftreten verschieden sein kann
$\mathcal{C}^r$	Menge der r-mal stetig differenzierbaren Funktionen
$\mathscr{C}_L$	Menge der linearen Dichteschätzer, vgl. (3.1)
$\mathcal{D}_{\sigma pq}(M)$	$\left\{f\colon \int_{\mathbb{R}} f(x)dx = 1,\ f \geqslant 0,\ \|f\|_{\sigma pq} \leqslant M\right\}$
$\mathcal{D}_{\sigma pq}(M,T)$	$\{f \in \mathcal{D}_{\sigma pq}(M)\colon \operatorname{supp} f \subset [-T,T]\}$
$D_j(f)$	Projektion auf den Raum W_j, vgl. (1.4)
$E_0(f)$	Projektion auf den Raum V_0, vgl. (1.3)
Eg	Erwartungswert der messbaren Abbildung g
ε	vgl. (2.5)
$\hat{f}$	Dichteschätzer
i.i.d.	unabhängig identisch verteilt (independent and identically distributed)
$J_{\sigma pq}(f)$	Norm im Besovraum $B_{\sigma pq}$, $\|E_0(f)\|_{L^p} + \left(\sum_{j\geqslant 0}\left(2^{j\sigma}\|D_j(f)\|_{L^p}\right)^q\right)^{\frac{1}{q}}$
$J'_{\sigma pq}(f)$	alternative Notation für $\|f\|_{\sigma pq}$
$L^p(\mathbb{R})$	Menge aller messbaren Funktionen $f:\mathbb{R}\to\mathbb{R}$ mit $\|f\|_{L^p} < \infty$
ℓ^p	Menge aller Folgen $(a_n)_{n\in\mathbb{N}}$ mit $\|a_n\|_{\ell^p} < \infty$
$\log n$	natürlicher Logarithmus von $n \in \mathbb{R}^+$
$\log_2 n$	Logarithmus von $n \in \mathbb{R}^+$ zur Basis 2
$\mathcal{N}(\mu,\sigma^2)$	Normalverteilung mit Erwartungswert $\mu \in \mathbb{R}$ und Varianz $\sigma^2 > 0$
R_n	Verlustfunktion, hier meist $E\left\|\hat{f} - f\right\|_{L^p}$
$S(x)$	vgl. (2.19)
$\operatorname{span}\{v_i\colon i \in I\}$	Vektorraum, der von den v_i, $i \in I$, erzeugt wird
$\operatorname{supp} f$	Träger der Funktion f; $\operatorname{supp} f := \overline{\{x \in \mathbb{R}\colon f(x) \neq 0\}}$
ξ	vgl. (2.19)

Kapitel 1

Grundlagen

1.1 Wavelets

In diesem und im nächsten Abschnitt sind die für das Folgende benötigten Aussagen zusammengestellt. Auf Beweise soll hier verzichtet werden, sie finden sich in den Standardwerken [12], [2], sowie in [9] und [8].

1.1 Definition: Es sei $\varphi \in L^2(\mathbb{R})$ eine reellwertige Funktion und $\varphi_{jk}(x) := 2^{\frac{j}{2}}\varphi\big(2^j x - k\big)$ für festes $j \in \mathbb{Z}$. Die Funktionenfolge $(\varphi_{jk})_{k\in\mathbb{Z}}$ sei orthonormal in $L^2(\mathbb{R})$ und es bezeichne $V_j := \operatorname{span}\big\{\varphi_{jk} \colon k \in \mathbb{Z}\big\}$. Sei schließlich $V_j \subset V_{j+1}$ für alle $j \in \mathbb{Z}$.

Dann heißt φ *Vater-Wavelet* oder *Skalenfunktion* und die Folge der Räume $(V_j)_{j\in\mathbb{Z}}$ eine *Multiskalenanalyse.* ❖

Die Konstruktion solcher Funktionen φ wird ausführlich in [9, Chapter 5-7] beschrieben. Im Folgenden soll φ stets eine $\mathcal{C}^l$-Funktion mit kompaktem Träger sein, $\operatorname{supp}\varphi \subset [-A, A]$. Die Multiskalenanalyse heißt in diesem Fall *l-regulär.*

1.2 Satz und Definition: Seien φ eine Skalenfunktion und $(V_j)_{j\in\mathbb{Z}}$ eine l-reguläre Multiskalenanalyse. Sei $W_j \subset L^2(\mathbb{R})$ das orthogonale Komplement von V_j in V_{j+1}, also $V_{j+1} = V_j \oplus W_j$. Dann gibt es eine $\mathcal{C}^l$-Funktion $\psi : \mathbb{R} \to \mathbb{R}$ mit kompaktem Träger, sodass die Folge $(\psi_{0k})_{k\in\mathbb{Z}}$ eine Orthonormalbasis von W_0 ist und die Funktionen $(\psi_{jk})_{j,k\in\mathbb{Z}}$ eine Orthonormalbasis von $L^2(\mathbb{R})$ bilden.

Dabei ist analog zu oben $\psi_{jk}(x) := 2^{\frac{j}{2}}\psi\big(2^j x - k\big)$.

Dieses ψ heißt *Mutter-Wavelet* oder einfach *Wavelet.* ❖

Für die Konstruktion von ψ sei wieder auf [9] verwiesen. Bei gegebenem φ ist ψ im Allgemeinen nicht eindeutig bestimmt.

Für ein fest gewähltes $j_0 \in \mathbb{Z}$ besteht die Zerlegung

$$L^2(\mathbb{R}) = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots . \tag{1.1}$$

Insbesondere gibt es für alle $f \in L^2(\mathbb{R})$ die eindeutige Darstellung

$$f = \sum_{k\in\mathbb{Z}} \alpha_{j_0k}\varphi_{j_0k} + \sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk},$$

wobei

$$\alpha_{jk} = \int_{\mathbb{R}} f(x)\varphi_{jk}(x)dx, \qquad \beta_{jk} = \int_{\mathbb{R}} f(x)\psi_{jk}(x)dx. \tag{1.2}$$

Meist wird $j_0 = 0$ gewählt, in dieser Arbeit ist j_0 immer nichtnegativ. Es gilt stets

$$\int_{\mathbb{R}} \varphi(x)dx = 1, \qquad \int_{\mathbb{R}} \psi(x)dx = 0.$$

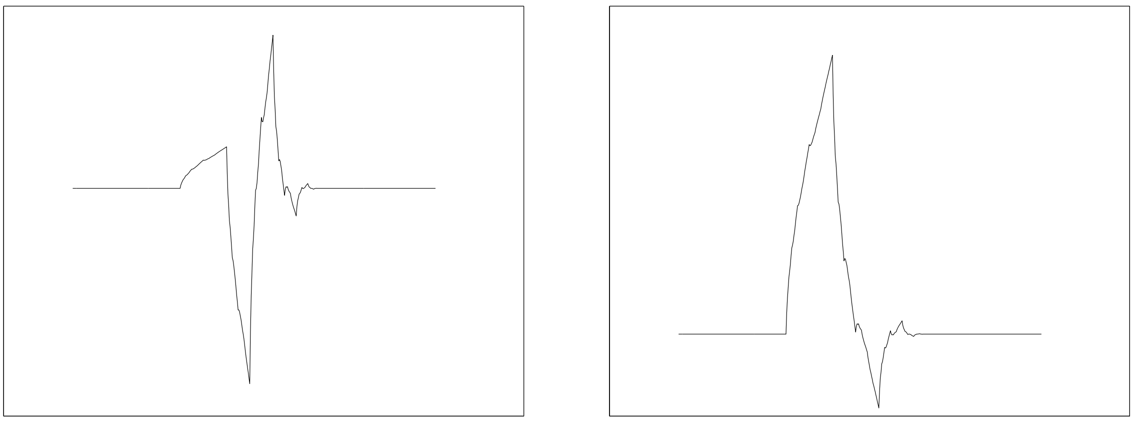
1.3 Beispiel (Haar- und Daubechies-Wavelets, vgl. [2, Chap. 6-7]): Das historisch erste und einfachste Wavelet ist das Haar-Wavelet mit

$$\varphi(x) = \mathbb{1}_{[0,1]}(x), \qquad \psi(x) = \begin{cases} 1 & \text{für } 0 \leqslant x < \frac{1}{2}, \\ -1 & \text{für } \frac{1}{2} \leqslant x \leqslant 1. \end{cases}$$

Es gibt keine $\mathcal{C}^\infty$-Wavelets mit kompaktem Träger. Für jedes $l \in \mathbb{N}$ existiert jedoch eine l-reguläre Multiskalenanalyse, deren zugehörige Funktionen φ und ψ kompakten Träger haben.
Es gibt mehrere verschiedene Familien von Wavelets mit diesen Eigenschaften. Die erstmals von Daubechies konstruierten Wavelets $D2N$, $N \in \mathbb{N}$, sind eine davon. Für $D2N$ gilt

$$\text{supp}\,\varphi \subset [0, 2N-1], \qquad \text{supp}\,\psi \subset [-N+1, N].$$

Für $N \geqslant 3$ sind die Daubechies-Wavelets stetig differenzierbar.
Das Haar-Wavelet entspricht in Daubechies' Konstruktion dem Fall $N = 1$. ❖



(a) Wavelet (b) Skalenfunktion

Abb. 1: Das Daubechies-Wavelet $D4$ ($N = 2$). Es ist stetig, aber nicht differenzierbar.

1.2 Besovräume

Viele der in den folgenden Kapiteln bewiesenen Abschätzungen sind Aussagen über Funktionen in Besovräumen. Diese Räume sind hier deshalb von Interesse, weil sich die in ihnen enthaltenen Funktionen allein mit Hilfe ihrer Wavelet-Koeffizienten beschreiben lassen. Dies ist etwa für Sobolevräume nicht der Fall. Außerdem erfüllen Funktionen aus Besovräumen Optimalitätseigenschaften in Bezug auf lineare Minimax-Schätzer (vgl. [11]). Darauf soll hier jedoch nicht näher eingegangen werden.

Beweise zu den angeführten Aussagen finden sich in [12], [9] und [1].

1.4 Definition: Der *Besovraum* $B_{\sigma pq}$ mit den Parametern $\sigma > 0$, $1 \leqslant p, q \leqslant \infty$ sei die Menge aller Funktionen $f : \mathbb{R} \to \mathbb{R}$, für die gilt

$$\|f\|_{\sigma pq} := \|\alpha_{0\bullet}\|_{\ell^p} + \left(\sum_{j \geqslant 0} \left(2^{j(\sigma + \frac{1}{2} - \frac{1}{p})} \|\beta_{j\bullet}\|_{\ell^p}\right)^q \right)^{\frac{1}{q}} < \infty.$$

Dabei sind $(\alpha_{0k})_{k \in \mathbb{Z}}$ und $(\beta_{jk})_{j,k \in \mathbb{Z}}$ die Koeffizienten in der Waveletentwicklung von f. Oft wird anstelle von $\|f\|_{\sigma pq}$ auch $J'_{\sigma pq}(f)$ geschrieben. ❖

Der Raum $B_{\sigma pq}$ ist dabei unabhängig von den verwendeten Skalen- und Waveletfunktionen. Eine alternative Definition von $B_{\sigma pq}$ ist mit der Norm

$$J_{\sigma pq}(f) := \|E_0(f)\|_{L^p} + \left(\sum_{j \geqslant 0} \left(2^{j\sigma} \|D_j f\|_{L^p}\right)^q \right)^{\frac{1}{q}}$$

möglich. Dabei sind E_i bzw. D_j die Projektionen auf die Räume V_i bzw. W_j (vgl. (1.2)):

$$E_i f(x) = \sum_{k \in \mathbb{Z}} \alpha_{ik} \varphi_{ik}(x) = \sum_{k \in \mathbb{Z}} \int_{\mathbb{R}} \varphi_{ik}(y) f(y) dy \, \varphi_{ik}(x), \tag{1.3}$$

$$D_j f(x) = \sum_{k \in \mathbb{Z}} \beta_{jk} \psi_{jk}(x) = \sum_{k \in \mathbb{Z}} \int_{\mathbb{R}} \psi_{jk}(y) f(y) dy \, \psi_{jk}(x). \tag{1.4}$$

Die Normen $J_{\sigma pq}$ und $J'_{\sigma pq}$ sind äquivalent (vgl. etwa [9, Cor. 9.1]).

1.5 Satz: *Sei g eine Skalen- oder Waveletfunktion, $\vartheta_g(x) = \sum_{k \in \mathbb{Z}} |g(x-k)|$. Sei weiter*

$$f(x) = \sum_{k \in \mathbb{Z}} \lambda_k 2^{\frac{j}{2}} g\big(2^j x - k\big),$$

$1 \leqslant p \leqslant \infty$ und q so gewählt, dass $\frac{1}{p} + \frac{1}{q} = 1$. Dann gilt

$$c_1 2^{j(\frac{1}{2} - \frac{1}{p})} \|\lambda\|_{\ell^p} \leqslant \|f\|_{L^p} \leqslant c_2 2^{j(\frac{1}{2} - \frac{1}{p})} \|\lambda\|_{\ell^p},$$

mit $c_1 = \left(\|\vartheta_g\|_1^{\frac{1}{q}} \|\vartheta_g\|_\infty^{\frac{1}{p}} \right)^{-1}$ und $c_2 = \|\vartheta_g\|_p$.

Der Beweis findet sich in [9, Prop. 8.3].
Man beachte dabei den Unterschied zwischen den Normen $\|\cdot\|_p$ und $\|\cdot\|_{L^p}$, $p \in [1, \infty)$:

$$\|f\|_p := \left(\int_0^1 |f(x)|^p \, dx \right)^{\frac{1}{p}} \qquad \|f\|_{L^p} := \left(\int_{\mathbb{R}} |f(x)|^p \, dx \right)^{\frac{1}{p}} \qquad \|f\|_\infty := \sup_{x \in \mathbb{R}} |f(x)| \, .$$

Die Norm $\|\cdot\|_{\ell^p}$ ist die übliche Norm für Folgenräume. Satz 1.5 wird im Folgenden stets dann benötigt, wenn zwischen den $\|\cdot\|_{L^p}$ und $\|\cdot\|_{\ell^p}$ gewechselt werden muss. Der Satz besagt gerade, dass die beiden Normen äquivalent sind, wenn sie auf Funktionen angewendet werden, die in einer Waveletbasis entwickelbar sind.

1.6 Satz (Besov-Einbettungen): *Es bestehen folgende Inklusionen:*

$$\begin{aligned}
B_{\sigma' p q'} &\subset B_{\sigma p q} && \text{für } \sigma' > \sigma \text{ oder für } \sigma' = \sigma \text{ und } q' \leqslant q. \\
B_{\sigma p q} &\subset B_{\sigma' r q} && \text{für } r > p \text{ und } \sigma' = \sigma - \frac{1}{p} + \frac{1}{r}. \\
B_{\sigma p q} &\subset B_{\sigma' \infty \infty} && \text{für } \sigma - \frac{1}{p} > 0 \text{ und } q > 1 \text{ oder für } \sigma - \frac{1}{p} \geqslant 0 \text{ und } q = 1. \\
B_{0 r (r \wedge 2)} &\subset L^r && \text{für } r \geqslant 1,
\end{aligned}$$

Dabei ist B_{0pq} definiert durch die Besovnorm $J'_{\sigma pq}$, mit $\sigma = 0$.
Außerdem gelten in diesen Räumen Ungleichungen für die zugehörigen Normen:

$$\begin{aligned}
\|f\|_{\sigma' p q'} &\geqslant \|f\|_{\sigma p q} && \text{für } \sigma' > \sigma \text{ oder für } \sigma' = \sigma \text{ und } q' \leqslant q. \\
\|f\|_{\sigma p q} &\geqslant \|f\|_{\sigma' r q} && \text{für } r > p \text{ und } \sigma' = \sigma - \frac{1}{p} + \frac{1}{r}. \\
\|f\|_{\sigma p q} &\geqslant \|f\|_{\sigma' \infty \infty} && \text{für } \sigma - \frac{1}{p} > 0 \text{ und } q > 1 \text{ oder für } \sigma - \frac{1}{p} \geqslant 0 \text{ und } q = 1. \\
\|f\|_{0 r (r \wedge 2)} &\geqslant \|f\|_{L^r} && \text{für } r \geqslant 1.
\end{aligned}$$

Beweis. Die Aussagen werden etwa in [9, Cor. 9.2] bewiesen. Hier soll nur exemplarisch gezeigt werden, dass $B_{\sigma pq} \subset B_{\sigma'' \infty q}$ für $\sigma'' = \sigma - \frac{1}{p}$ gilt; dies ist ein Spezialfall der Aussage $B_{\sigma pq} \subset B_{\sigma' rq}$ mit $r = \infty$.
Für alle Folgen $(a_n)_{n \geqslant 0}$ gilt die bekannte Abschätzung

$$\|a_n\|_{\ell^\infty} = \sup_{n \in \mathbb{N}} |a_n| \leqslant \left(\sum_{n \in \mathbb{N}} |a_n|^p \right)^{\frac{1}{p}} = \|a_n\|_{\ell^p}$$

und daher folgt für $f \in B_{\sigma pq}$

$$\begin{aligned}
\|f\|_{\sigma'' \infty q} &= \|\alpha_{0\bullet}\|_{\ell^\infty} + \left(\sum_{j \geqslant 0} \left(2^{j(\sigma'' + \frac{1}{2})} \|\beta_{j\bullet}\|_{\ell^\infty} \right)^q \right)^{\frac{1}{q}} \\
&\leqslant \|\alpha_{0\bullet}\|_{\ell^p} + \left(\sum_{j \geqslant 0} \left(2^{j(\sigma + \frac{1}{2} - \frac{1}{p})} \|\beta_{j\bullet}\|_{\ell^p} \right)^q \right)^{\frac{1}{q}} = \|f\|_{\sigma pq} < \infty,
\end{aligned}$$

Also gilt $\|f\|_{\sigma pq} \geqslant \|f\|_{\sigma'' \infty q}$ und $f \in B_{\sigma'' \infty q}$. ▨

Kapitel 2

Waveletschätzer mit Thresholding

2.1 Problemstellung

Es seien $X_1, \dots, X_n$ unabhängige und identisch verteilte Zufallsvariablen mit Dichtefunktion $f(x)$. Diese Dichtefunktion soll unter der Voraussetzung geschätzt werden, dass sie in einem Besovraum $B_{\sigma pq}$, $\sigma > \frac{1}{p}$, $p, q \geqslant 1$, liegt und kompakten Träger hat. Hierzu setzt man die empirischen Wavelet-Koeffizienten von f mit

$$\hat{\alpha}_{j_0 k} = \frac{1}{n}\sum_{i=1}^{n} \varphi_{j_0 k}(X_i), \qquad \hat{\beta}_{jk} = \frac{1}{n}\sum_{i=1}^{n} \psi_{jk}(X_i)$$

und die Dichte f selbst mit

$$f(x) = \sum_{k\in\mathbb{Z}} \alpha_{j_0 k}\varphi_{j_0 k}(x) + \sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x), \qquad \operatorname{supp} f \subset [-T, T]$$

an. Dabei seien φ und ψ Skalen- bzw. Waveletfunktion einer l-regulären Multiskalenanalyse, $l \in \mathbb{N}$, mit kompaktem Träger. Der Schätzer $\hat{f}$ hat dann die Gestalt

$$\hat{f}(x) = \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_0 k}\varphi_{j_0 k}(x) + \sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}} \hat{\beta}_{jk}\psi_{jk}(x). \tag{2.1}$$

Der Fehler wird in einer L^r-Norm angegeben:

$$R_n(\hat{f}, f) = E\left\|\hat{f} - f\right\|_{L^r}, \qquad 1 \leqslant r \leqslant \infty.$$

E steht hier für den Erwartungswert, der bezüglich der „wahren" Dichte f gebildet wird. Dies ist nicht mit der Projektionsabbildung E_{j_0} zu verwechseln.

Um $\hat{f}$ zu verbessern, werden die Koeffizienten $\hat{\beta}_{jk}$ später bei einem festen Wert abgeschnitten („*hartes Thresholding*") oder in ihrem Absolutbetrag verringert („*weiches Thresholding*"). In Ermangelung einer treffenden Übersetzung wird im Folgenden für diese Technik stets der Begriff *Thresholding* verwendet. Für einen festen Schwellenwert („*threshold*") $\lambda > 0$ wird für das harte Thresholding

$$\tilde{\beta}_{jk} := \delta_h\big(\hat{\beta}_{jk}, \lambda\big) := \hat{\beta}_{jk}\mathbb{1}_{\{|\hat{\beta}_{jk}|>\lambda\}}, \tag{2.2}$$

und für das weiche Thresholding

$$\tilde{\beta}_{jk} := \delta_s\big(\hat{\beta}_{jk}, \lambda\big) := \mathrm{sgn}\hat{\beta}_{jk}\big(\big|\hat{\beta}_{jk}\big| - \lambda\big)^+ \tag{2.3}$$

definiert.

Schließlich werden geeignete Indizes $j_0, j_1 \in \mathbb{Z}$ zu bestimmen sein, sodass in (2.1) die Summe $\sum_{j \geqslant j_0}$ zu einer endlichen Summe $\sum_{j=j_0}^{j_1}$ wird. Die über k indizierten Summen sind hier stets endlich, da die Wavelet-Funktionen φ und ψ mit kompaktem Träger im Intervall $[-A, A]$ vorausgesetzt werden.

Insgesamt wird also im Folgenden ein spezieller nichtlinearer Schätzer betrachtet, ein Waveletschätzer mit (hartem) Thresholding:

$$\mathrm{TW}(x) = \sum_{k \in \mathbb{Z}} \hat{\alpha}_{j_0 k}\varphi_{j_0 k}(x) + \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \tilde{\beta}_{jk}\psi_{jk}(x), \tag{2.4}$$

wobei für den Schwellenwert aus (2.2) mit einer Konstanten K, die in Satz 2.16 näher bezeichnet wird,

$$\tilde{\beta}_{jk} = \begin{cases} \hat{\beta}_{jk} & \text{für } |\hat{\beta}_{jk}| > K\sqrt{\frac{j}{n}}, \\ 0 & \text{sonst,} \end{cases}$$

gelten soll. Man beachte, dass nach Definition der empirischen Koeffizienten $\hat{\alpha}_{jk}$ und $\tilde{\beta}_{jk}$ der Schätzer TW auch von der Zahl n der Beobachtungen abhängt. Um die Notation zu entlasten, wird dieses n nicht explizit notiert. Die Funktionen φ und ψ sollen kompakten Träger im Intervall $[-A, A]$ haben.

Die Funktionenräume, die im Folgenden betrachtet werden, werden von nun an stets bezeichnet mit

$$\begin{aligned} \mathcal{D}_{\sigma pq}(M) &:= \Big\{f\colon \int_{\mathbb{R}} f(x)dx = 1,\ f \geqslant 0,\ \|f\|_{\sigma pq} \leqslant M\Big\}, \\ \mathcal{D}_{\sigma pq}(M, T) &:= \big\{f \in \mathcal{D}_{\sigma pq}(M)\colon \mathrm{supp}\, f \subset [-T, T]\big\}. \end{aligned}$$

Die Grundidee bei diesem Schätzer TW ist, mit einem linearen Anteil zu starten (die erste Summe, vgl. zur Definition von linearen Schätzern auch (3.1)) und diesen linearen Schätzer zu verbessern, indem man einige nichtlineare Summanden hinzu nimmt. Später, in Kapitel 3, wird Näheres zu linearen Schätzern ausgeführt. Man kann zeigen, dass ein linearer Waveletschätzer im Allgemeinen nicht optimal sein kann und daher diese Verbesserung durch nichtlineare Anteile notwendig ist. Die Wahl von j_0 ist derart, dass „grobe Züge“ der zu schätzenden Dichtefunktion f richtig dargestellt werden. Die Details und mögliche lokale Fluktuationen werden dann durch weitere Koeffizienten, eben durch den nichtlinearen Anteil, aufgelöst. Damit nicht zu viele kleine Fluktuationen dazukommen, die die Varianz des Schätzers unnötig verstärken würden, werden erst „große“ Waveletkoeffizienten (namentlich solche, die über den Schwellenwert $K\sqrt{\frac{j}{n}}$ hinauskommen) berücksichtigt.

Hierin liegt die Stärke des Waveletschätzers wie überhaupt der Waveletentwicklung von Funktionen: Es ist möglich, lokale Fluktuationen von Funktionen präzise darzustellen, wogegen eine

Fourierreihe eine Änderung der Frequenz einer Funktion nicht oder nur mit sehr vielen Koeffizienten auflösen kann. Die Waveletentwicklung kann auf dem Level j_0 (also nur mit den α_{j_0k}) den groben Verlauf der Funktion darstellen und mit den β_{jk} immer mehr kleine Details hinzufügen. Die Größe der darstellbaren „kleinen Details“ hängt dabei vom Level j ab.

Die obere Grenze für den nichtlinearen Anteil wird so gewählt, dass auch der Biasterm eine gute Konvergenzrate einhalten kann. Die Zahlen j_0 und j_1 werden also, plakativ gesprochen, im Wesentlichen auf die Bedürfnisse der Konvergenz zugeschnitten, um Bias und Varianz des Schätzers nicht allzu groß werden zu lassen.

Zunächst wird in Abschnitt 2.2 der Satz über die Konvergenzgeschwindigkeit des Schätzers TW formuliert und in den Abschnitten 2.3 und 2.4 bewiesen. Der hier dargestellte Beweis folgt den Argumenten in [4]. Der besseren Lesbarkeit und Verständlichkeit wegen wurde die Notation leicht angepasst. In Abschnitt 2.5 werden Verallgemeinerungen dieses Hauptsatzes angegeben, bevor in Kapitel 3 dieses Ergebnis in einen Zusammenhang mit anderen Schätzern und allgemeinen Optimalitätsaussagen gestellt wird.

2.2 Die asymptotische Konvergenzgeschwindigkeit des Schätzers TW

Seien $\sigma > \frac{1}{p}$ und $1 \leqslant p \leqslant r < \infty$. Außerdem seien für alles Folgende definiert:

$$\begin{aligned} \sigma' &:= \sigma - \frac{1}{p} + \frac{1}{r} \\ \alpha &:= \min\left\{ \frac{\sigma}{1+2\sigma}, \frac{\sigma - \frac{1}{p} + \frac{1}{r}}{1 + 2\sigma - \frac{2}{p}} \right\}, \\ \varepsilon &:= \sigma p - \frac{r-p}{2}. \end{aligned} \tag{2.5}$$

2.1 Hauptsatz: *Sei $\sigma > \frac{1}{p}$ und $1 \leqslant p \leqslant r < \infty$, $r \leqslant \frac{1+2\sigma}{\sigma}$. Dann gibt es positive, reelle Konstanten C und K_0, die nicht von n abhängen, sodass für*

$$2^{j_0(n)} \asymp \left(n (\log n)^{\frac{r-p}{p} \mathbb{1}_{\{\varepsilon \geqslant 0\}}} \right)^{1-2\alpha}, \qquad 2^{j_1(n)} \asymp \left(\frac{n}{\log n} \right)^{\frac{\alpha}{\sigma'}} \tag{2.6}$$

und für alle $K \geqslant K_0$ gilt

$$\sup_{f \in \mathcal{D}_{\sigma pq}(M,T)} E \left\| \mathrm{TW} - f \right\|_{L^r}^r \leqslant \begin{cases} C \left(\frac{\log n}{n} \right)^{\alpha r}, & \varepsilon \neq 0, \\ C (\log n)^{\left(\frac{r}{2} - \frac{p}{q}\right)^+} \left(\frac{\log n}{n} \right)^{\alpha r}, & \varepsilon = 0. \end{cases} \tag{2.7}$$

Die Notation $\log n$ bezeichnet hier und im Folgenden stets den natürlichen Logarithmus. Für die asymptotischen Betrachtungen im Beweis ist die Basis des Logarithmus jedoch unerheblich, da sich Logarithmen zu verschiedenen Basen nur um Konstanten unterscheiden. Der Logarithmus von n zur Basis 2, der gelegentlich verwendet wird, ist mit $\log_2 n$ bezeichnet.

Der Schätzer TW konvergiert also im L^r-Sinn gegen die zu schätzende Dichtefunktion f. Die Konvergenzgeschwindigkeit liegt im Wesentlichen bei $\left(\frac{\log n}{n}\right)^{\alpha r}$ und hängt von dem Besovraum ab, in dem f liegt. Nur im Fall $\varepsilon = 0$ und für geeignete Werte von p, q und r (beispielsweise „große Werte“ von q) kommt noch ein zusätzlicher $(\log n)$-Term dazu. Die Konvergenzraten sind in diesem Sinne unstetig in den Koeffizienten σ, p und q, was sich auch in der Konstanten C niederschlägt, die für $\varepsilon \to 0$ und für $\frac{r}{2} - \frac{p}{q} \to 0$ sehr groß werden kann. Auf dieses Verhalten, das in [4] knapp vermerkt wird, soll hier nicht näher eingegangen werden.

Die Forderung (2.6) führt in den meisten Fällen dazu, dass tatsächlich $j_0(n) \leqslant j_1(n)$ gilt und also im Schätzer TW auch Waveletkoeffizienten mit Thresholding verwendet werden (vgl. Satz 2.27). Nur wenn $r = p$ (und folglich $\varepsilon > 0$) gilt, kann die entsprechende Summe in TW leer sein und daher wegfallen. Näheres zu diesen Fällen wird in Kapitel 3 angegeben werden. Meist wird der Einfachheit halber statt $j_0(n)$ nur j_0 geschrieben (analog für $j_1(n)$), die ausführlichere Schreibweise dient nur zur Verdeutlichung.

Für die Aussage des Hauptsatzes ist die Voraussetzung $r \geqslant p$ unwesentlich. Da hier nur Funktionen mit kompaktem Träger betrachtet werden, ist die Abbildung $r \mapsto \|f\|_{L^r}$ nach der Hölderschen Ungleichung monoton wachsend (vgl. [7, Satz IV.2.10]). Folglich gilt auch für $r < p$ die Schranke, die für $r = p$ in (2.7) angegeben ist. Um die im Beweis des Hauptsatzes notwendigen Abschätzungen zu zeigen, ist die Voraussetzung $r \geqslant p$ dennoch nötig. Ohne sie wäre eine andere Beweisidee notwendig.

Die hier angegebene Konvergenzrate unterscheidet sich leicht von der, die in [4, Thm. 3] angegeben wird. Dort kommt im Fall $\varepsilon > 0$ ein logarithmischer Term $(\log n)^{-\frac{\alpha\varepsilon r}{\sigma p}}$ hinzu. Allerdings wird sich im Beweis des Satzes zeigen, dass die in [4] angegebene Schranke nicht eingehalten werden kann (vgl. Satz 2.26) und daher eine leichte Abschwächung der Behauptung notwendig ist.

Auch die Voraussetzung $r \leqslant \frac{1+2\sigma}{\sigma}$ ist hier zusätzlich zu [4, Thm. 3] eingefügt. Sie ist zum Beweis einiger der notwendigen Abschätzungen hinzugenommen worden (vgl. etwa Hilfssatz 2.8 und Satz 2.16). Für gewisse σ und p (etwa für „große Werte" von p) kann kein r angegeben werden, das einerseits $r \geqslant p$ und andererseits $r \leqslant \frac{1+2\sigma}{\sigma}$ erfüllt. In diesem Fällen kann der Satz nicht angewendet werden.

Der Ausdruck „C" bezeichnet stets eine positive, reelle Konstante, in der alle Terme gesammelt werden, die nicht von n abhängen. Da hier nur eine asymptotische Aussage für $n \to \infty$ zu zeigen ist, spielt der genaue Wert von C keine Rolle. Insbesondere kann C jedes Mal, wenn es auftritt, verschieden sein. Man beachte dabei, dass j_0 und j_1 nie als Konstanten aufgefasst werden, da sie in (2.6) durch n definiert wurden.

2.3 Vorbereitungen

2.3.1 Elementare Umformungen mit α, ε und σ'

Im Abschnitt 2.3.1 seien stets die Voraussetzungen von Hauptsatz 2.1 erfüllt. Gelegentlich wird zur Verdeutlichung in der Formulierung der folgenden Hilfssätze nochmals auf einige der dort relevanten Voraussetzungen hingewiesen.

2.2 Hilfssatz: *Es gilt*

$$\alpha = \begin{cases} \dfrac{\sigma}{1+2\sigma} & \text{für } \varepsilon \geqslant 0, \\ \dfrac{\sigma'}{1+2\sigma-\frac{2}{p}} & \text{für } \varepsilon \leqslant 0. \end{cases} \tag{2.8}$$

Außerdem ist stets $\alpha \geqslant 0$.

Beweis. Wegen $1+2\sigma > 0$ und $p+2\sigma p - 2 > 1+2-2 = 1 > 0$ gilt

$$\begin{aligned} \alpha = \frac{\sigma}{1+2\sigma} \quad &\Longleftrightarrow \quad \frac{\sigma}{1+2\sigma} \leqslant \frac{\sigma - \frac{1}{p} + \frac{1}{r}}{1+2\sigma-\frac{2}{p}} \\ &\Longleftrightarrow \quad \frac{\sigma}{1+2\sigma} \leqslant \frac{\sigma p - 1 + \frac{p}{r}}{p+2\sigma p - 2} \\ &\Longleftrightarrow \quad \sigma p + 2\sigma^2 p - 2\sigma \leqslant \sigma p - 1 + \frac{p}{r} + 2\sigma^2 p - 2\sigma + \frac{2\sigma p}{r} \\ &\Longleftrightarrow \quad 0 \leqslant p + 2\sigma p - r \\ &\Longleftrightarrow \quad p + 2\sigma p \geqslant r \\ &\Longleftrightarrow \quad \sigma p \geqslant \frac{r-p}{2} \\ &\Longleftrightarrow \quad \varepsilon \geqslant 0. \end{aligned}$$

Also gilt für $\varepsilon \leqslant 0$ aufgrund der gerade gezeigten Äquivalenz $\frac{\sigma}{1+2\sigma} \geqslant \frac{\sigma-\frac{1}{p}+\frac{1}{r}}{1+2\sigma-\frac{2}{p}}$.

Insbesondere gilt für $\varepsilon = 0$, dass $\frac{\sigma}{1+2\sigma} = \frac{\sigma'}{1+2\sigma-\frac{2}{p}}$.

Zu zeigen ist noch, dass unter den Voraussetzungen von Hauptsatz 2.1 stets $\alpha \geqslant 0$ gilt. Da $\sigma > 0$ ist, ist auch $\frac{\sigma}{1+2\sigma} > 0$. Für den Fall $\varepsilon < 0$ ist nach dem bereits Gezeigten

$$\alpha = \frac{\sigma'}{1+2\sigma-\frac{2}{p}} = \frac{\sigma' p}{p+2\sigma p-2}.$$

Wegen $\sigma > \frac{1}{p}$ sind hier Zähler und Nenner positiv. ▨

2.3 Hilfssatz: *Es gilt*

$$\alpha r = \begin{cases} \dfrac{r-p}{2} + \varepsilon(1-2\alpha) & \text{für } \varepsilon \geqslant 0, \\ \dfrac{r-p}{2} + \dfrac{\varepsilon\alpha}{\sigma'} & \text{für } \varepsilon \leqslant 0. \end{cases} \tag{2.9}$$

Beweis. Sei $\varepsilon \geqslant 0$. Dann gilt nach (2.8) $\alpha = \frac{\sigma}{1+2\sigma}$ und damit

$$\begin{aligned}
\frac{r-p}{2} + \varepsilon(1-2\alpha) &= \frac{r-p}{2} + \left(\sigma p - \frac{r-p}{2}\right) - 2\frac{\sigma}{1+2\sigma}\left(\sigma p - \frac{r-p}{2}\right) \\
&= \sigma p - \frac{2\sigma^2 p}{1+2\sigma} + \frac{\sigma(r-p)}{1+2\sigma} \\
&= \frac{\sigma p(1+2\sigma) - 2\sigma^2 p + \sigma r - \sigma p}{1+2\sigma} = \frac{\sigma r}{1+2\sigma} = \alpha r.
\end{aligned}$$

Für $\varepsilon \leqslant 0$ ist $\alpha = \frac{\sigma'}{1+2\sigma-\frac{2}{p}}$ und deshalb

$$\begin{aligned}
\frac{r-p}{2} + \frac{\varepsilon\alpha}{\sigma'} &= \frac{r-p}{2} + \left(\sigma p - \frac{r-p}{2}\right)\frac{\sigma'}{1+2\sigma-\frac{2}{p}}\frac{1}{\sigma'} \\
&= \frac{r-p}{2} + \frac{\sigma p}{1+2\sigma-\frac{2}{p}} - \frac{r-p}{2(1+2\sigma-\frac{2}{p})} \\
&= \frac{r+2\sigma r - \frac{2r}{p} - p - 2\sigma p + 2 + 2\sigma p - r + p}{2(1+2\sigma-\frac{2}{p})} \\
&= \frac{\sigma r - \frac{r}{p} + 1}{1+2\sigma-\frac{2}{p}} = \frac{r(\sigma - \frac{1}{p} + \frac{1}{r})}{1+2\sigma-\frac{2}{p}} = r\alpha.
\end{aligned}$$

▨

2.4 Hilfssatz: *Sei $\varepsilon = 0$. Dann gilt $1-2\alpha = \frac{p}{r}$.*

Beweis. Es gilt die Äquivalenz

$$\begin{aligned}
\varepsilon = 0 \iff \sigma p = \frac{r-p}{2} &\iff \frac{r}{p} - 1 = 2\sigma \\
&\iff \frac{p}{r} = \frac{1}{1+2\sigma} = 1 - \frac{2\sigma}{1+2\sigma} = 1-2\alpha.
\end{aligned}$$

▨

2.5 Hilfssatz: *Ist $\varepsilon > 0$, so gilt $\alpha - \frac{\alpha\varepsilon}{\sigma p} = \frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha$.*
Für $\frac{\alpha}{\sigma p} \geqslant 0$ folgt insbesondere $\alpha \geqslant \frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha$.

Beweis. Dies sieht man mit (2.9) und (2.8) folgendermaßen:

$$\begin{aligned}
\alpha - \frac{\alpha\varepsilon}{\sigma p} = \frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha &\iff -2\alpha\varepsilon = \sigma r - \sigma p - 2\sigma\alpha r \\
&\iff -2\alpha\varepsilon = \sigma r - \sigma p - 2\sigma\left(\frac{r-p}{2} + \varepsilon - 2\alpha\varepsilon\right) \\
&\iff -2\alpha\varepsilon = -2\sigma\varepsilon + 4\alpha\sigma\varepsilon \\
&\iff -\alpha\varepsilon = \sigma\varepsilon(2\alpha - 1) \\
&\iff -\frac{\sigma\varepsilon}{1+2\sigma} = \sigma\varepsilon\left(\frac{2\sigma}{2\sigma+1} - 1\right) = \sigma\varepsilon\frac{-1}{2\sigma+1}.
\end{aligned}$$

In der letzten Zeile steht eine offensichtlich wahre Aussage. ▨

2.6 Hilfssatz: *Sei $\varepsilon > 0$. Dann ist $-\frac{r-p}{p}\varepsilon(1-2\alpha) \leqslant \alpha r$.*

Beweis. Zunächst gilt

$$1 - 2\alpha = 1 - \frac{2\sigma}{1+2\sigma} = \frac{1}{1+2\sigma} = \frac{\alpha}{\sigma} \qquad \text{für alle } \varepsilon \geqslant 0 \tag{2.10}$$

und daher auch $1 - 2\alpha \geqslant 0$. Insbesondere ist wegen $r \geqslant p \geqslant 0$ auch

$$\frac{r-p}{2} + \varepsilon(1-2\alpha)\frac{r}{p} = \frac{r-p}{2} + \varepsilon(1-2\alpha)\left(1 + \frac{r-p}{p}\right) \geqslant 0.$$

Außerdem ist

$$\frac{r-p}{2} + \varepsilon(1-2\alpha)\left(1 + \frac{r-p}{p}\right) \geqslant 0 \quad \Longleftrightarrow \quad -\frac{r-p}{p}\varepsilon(1-2\alpha) \leqslant \frac{r-p}{2} + \varepsilon(1-2\alpha),$$

was wegen (2.9) äquivalent zur Behauptung ist. ▨

2.7 Hilfssatz: *Sei $\varepsilon < 0$. Dann gilt $1 - 2\alpha < \frac{\alpha}{\sigma'}$.*

Beweis. Nach Hilfssatz 2.2 ist $\alpha \geqslant 0$ und wegen $\sigma - \frac{1}{p} > 0$ ist $\sigma' = \sigma - \frac{1}{p} + \frac{1}{r} \geqslant 0$, also

$$\begin{aligned} 1 - 2\alpha < \frac{\alpha}{\sigma'} \iff \frac{\alpha}{1-2\alpha} > \sigma' &\iff \frac{\frac{\sigma'}{1+2\sigma-\frac{2}{p}}}{\frac{1+2\sigma-\frac{2}{p}-2\sigma'}{1+2\sigma-\frac{2}{p}}} > \sigma' \\ &\iff \frac{1}{1+2\sigma-\frac{2}{p}-2\sigma'} > 1 \\ &\iff 0 > \sigma - \frac{1}{p} - \sigma' = -\frac{1}{r}, \end{aligned}$$

was wegen $r \geqslant 1$ offensichtlich erfüllt ist. ▨

2.8 Hilfssatz: *Seien $r \leqslant \frac{1+2\sigma}{\sigma}$ und $\sigma > \frac{1}{p}$. Dann ist $\frac{\alpha}{\sigma'} \leqslant 1$.*

Beweis. Es wird eine Fallunterscheidung nach dem Wert von ε durchgeführt.

Sei zunächst $\varepsilon \leqslant 0$. Dann ist nach (2.8) $\alpha = \frac{\sigma'}{1+2\sigma-\frac{2}{p}}$, also $\frac{\alpha}{\sigma'} = \frac{1}{1+2\sigma-\frac{2}{p}}$. Es reicht also zu zeigen, dass $1 + 2\sigma - \frac{2}{p} \geqslant 1$ ist. Es gilt aber

$$\sigma p > 1 \quad \Longrightarrow \quad 2\sigma p > 2 \quad \Longrightarrow \quad 2\sigma - \frac{2}{p} \geqslant 0 \quad \Longrightarrow \quad 1 + 2\sigma - \frac{2}{p} \geqslant 1.$$

Sei nun $\varepsilon > 0$. Dann ist $\alpha = \frac{\sigma}{1+2\sigma}$ und daher $\frac{\alpha}{\sigma'} = \frac{\sigma}{(1+2\sigma)(\sigma-\frac{1}{p}+\frac{1}{r})}$. Es folgt, da $\sigma - \frac{1}{p} > 0$ ist,

$$\begin{aligned} r \leqslant \frac{1+2\sigma}{\sigma} \quad \Longrightarrow \quad \frac{r\sigma}{1+2\sigma} \leqslant 1 \quad &\Longrightarrow \quad \frac{\sigma}{(1+2\sigma)\frac{1}{r}} \leqslant 1 \\ &\Longrightarrow \quad \frac{\sigma}{(1+2\sigma)(\sigma - \frac{1}{p} + \frac{1}{r})} \leqslant 1. \end{aligned}$$

▨

2.9 Hilfssatz: *Sei $p \leqslant r$ und $\sigma p > 1$. Dann ist $1 - \frac{\varepsilon}{\sigma p} \geqslant 0$.*

Beweis. Es gilt nach Definition von ε die Implikationskette

$$p \leqslant r \quad \Longrightarrow \quad \frac{r-p}{2} \geqslant 0 \quad \Longrightarrow \quad \sigma p - \varepsilon \geqslant 0 \quad \Longrightarrow \quad 1 \geqslant \frac{\varepsilon}{\sigma p} \qquad ▨$$

Schließlich ist noch eine Aussage über die Anzahl der Terme zu zeigen, in denen Thresholding verwendet wird.

2.10 Hilfssatz: *Sei $\varepsilon = 0$. Mit der Wahl von j_0 und j_1 wie in (2.6) gilt dann für $n \to \infty$*

$$\frac{j_1(n)}{j_0(n)} \asymp \frac{r}{r-2}.$$

Beweis. Zunächst ist

$$j_0 \asymp (1-2\alpha)\left(\log n + \frac{r-p}{p}\log\log n\right) \qquad j_1 \asymp \frac{\alpha}{\sigma'}\big(\log n - \log\log n\big).$$

Folglich gilt für den Quotienten

$$\frac{j_1}{j_0} \asymp \frac{\alpha}{\sigma'(1-2\alpha)} \cdot \frac{\log n - \log\log n}{\log n + \frac{r-p}{p}\log\log n}.$$

Der zweite Faktor konvergiert nach der Regel von de l'Hospital gegen 1, für den ersten Faktor gilt aber

$$\begin{aligned}
\varepsilon = 0 \quad &\Longleftrightarrow \quad \frac{r-p}{2} = \sigma p \\
&\Longleftrightarrow \quad r - p = 2\sigma p \\
&\Longleftrightarrow \quad \frac{r}{\sigma p} - \frac{1}{\sigma} = 2 \\
&\Longleftrightarrow \quad r - 2 = r - \frac{r}{\sigma p} + \frac{1}{\sigma} \\
&\Longleftrightarrow \quad \alpha(r-2) = r\left(\sigma - \frac{1}{p} + \frac{1}{r}\right)\frac{\alpha}{\sigma}.
\end{aligned}$$

Wegen (2.10) und nach Definition von σ' ist dies äquivalent zu

$$\alpha(r-2) = r\sigma'(1-2\alpha)$$

also auch zu

$$\frac{\alpha}{\sigma'(1-2\alpha)} = \frac{r}{r-2}.$$

Das war zu zeigen. ▨

2.3.2 Abschätzungen für die Momente

2.11 Hilfssatz: *Es seien $X_1, \ldots, X_n$ i.i.d. Beobachtungen einer (wesentlich) beschränkten Dichtefunktion f, $m \geqslant 1$ und sei $g \in L^m(\mathbb{R})$ eine beschränkte Funktion mit $\int g^2 = 1$. Sei weiter*

$$g_{jk}(x) = 2^{\frac{j}{2}} g\big(2^j x - k\big), \qquad \gamma_{jk} = \int_{\mathbb{R}} g_{jk}(x) f(x) dx, \qquad \hat{\gamma}_{jk} = \frac{1}{n} \sum_{i=1}^{n} g_{jk}(X_i).$$

Dann gibt es eine Konstante $C > 0$, sodass für alle $j, k \in \mathbb{Z}$ gilt

$$E \left| \hat{\gamma}_{jk} - \gamma_{jk} \right|^m \leqslant C n^{-\frac{m}{2}} \left(\|f\|_\infty^{\frac{m}{2}} + 2^{m+1} \|f\|_\infty \|g\|_{L^m}^m \left(\frac{2^j}{n} \right)^{(\frac{m}{2}-1)^+} \right). \tag{2.11}$$

Die Notation legt nahe, dass g eine Waveletfunktion ist. In der Tat wird der Hilfssatz später auf Wavelets angewendet, aber hier wird nicht gefordert, dass die g_{jk} paarweise orthogonal sind.

Für den Beweis wird die Ungleichung von Rosenthal benötigt, die hier nicht gezeigt werden soll. Ein ausführlicher Beweis findet sich bei [9, Thm. C.2].

2.12 Satz (Rosenthal): *Seien $Y_1, \ldots, Y_n$ i.i.d. Zufallsvariablen und $\zeta^2 \in \mathbb{R}^+$ mit $EY_i = 0$ und $EY_i^2 \leqslant \zeta^2 < \infty$. Dann gibt es eine Konstante $C > 0$, sodass*

$$E \left| \frac{1}{n} \sum_{i=1}^{n} Y_i \right|^m \leqslant \begin{cases} C \left(\frac{\zeta^m}{n^{\frac{m}{2}}} + \frac{E|Y_1|^m}{n^{m-1}} \right), & \text{für } m \geqslant 2, \\ \zeta^m n^{-\frac{m}{2}}, & \text{für } 1 \leqslant m \leqslant 2. \end{cases} \tag{2.12}$$

Beweis (von Hilfssatz 2.11). Betrachte die Zufallsvariablen $Y_i := g_{jk}(X_i) - E g_{jk}(X_i)$; offensichtlich ist $EY_i = 0$. Außerdem gilt

$$\begin{aligned} E\left|Y_1\right|^m = E\left|g_{jk}(X_1) - E g_{jk}(X_1)\right|^m &\leqslant 2^m \left(E\left|g_{jk}(X_1)\right|^m + E\left|E g_{jk}(X_1)\right|^m\right) \\ &\leqslant 2^m \cdot 2E\left|g_{jk}(X_1)\right|^m \\ &= 2^{m+1} 2^{m\frac{j}{2}} \int_{\mathbb{R}} \left|g(2^j x - k)\right|^m f(x) dx \\ &\leqslant 2^{m+1} 2^{m\frac{j}{2}} \|f\|_\infty \int_{\mathbb{R}} 2^{-j} \left|g(y)\right|^m dy \\ &= 2^{m+1} 2^{j(\frac{m}{2}-1)} \|f\|_\infty \|g\|_{L^m}^m . \end{aligned}$$

Für die zweiten Momente ergibt sich die bessere Abschätzung

$$\begin{aligned}
\zeta^2 := EY_i^2 = E\big(g_{jk}(X_i) - Eg_{jk}(X_1)\big)^2 &= Eg_{jk}^2(X_1) - E^2 g_{jk}(X_1) \leqslant Eg_{jk}^2(X_1) \\
&= \int_{\mathbb{R}} 2^j \left(g(2^j x - k)\right)^2 f(x)dx \\
&= \int_{\mathbb{R}} \left(g(y-k)\right)^2 f(2^{-j}y) 2^{-j+j} dy \\
&\leqslant \|f\|_\infty \int_{\mathbb{R}} \left(g(y-k)\right)^2 dy \\
&= \|f\|_\infty \|g\|_{L^2}^2 = \|f\|_\infty < \infty.
\end{aligned}$$

Durch Umschreiben erhält man noch

$$\begin{aligned}
E\left|\hat{\gamma}_{jk} - \gamma_{jk}\right|^m &= E\left|\frac{1}{n}\sum_{i=1}^n g_{jk}(X_i) - \int_{\mathbb{R}} g_{jk}(x)f(x)dx\right|^m \\
&= E\left|\frac{1}{n}\sum_{i=1}^n g_{jk}(X_i) - Eg_{jk}(X_1)\right|^m = E\left|\frac{1}{n}\sum_{i=1}^n Y_i\right|^m.
\end{aligned}$$

Satz 2.12 zeigt daher nun die Existenz einer Konstanten C mit

$$\begin{aligned}
E\left|\hat{\gamma}_{jk} - \gamma_{jk}\right|^m &\leqslant \begin{cases} C\left(\frac{\zeta^m}{n^{\frac{m}{2}}} + \frac{E|Y_1|^m}{n^{m-1}}\right) & \text{für } m \geqslant 2, \\ \zeta^m n^{-\frac{m}{2}} & \text{für } 1 \leqslant m \leqslant 2, \end{cases} \\
&\leqslant \begin{cases} C\left(\|f\|_\infty^{\frac{m}{2}} n^{-\frac{m}{2}} + 2^{m+1} 2^{j(\frac{m}{2}-1)} \|f\|_\infty \|g\|_{L^m}^m n^{-m+1}\right), & m \geqslant 2, \\ \|f\|_\infty^{\frac{m}{2}} n^{-\frac{m}{2}}, & 1 \leqslant m \leqslant 2 \end{cases} \\
&\leqslant \begin{cases} Cn^{-\frac{m}{2}}\left(\|f\|_\infty^{\frac{m}{2}} + 2^{m+1} 2^{j(\frac{m}{2}-1)} \|f\|_\infty \|g\|_{L^m}^m n^{-\frac{m}{2}+1}\right), & m \geqslant 2, \\ 1 \cdot n^{-\frac{m}{2}}\left(\|f\|_\infty^{\frac{m}{2}} + 2^{m+1} \|f\|_\infty \|g\|_{L^m}^m\right), & 1 \leqslant m \leqslant 2, \end{cases} \\
&= Cn^{-\frac{m}{2}}\left(\|f\|_\infty^{\frac{m}{2}} + 2^{m+1} \|f\|_\infty \|g\|_{L^m}^m \left(\frac{2^j}{n}\right)^{(\frac{m}{2}-1)^+}\right).
\end{aligned}$$

▨

Ist bekannt, dass $f \in B_{\sigma pq}(M)$, so kann man auch für die Funktionswerte von f eine Abschätzung angeben, die nur von den Besov-Parametern und der Schranke M für die Besov-Norm abhängt. Der folgende Hilfssatz zeigt, dass alle Funktionen in $B_{\sigma pq}(M)$ wesentlich beschränkt sind, und gibt eine (im Allgemeinen nicht scharfe) Schranke an.

2.13 Hilfssatz: *Seien $f \in \mathcal{D}_{\sigma pq}(M)$, $\sigma'' := \sigma - \frac{1}{p}$, $q \geqslant 1$ und $\frac{1}{q} + \frac{1}{q'} = 1$. Dann gibt es eine Konstante $C > 0$, sodass*

$$\|f\|_\infty \leqslant C\big(1 - 2^{-\sigma'' q'}\big)^{-\frac{1}{q'}} \|f\|_{\sigma'' \infty q} \leqslant CM\big(1 - 2^{-\sigma'' q'}\big)^{-\frac{1}{q'}}. \tag{2.13}$$

Beweis. Nach Satz 1.6 ist $\mathcal{D}_{\sigma pq} \subset \mathcal{D}_{\sigma''\infty q}$ und $\|f\|_{\sigma''\infty q} \leqslant \|f\|_{\sigma pq} \leqslant M$, also $f \in \mathcal{D}_{\sigma''\infty q}(M)$. Dies zeigt die zweite Ungleichung.

Für die erste Ungleichung aus der Behauptung ist, da f, φ und ψ beschränkt sind, kompakten Träger haben und daher alle auftretenden Summen über k endlich sind,

$$\begin{aligned}
\|f\|_\infty = \sup_{x\in\mathbb{R}} |f(x)| &= \sup_{x\in\mathbb{R}} \left| \sum_{k\in\mathbb{Z}} \alpha_{0k}\varphi_{0k}(x) + \sum_{j\geqslant 0}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x) \right| \\
&\leqslant C_\varphi \sup_{k\in\mathbb{Z}} |\alpha_{0k}|\, \|\varphi_{0k}\|_\infty + C_\psi \sum_{j\geqslant 0} \sup_{k\in\mathbb{Z}} |\beta_{jk}|\, \|\psi_{jk}\|_\infty \\
&\leqslant C \left(\sup_{k\in\mathbb{Z}} |\alpha_{0k}|\, \|\varphi\|_\infty + \sum_{j\geqslant 0} 2^{\frac{j}{2}} \sup_{k\in\mathbb{Z}} |\beta_{jk}|\, \|\psi\|_\infty \right) \\
&\leqslant C \max\{\|\varphi\|_\infty, \|\psi\|_\infty\} \left(\sup_{k\in\mathbb{Z}} |\alpha_{0k}| + \sum_{j\geqslant 0} \left| 2^{\frac{j}{2}} \sup_{k\in\mathbb{Z}} |\beta_{jk}| \right| \right) \\
&\leqslant C \left(\|\alpha_{0\bullet}\|_{\ell^\infty} + \left\| \left(2^{\frac{j}{2}} \sup_{k\in\mathbb{Z}} |\beta_{jk}|\right)_{j\geqslant 0} \right\|_{\ell^1} \right).
\end{aligned}$$

Man beachte, dass die Konstante C sich während der Rechnung verändert, aber nur von φ und ψ abhängt (zunächst werden die Summen über k durch C_φ bzw. C_ψ ersetzt, da die Träger der Funktionen endlich sind und daher auch nur endlich viele Summanden auftreten; danach werden zusätzlich die Suprema der Funktionen in einer Konstanten C zusammengefasst).

Mit $\eta_j := 2^{\frac{j}{2}} \sup_k |\beta_{jk}|$ kann auf den letzten Ausdruck die Höldersche Ungleichung angewendet werden:

$$\begin{aligned}
\|\eta_j\|_{\ell^1} = \left\| 2^{-j\sigma''} 2^{j\sigma''} \eta_j \right\|_{\ell^1} &\leqslant \left\| 2^{-j\sigma''} \right\|_{\ell^{q'}} \left\| 2^{j\sigma''} \eta_j \right\|_{\ell^q} \\
&= \left(\sum_{j=0}^{\infty} \left(2^{-\sigma'' q'}\right)^j \right)^{\frac{1}{q'}} \left(\sum_{j\geqslant 0} 2^{j\sigma'' q} \left(2^{\frac{j}{2}} \sup_{k\in\mathbb{Z}} |\beta_{jk}| \right)^q \right)^{\frac{1}{q}} \\
&= \left(\frac{1}{1-2^{-\sigma'' q'}} \right)^{\frac{1}{q'}} \left(\sum_{j\geqslant 0} \left(2^{j(\sigma''+\frac{1}{2})} \sup_{k\in\mathbb{Z}} |\beta_{jk}| \right)^q \right)^{\frac{1}{q}}
\end{aligned}$$

Dabei ist auch der letzte Ausdruck endlich, da $f \in \mathcal{D}_{\sigma''\infty q}$ (wäre er nicht endlich, wäre auch $\|f\|_{\sigma''\infty q} = \infty$, vgl. Satz 1.6). Insgesamt gilt also

$$\begin{aligned}
\|f\|_\infty &\leqslant C \left(\|\alpha_{0\bullet}\|_{\ell^\infty} + \left\| \left(2^{\frac{j}{2}} \sup_{k\in\mathbb{Z}} |\beta_{jk}|\right)_{j\geqslant 0} \right\|_{\ell^1} \right) \\
&\leqslant C \left(\|\alpha_{0\bullet}\|_{\ell^\infty} + \left(\frac{1}{1-2^{-\sigma'' q'}} \right)^{\frac{1}{q'}} \left(\sum_{j\geqslant 0} \left(2^{j(\sigma''+\frac{1}{2})} \sup_{k\in\mathbb{Z}} |\beta_{jk}| \right)^q \right)^{\frac{1}{q}} \right)
\end{aligned}$$

$$
\begin{aligned}
&\leqslant C\left(\frac{1}{1-2^{-\sigma'' q'}}\right)^{\frac{1}{q'}}\left(\|\alpha_{0\bullet}\|_{\ell^\infty}+\left(\sum_{j\geqslant 0}\left(2^{j(\sigma''+\frac{1}{2})}\sup_{k\in\mathbb{Z}}|\beta_{jk}|\right)^q\right)^{\frac{1}{q}}\right)\\
&= C\left(1-2^{-\sigma'' q'}\right)^{-\frac{1}{q'}}\|f\|_{\sigma''\infty q}\,.
\end{aligned}
$$

Aus (2.11) und (2.13) folgt nun für alle Dichtefunktionen $f \in \mathcal{D}_{\sigma pq}(M)$ direkt

2.14 Satz: *Sei entweder $m > 2$ und $n \geqslant c2^j$ für eine Konstante $c \in \mathbb{R}^+$ oder $m \in [1,2]$ und $n \in \mathbb{N}$ beliebig. Dann gilt mit der Notation aus Hilfssatz 2.11 für i.i.d. Zufallsvariablen $X_1,\ldots,X_n$ mit der Dichtefunktion $f \in \mathcal{D}_{\sigma pq}(M)$*

$$
E\,|\hat{\gamma}_{jk}-\gamma_{jk}|^m \leqslant Cn^{-\frac{m}{2}}, \tag{2.14}
$$

wobei C nicht von n oder j abhängt.

2.15 Hilfssatz: *Sei $j = j_0$. Unter den Voraussetzungen von Hauptsatz 2.1 gilt $n \geqslant c2^j$ (die Voraussetzung aus Satz 2.14) für alle genügend großen n.*

Beweis. Nach (2.6) gibt es eine Konstante $C > 0$, sodass für alle $n \gg 0$ gilt

$$
2^{j_0} \leqslant C\left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon\geqslant 0\}}}\right)^{1-2\alpha}.
$$

Da α nach Hilfssatz 2.2 unter den Voraussetzungen von Hauptsatz 2.1 stets positiv ist, folgt die Behauptung.

2.3.3 Große Abweichungen

2.16 Satz: *Sei $j_0 \leqslant j \leqslant j_1$ und $2^{j_1} \leqslant \tilde{C}\frac{n}{j_1}$ für ein konstantes $\tilde{C} \in \mathbb{R}^+$. Dann gibt es für alle $\gamma \geqslant 1$ positive Konstanten K, C, die nicht von n oder j abhängen, sodass gilt*

$$
P\left(\left|\hat{\beta}_{jk}-\beta_{jk}\right| > \frac{K}{2}\sqrt{\frac{j}{n}}\right) \leqslant 2C\cdot 2^{-\gamma j}. \tag{2.15}
$$

Man kann hier $K = c\gamma\sqrt{M \vee 1}$ mit einem $c \in \mathbb{R}^+$ wählen, für das $c^2 \geqslant 8\log(2)\left(1+\frac{1}{3}c\,\|\psi\|_\infty\right)$ gilt. Dabei ist M der Parameter der Klasse $B_{\sigma pq}(M)$.

Dieses K ist als das K aus der Behauptung von Hauptsatz 2.1 festzulegen, das den Schwellenwert für die Schätzer $\tilde{\beta}_{jk}$ bestimmt. Welche γ zu verwenden sind, wird in Satz 2.28 und Satz 2.29 näher beschrieben.

Der Satz gibt eine Schranke für das Ereignis an, dass der empirische Waveletkoeffizient $\hat{\beta}_{jk}$ und der wahre Waveletkoeffizient β_{jk} nicht nahe beieinander liegen. Für den späteren Beweis von Hauptsatz 2.1 wird dieser Satz in dem Sinne verwendet, dass „große Fehler des Schätzers selten sind".

Man beachte, dass die Bezeichnung $\log(2)$ wieder den natürlichen Logarithmus von 2 darstellt, nicht etwa die Logarithmusfunktion zur Basis 2.

2.17 Hilfssatz: *Unter den Voraussetzungen von Hauptsatz 2.1 gilt* $2^{j_1} \leqslant \tilde{C}\frac{n}{j_1}$ *(die Voraussetzung von Satz 2.16) für alle genügend großen* n.
Insbesondere ist auch $n \geqslant c2^j$ *(die Voraussetzung aus Satz 2.14) für alle genügend großen* n *erfüllt, sofern* $j_0 \leqslant j \leqslant j_1$ *gilt.*

Beweis. Mit (2.6) gibt es eine Konstante $C > 0$, sodass für alle $n \gg 0$ gilt

$$2^{j_1} \leqslant C\left(\frac{n}{\log n}\right)^{\frac{\alpha}{\sigma'}} \qquad \text{also} \qquad j_1 \leqslant C\frac{\alpha}{\sigma'}\left(\log n - \log\log n\right).$$

Da nach Hilfssatz 2.8 auch $\frac{\alpha}{\sigma'} \leqslant 1$ ist, folgt

$$j_1 2^{j_1} \leqslant C\frac{\alpha}{\sigma'}\left(\log n - \log\log n\right) C\frac{n^{\frac{\alpha}{\sigma'}}}{(\log n)^{\frac{\alpha}{\sigma'}}} \leqslant C\log n\frac{n^{\frac{\alpha}{\sigma'}}}{(\log n)^{\frac{\alpha}{\sigma'}}} \leqslant C\log n\frac{n}{\log n} \leqslant Cn.$$

Zum Beweis von Satz 2.16 wird eine ähnliche Aussage wie die Ungleichung von Rosenthal benötigt, allerdings nicht für Erwartungswerte, sondern für die Wahrscheinlichkeit, dass eine Zufallsvariable große Werte annimmt.

2.18 Satz (Bernstein): *Es seien* $Y_1, \ldots, Y_n$ *beschränkte i.i.d. Zufallsvariablen mit* $EY_i = 0$ *und* $EY_i^2 := \zeta^2 < \infty$ *und* $|Y_i| \leqslant N < \infty$ *für* $i = 1, \ldots, n$. *Dann gilt für alle* $\lambda > 0$:

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} Y_i\right| > \lambda\right) \leqslant 2\exp\left(-\frac{n\lambda^2}{2(\zeta^2 + \frac{1}{3}N\lambda)}\right).$$

Der Beweis ist bei [9, Thm. C.1] zu finden.

2.19 Hilfssatz: *Seien* $X_1, \ldots, X_n$ *i.i.d. Zufallsvariable mit der Dichtefunktion* $f \in \mathcal{D}_{\sigma pq}(M)$ *und sei* $Y_i := \psi_{jk}(X_i) - E\psi_{jk}(X_i)$. *Dann gilt* $\zeta^2 := EY_i^2 \leqslant C\|f\|_\infty \leqslant CM < \infty$ *für eine Konstante* $C > 0$.

Beweis. Die erste Ungleichung folgt wie im Beweis von Hilfssatz 2.11. Die Voraussetzungen von Hilfssatz 2.11 sind hier ebenfalls erfüllt, da ψ eine Waveletfunktion ist. Aus (2.13) folgt $\|f\|_\infty \leqslant CM < \infty$.

Beweis (von Satz 2.16). Ähnlich wie bei der Ungleichung von Rosenthal wird die Ungleichung von Bernstein angewendet auf $Y_i := \psi_{jk}(X_i) - E\psi_{jk}(X_i)$. Man beachte, dass die Y_i dann beschränkt sind, da die Waveletfunktionen ψ_{jk} kompakten Träger haben und stetig sind (die Multiskalenanalyse ist nach Konstruktion von TW l-regulär für ein $l \in \mathbb{N}$). Offensichtlich ist $EY_i = 0$ und nach Hilfssatz 2.19 gilt $EY_i^2 < \infty$.
Genauso wie im Beweis zu Hilfssatz 2.11 geht die linke Seite von (2.15) über in die linke Seite der Bernsteinschen Ungleichung mit $\lambda := \frac{K}{2}\sqrt{\frac{j}{n}}$. Zunächst gilt

$$\begin{aligned} N = \max_{i=1,\ldots,n} |Y_i| &= \max_{i=1,\ldots,n} \left|\psi_{jk}(X_i) - E\big(\psi_{jk}(X_i)\big)\right| \\ &\leqslant \max_{i=1,\ldots,n} |\psi_{jk}(X_i)| + \max_{i=1,\ldots,n} |E\psi_{jk}(X_i)| \\ &\leqslant \|\psi_{jk}\|_\infty + \|\psi_{jk}\|_\infty = 2^{\frac{j}{2}}\|\psi\|_\infty + 2^{\frac{j}{2}}\|\psi\|_\infty = 2\cdot 2^{\frac{j}{2}}\|\psi\|_\infty. \end{aligned} \tag{2.16}$$

Mit der Ungleichung von Bernstein ergibt sich dann

$$\begin{aligned}
P\left(\left|\hat{\beta}_{jk}-\beta_{jk}\right| > \frac{K}{2}\sqrt{\frac{j}{n}}\right) &\leqslant 2\exp\left(-\frac{n\frac{K^2}{4}\frac{j}{n}}{2\left(\zeta^2+2\cdot 2^{\frac{j}{2}}\left\|\psi\right\|_\infty\frac{K}{2}\sqrt{\frac{j}{n}}\frac{1}{3}\right)}\right)\\
&\leqslant 2\exp\left(-\frac{n\frac{K^2}{4}\frac{j}{n}}{2\left(\zeta^2+2\cdot 2^{\frac{j_1}{2}}\left\|\psi\right\|_\infty\frac{K}{2}\sqrt{\frac{j_1}{n}}\frac{1}{3}\right)}\right)\\
&\leqslant 2\exp\left(-\frac{K^2 j}{8\left(\zeta^2+2^{\frac{j_1}{2}}\left\|\psi\right\|_\infty\frac{K}{3}C2^{-\frac{j_1}{2}}\right)}\right).
\end{aligned}$$

Dabei wurde im ersten Schritt (2.16) verwendet, im zweiten Schritt $j \leqslant j_1$ und im dritten Schritt $C2^{-\frac{j}{2}} \geqslant \sqrt{\frac{j}{n}}$, also die Voraussetzung $2^j \leqslant C\frac{n}{j}$.
Nun wählt man ein $\gamma \geqslant 1$, ein $c \in \mathbb{R}^+$ derart, dass $c^2 \geqslant 8\log(2)\left(1+\frac{1}{3}c\left\|\psi\right\|_\infty\right)$ gilt, und setzt $K := c\gamma\sqrt{M \vee 1}$. Aus Hilfssatz 2.19 folgt außerdem $\zeta^2 \leqslant CM$ und daher insgesamt

$$\begin{aligned}
P\left(\left|\hat{\beta}_{jk}-\beta_{jk}\right| > \frac{K}{2}\sqrt{\frac{j}{n}}\right) &\leqslant 2\exp\left(-\frac{c^2(M\vee 1)\gamma^2 j}{8(CM+\frac{1}{3}c\sqrt{M\vee 1}\gamma C\left\|\psi\right\|_\infty)}\right)\\
&\leqslant 2\exp\left(-\frac{8\log(2)(1+\frac{1}{3}c\left\|\psi\right\|_\infty)(M\vee 1)\gamma^2 j}{8(CM+\frac{1}{3}c\sqrt{M\vee 1}\gamma C\left\|\psi\right\|_\infty)}\right)\\
&= 2\exp\left(-\frac{3\log(2)(1+\frac{1}{3}c\left\|\psi\right\|_\infty)(M\vee 1)\gamma^2 j}{3CM+c\sqrt{M\vee 1}\gamma C\left\|\psi\right\|_\infty}\right)\\
&\leqslant 2\exp\left(-\frac{\log(2)(3+c\left\|\psi\right\|_\infty)(M\vee 1)\gamma^2 j}{3CM+c(M\vee 1)\gamma C\left\|\psi\right\|_\infty}\right)\\
&\leqslant 2\exp\left(-\frac{\log(2)(3+c\left\|\psi\right\|_\infty)(M\vee 1)\gamma^2 j}{3CM\gamma+c(M\vee 1)\gamma C\left\|\psi\right\|_\infty}\right)\\
&\leqslant 2\exp\left(-\frac{\log(2)(3+c\left\|\psi\right\|_\infty)(M\vee 1)\gamma^2 j}{3C(M\vee 1)\gamma+c(M\vee 1)\gamma C\left\|\psi\right\|_\infty}\right)\\
&= 2\exp\left(-\frac{\gamma j\log(2)(3+c\left\|\psi\right\|_\infty)(M\vee 1)\gamma}{C(M\vee 1)(3+c\left\|\psi\right\|_\infty)\gamma}\right)\\
&= 2C\exp\left(\log 2^{-\gamma j}\right) = 2C\cdot 2^{-\gamma j}.
\end{aligned}$$

⧄

2.3.4 Normungleichungen

Sei $r \geqslant 1$ und $\hat{f} := \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\hat{f}_{jk}\psi_{jk}$ eine zufällige Funktion (eine Funktion, deren Waveletkoeffizienten nicht deterministisch, sondern von den Beobachtungen der Dichtefunktion f abhängig, also $\sigma(X_1, \ldots, X_n)$-messbar, sind). Später, im Beweis von Hauptsatz 2.1 werden diese Koeffizienten $\hat{f}_{jk}$ Summen und Differenzen aus den $\tilde{\beta}_{jk}$ des Schätzers TW und den wahren Koeffizienten β_{jk} sein. Um eine für den Beweis von Hauptsatz 2.1 brauchbare Abschätzung zu erhalten, sind zunächst zwei Hilfssätze nötig.

2.20 Hilfssatz: *Sei $r \geqslant 1$. Dann gilt*

$$\|\hat{f}\|_{L^r}^r \leqslant \left(\sum_{j=j_0}^{j_1} \|D_j \hat{f}\|_{L^r}^{(r\wedge 2)} \right)^{\frac{r}{r\wedge 2}}. \tag{2.17}$$

Beweis. Die rechte Seite ist die r-te Potenz eines Summanden der Besovnorm $J_{0r(r\wedge 2)}(\hat{f})$; falls dies ein endlicher Ausdruck ist, muss wegen $B_{0r(r\wedge 2)} \subset L^r$ (vgl. Satz 1.6) auch die linke Seite endlich sein. Nach Definition der Projektionsabbildung D_j und mit der Jensenschen Ungleichung folgt

$$\begin{aligned} \|\hat{f}\|_{L^r} = \left\| \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \hat{f}_{jk}\psi_{jk} \right\|_{L^r} \leqslant \sum_{j=j_0}^{j_1} \|D_j \hat{f}\|_{L^r} &= \left(\sum_{j=j_0}^{j_1} \|D_j \hat{f}\|_{L^r} \right)^{\frac{r\wedge 2}{r\wedge 2}} \\ &\leqslant \left(\sum_{j=j_0}^{j_1} \|D_j \hat{f}\|_{L^r}^{r\wedge 2} \right)^{\frac{1}{r\wedge 2}}. \end{aligned}$$

Potenziert man beide Seiten mit r, folgt die Behauptung. ▨

2.21 Hilfssatz: *Es gibt eine Konstante $C > 0$, sodass für $r \geqslant 1$ gilt*

$$\|D_j \hat{f}\|_{L^r}^r \leqslant C 2^{j\left(\frac{r}{2}-1\right)} \sum_{k\in\mathbb{Z}} |\hat{f}_{jk}|^r. \tag{2.18}$$

Beweis. Mit Satz 1.5, angewendet auf $D_j \hat{f}$ und mit $\lambda_k = \hat{f}_{jk}$, $p = r$, gilt

$$\|D_j \hat{f}\|_{L^r}^r = \left\| \sum_{k\in\mathbb{Z}} \hat{f}_{jk}\psi_{jk} \right\|_{L^r}^r \leqslant \left(c_2 2^{j(\frac{1}{2}-\frac{1}{r})} \right)^r \|\hat{f}_{j\bullet}\|_{\ell^r}^r = C 2^{j\left(\frac{r}{2}-1\right)} \sum_{k\in\mathbb{Z}} |\hat{f}_{jk}|^r.$$ ▨

Zur kürzeren Schreibweise seien für $x \in \mathbb{R}$ und für das r aus der Behauptung von Hauptsatz 2.1 stets

$$S(x) := \sum_{j=j_0}^{j_1} 2^{jx} \qquad \text{und} \qquad \xi := \frac{r}{r-2} \tag{2.19}$$

definiert. Die beiden folgenden Aussagen werden an zentraler Stelle im Beweis von Hauptsatz 2.1 mehrfach nützlich sein.

2.22 Satz: *Sei $j_0 \leqslant j_1$. Dann gibt es eine Konstante $C > 0$, sodass*

$$S(x) \leqslant \begin{cases} C 2^{\max\{j_0 x, j_1 x\}}, & x \neq 0, \\ j_1 - j_0, & x = 0. \end{cases} \tag{2.20}$$

Zur Voraussetzung $j_0 \leqslant j_1$ vergleiche Satz 2.27.

Beweis. Der Fall $x = 0$ ist klar, es gilt dann sogar Gleichheit. Sei also $x \neq 0$. Dann gilt mit der geometrischen Reihe für den Fall $xj_1 \leqslant xj_0$:

$$\sum_{j=j_0}^{j_1} 2^{jx} = \frac{2^{xj_0} - 2^{x(j_1+1)}}{1-2^x} \leqslant \frac{2^{xj_0}}{1-2^x} = C2^{xj_0}.$$

Für $xj_1 > xj_0$ ergibt sich insbesondere $2^{x(j_0-j_1)} < 1$ und $x > 0$ (da nach Voraussetzung $j_1 \geqslant j_0$ gilt) und folglich

$$\sum_{j=j_0}^{j_1} 2^{jx} \leqslant \frac{1}{1-2^x} 2^{xj_1} \left(2^{x(j_0-j_1)} - 2^x\right) = \frac{2^x - 2^{x(j_0-j_1)}}{2^x - 1} 2^{xj_1} \leqslant \frac{2^x}{2^x-1} 2^{xj_1} = C2^{xj_1}.$$

◪

2.23 Hilfssatz: *Seien $j_0 \leqslant j_1$, $r > 2$, und ξ wie in (2.19). Dann gibt es eine Konstante $C > 0$, sodass*

$$S(\kappa\xi)^{\left(\frac{r}{2}-1\right)} S(b) \leqslant \begin{cases} C2^{j_0\left(\kappa\frac{r}{2}+b\right)}, & \text{falls } \kappa < 0, b < 0; \\ C2^{j_1\left(\kappa\frac{r}{2}+b\right)} & \text{falls } \kappa > 0, b > 0. \end{cases} \tag{2.21}$$

Beweis. Wegen (2.20) gilt, da ξ positiv ist,

$$\begin{aligned} S(\kappa\xi)^{\left(\frac{r}{2}-1\right)} S(b) &\leqslant C \left(2^{\max\{j_0\kappa\xi, j_1\kappa\xi\}}\right)^{\left(\frac{r}{2}-1\right)} 2^{\max\{j_0 b, j_1 b\}} \\ &= C2^{\max\left\{j_0\kappa\frac{r}{r-2}\left(\frac{r-2}{2}\right), j_1\kappa\frac{r}{r-2}\left(\frac{r-2}{2}\right)\right\} + \max\{j_0 b, j_1 b\}} \\ &= C2^{\max\{j_0\kappa\frac{r}{2}, j_1\kappa\frac{r}{2}\} + \max\{j_0 b, j_1 b\}}. \end{aligned}$$

Mit der Voraussetzung $j_0 \leqslant j_1$ folgt nun die Behauptung. ◪

2.24 Satz: *Sei $r \geqslant 1$ und $j_0 \leqslant j_1$. Es gibt eine Konstante $C > 0$, sodass für alle $\kappa \in \mathbb{R}$ und für ξ aus (2.19) gilt*

$$E\left\|\hat{f}\right\|_{L^r}^r \leqslant \begin{cases} C \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1\right)} \sum_{k\in\mathbb{Z}} E\left|\hat{f}_{jk}\right|^r, & 1 \leqslant r \leqslant 2, \\ CS(\kappa\xi)^{\left(\frac{r}{2}-1\right)^+} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)} \sum_{k\in\mathbb{Z}} E\left|\hat{f}_{jk}\right|^r, & r > 2. \end{cases} \tag{2.22}$$

Wird $\kappa = 0$ gesetzt, kann man die zweite Zeile für alle $r \geqslant 1$ *verwenden.*

Beweis. Zunächst zum Fall $r \in [1, 2]$. Mit (2.17) und (2.18) ergibt sich

$$E\left\|\hat{f}\right\|_{L^r}^r \leqslant E\left(\sum_{j=j_0}^{j_1} \left\|D_j\hat{f}\right\|_{L^r}^r\right)^{\frac{r}{r}} = \sum_{j=j_0}^{j_1} E\left\|D_j\hat{f}\right\|_{L^r}^r \leqslant C\sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1\right)} \sum_{k\in\mathbb{Z}} E\left|\hat{f}_{jk}\right|^r.$$

Für $r > 2$ wird die Höldersche Ungleichung, angewendet auf die zueinander konjugierten Exponenten $\frac{r}{2}$ und $\frac{r}{r-2}$, benötigt. Für alle $\kappa \in \mathbb{R}$ gilt nämlich

$$
\begin{aligned}
\sum_{j=j_0}^{j_1} \big\|D_j\hat{f}\big\|_{L^r}^2 &= \sum_{j=j_0}^{j_1} \Big|2^{j\kappa}2^{-j\kappa}\big\|D_j\hat{f}\big\|_{L^r}^2\Big| \\
&\leqslant \left(\sum_{j=j_0}^{j_1} 2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}} \left(\sum_{j=j_0}^{j_1} \left(2^{-j\kappa}\big\|D_j\hat{f}\big\|_{L^r}^2\right)^{\frac{r}{2}}\right)^{\frac{2}{r}} \\
&= \left(\sum_{j=j_0}^{j_1} 2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}} \left(\sum_{j=j_0}^{j_1} 2^{-j\kappa\frac{r}{2}}\big\|D_j\hat{f}\big\|_{L^r}^r\right)^{\frac{2}{r}}.
\end{aligned}
$$

Damit ergibt sich

$$
\begin{aligned}
E\left(\sum_{j=j_0}^{j_1}\big\|D_j\hat{f}\big\|_{L^r}^2\right)^{\frac{r}{2}} &\leqslant E\left(\left(\sum_{j=j_0}^{j_1} 2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}} \left(\sum_{j=j_0}^{j_1} 2^{-j\kappa\frac{r}{2}}\big\|D_j\hat{f}\big\|_{L^r}^r\right)^{\frac{2}{r}}\right)^{\frac{r}{2}} \\
&= \left(\sum_{j=j_0}^{j_1} 2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r-2}{r}\cdot\frac{r}{2}} E\left(\sum_{j=j_0}^{j_1} 2^{-j\kappa\frac{r}{2}}\big\|D_j\hat{f}\big\|_{L^r}^r\right)^{\frac{2}{r}\cdot\frac{r}{2}} \\
&= \left(\sum_{j=j_0}^{j_1} 2^{j\kappa\frac{r}{r-2}}\right)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{-j\kappa\frac{r}{2}} E\big\|D_j\hat{f}\big\|_{L^r}^r \\
&= \left(\sum_{j=j_0}^{j_1} 2^{j\kappa\xi}\right)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{-j\kappa\frac{r}{2}} E\big\|D_j\hat{f}\big\|_{L^r}^r \\
&\leqslant CS(\kappa\xi)^{\frac{r}{2}-1}\sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1-\frac{\kappa r}{2})}\sum_{k\in\mathbb{Z}} E\big|\hat{f}_{jk}\big|^r.
\end{aligned}
$$

Dabei ging in der letzten Abschätzung Hilfssatz 2.21 und die Definition von S aus (2.19) ein. Mit (2.17) und $r \wedge 2 = 2$ folgt dann die Behauptung. ▨

2.4 Beweis des Hauptsatzes

Zur Erinnerung an die Voraussetzungen von Hauptsatz 2.1 sei $f \in \mathcal{D}_{\sigma pq}(M)$ mit $\sigma p > 1$ und $1 \leqslant p \leqslant r < \infty$, $r \leqslant \frac{1+2\sigma}{\sigma}$ und

$$2^{j_0(n)} \asymp \left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon \geqslant 0\}}}\right)^{1-2\alpha}, \qquad 2^{j_1(n)} \asymp \left(\frac{n}{\log n}\right)^{\frac{\alpha}{\sigma'}}.$$

Es gilt zunächst $f = E_{j_0}f + D_{j_0 j_1}f + f - E_{j_1+1}f$, wobei wie üblich E_{j_i} die Projektion auf den Raum V_{j_i} ist und $D_{j_0 j_1}f$ die Projektion von f auf den Teilraum $W_{j_0} \oplus \cdots \oplus W_{j_1}$:

$$E_{j_i}f(x) = \sum_{k\in\mathbb{Z}} \alpha_{j_i k}\varphi_{j_i k}(x) \qquad D_{j_0 j_1}f(x) = \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}(x).$$

Es besteht nämlich die Zerlegung (vgl. die Definition der Räume V_i und W_j in Kapitel 1)

$$V_{j_1+1} = V_{j_1} \oplus W_{j_1} = \cdots = V_{j_0} \oplus W_{j_0} \oplus W_{j_0+1} \oplus \cdots \oplus W_{j_1}$$

und daher die Beziehung $E_{j_1+1} = E_{j_0} + D_{j_0 j_1}$. Der abzuschätzende Fehlerterm von (2.4) wird also zunächst aufgeteilt in

$$\begin{aligned} E\,\|\mathrm{TW} - f\|_{L^r}^r \leqslant 3^r \Bigg(& E \left\| \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_0 k}\varphi_{j_0 k}(x) - E_{j_0}f \right\|_{L^r}^r + \\ & + E \left\| \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \tilde{\beta}_{jk}\psi_{jk}(x) - D_{j_0 j_1}f \right\|_{L^r}^r + \|f - E_{j_1+1}f\|_{L^r}^r \Bigg). \end{aligned} \tag{2.23}$$

Der erste Term beschreibt den Fehler des linearen Anteils von TW, in dem ohne Thresholding gearbeitet wird. Die Abschätzung des zweiten Terms wird den meisten Aufwand erfordern, er gibt den Fehler des nichtlinearen Anteils von TW an. Der dritte Term ist ein Bias-Term. Die Konstante 3^r wird im weiteren Beweis, wie alle anderen Konstanten, keine Rolle spielen.

2.4.1 Linearer Term

2.25 Satz: *(i) Es gibt eine Konstante $C > 0$, sodass*

$$E \left\| \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_0 k}\varphi_{j_0 k}(x) - E_{j_0}f \right\|_{L^r}^r \leqslant C\left(\frac{2^{j_0}}{n}\right)^{\frac{r}{2}}. \tag{2.24}$$

(ii) Nach der Wahl von $j_0(n)$ in (2.6) erreicht dieser Term die Konvergenzrate aus (2.7), falls $\varepsilon = 0$ und $\frac{r}{2p} \leqslant \frac{1}{q}$ ist. In den übrigen Fällen ist er vernachlässigbar gegen die Terme aus (2.7).

BEWEIS. *(i)* Zunächst folgt mit Satz 1.5:

$$\begin{aligned} E\left\|\sum_{k\in\mathbb{Z}}\hat{\alpha}_{j_0k}\varphi_{j_0k}(x)-E_{j_0}f\right\|_{L^r}^r &= E\left\|\sum_{k\in\mathbb{Z}}\big(\hat{\alpha}_{j_0k}-\alpha_{j_0k}\big)\varphi_{j_0k}(x)\right\|_{L^r}^r \\ &\leqslant c_2 2^{j_0r\left(\frac{1}{2}-\frac{1}{r}\right)}E\left\|\hat{\alpha}_{j_0\bullet}-\alpha_{j_0\bullet}\right\|_{\ell^r}^r \\ &= C2^{j_0\left(\frac{r}{2}-1\right)}\sum_{k\in\mathbb{Z}}E\left|\hat{\alpha}_{j_0k}-\alpha_{j_0k}\right|^r. \end{aligned}$$

Löst man die Summe auf, kommt der Faktor 2^{j_0} dazu. Da $|\mathrm{supp}\,\varphi_{j_0k}| = 2^{-j_0}\,|\mathrm{supp}\,\varphi|$, läuft die Summe über insgesamt $C\cdot 2^{j_0}$ Summanden, wobei in der Konstanten C die Träger von f und φ eingehen. Plakativer ausgedrückt: Da der Träger von φ_{j_0k} kleiner ist als der Träger von φ, sind im j_0-ten Level 2^{j_0}-mal so viele Summanden in der Wavelet-Entwicklung nötig. Der Träger von f kommt dabei implizit in den α_{jk} vor, die außerhalb von supp f verschwinden müssen. Daher können außerdem alle Summanden nach (2.14) abgeschätzt werden (vgl. zur Anwendbarkeit von (2.14) auch Hilfssatz 2.15)

$$\begin{aligned} E\left\|\sum_{k\in\mathbb{Z}}\hat{\alpha}_{j_0k}\varphi_{j_0k}(x)-E_{j_0}f\right\|_{L^r}^r &\leqslant C2^{j_0}2^{j_0\left(\frac{r}{2}-1\right)}\sup_{k\in\mathbb{Z}}E\left|\hat{\alpha}_{j_0k}-\alpha_{j_0k}\right|^r \\ &\leqslant C2^{j_0\frac{r}{2}}n^{-\frac{r}{2}} \\ &= C\left(\frac{2^{j_0}}{n}\right)^{\frac{r}{2}}. \end{aligned}$$

(ii) Sei $\varepsilon = 0$, also $\sigma p = \frac{r-p}{2}$, und $\frac{r}{2p} \leqslant \frac{1}{q}$. Nach (2.6) wächst $2^{j_0(n)}$ asymptotisch wie

$$\left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon\geqslant 0\}}}\right)^{1-2\alpha} = \left(n(\log n)^{\frac{r-p}{p}}\right)^{1-2\alpha},$$

und daher gilt

$$\begin{aligned} \left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp \left(\frac{\left(n(\log n)^{\frac{r-p}{p}}\right)^{1-2\alpha}}{n}\right)^{\frac{r}{2}} &= n^{-\alpha r}(\log n)^{\left(\frac{r}{2p}-\frac{1}{2}-\frac{\alpha r}{p}+\alpha\right)r} \\ &= \left(\frac{\log n}{n}\right)^{\alpha r}(\log n)^{r\left(\frac{r}{2p}-\frac{1}{2}-\frac{\alpha r}{p}\right)}. \end{aligned} \tag{2.25}$$

Nun ist wegen $\varepsilon = 0$ und (2.9)

$$\frac{r}{2p}-\frac{1}{2}-\frac{\alpha r}{p} = 0 \iff r-p-2\alpha r = 0 \iff r-p-2\frac{r-p}{2} = 0,$$

was offensichtlich erfüllt ist. Folglich verschwindet der letzte $(\log n)$-Term und es bleibt

$$\left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp \left(\frac{\log n}{n}\right)^{\alpha r}.$$

Das ist aber die gleiche Konvergenzrate wie in der Behauptung (2.7), da wegen $\frac{r}{2p} \leqslant \frac{1}{q}$ und also $-\frac{2p}{r} \leqslant -q$ gilt

$$1 - \frac{2p}{rq} \leqslant 1 - \frac{q}{q} = 0 \implies \left(\frac{1}{2} - \frac{p}{qr}\right)^+ = 0.$$

Der $(\log n)$-Term fällt also in diesem Fall auch in der Behauptung weg.

Ist $\varepsilon = 0$, aber $\frac{r}{2p} > \frac{1}{q}$, so fällt der $(\log n)$-Term zwar wie in (2.25) weg, aber nicht in (2.7). Die Konvergenzrate ist damit also echt besser als in der Behauptung und die rechte Seite von (2.24) folglich vernachlässigbar gegen die rechte Seite von (2.7).

Ist $\varepsilon < 0$, so verhält sich $2^{j_0(n)}$ asymptotisch wie $n^{1-2\alpha}$ und daher

$$\left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp \left(\frac{n^{1-2\alpha}}{n}\right)^{\frac{r}{2}} = n^{-\alpha r},$$

was vernachlässigbar gegen die in Hauptsatz 2.1 behauptete Schranke $(\log n)^{\alpha r} n^{-\alpha r}$ ist.

Für $\varepsilon > 0$ gilt wieder wie oben

$$\left(\frac{2^{j_0(n)}}{n}\right)^{\frac{r}{2}} \asymp n^{-\alpha r} (\log n)^{\left(\frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha\right) r},$$

wogegen in Hauptsatz 2.1 $(\log n)^{\alpha r} n^{-\alpha r}$ als obere Schranke angegeben wird. Da die $n^{-\alpha r}$-Terme übereinstimmen, ist zu zeigen, dass gilt

$$\alpha r \geqslant \left(\frac{r}{2p} - \frac{1}{2} - \frac{\alpha r}{p} + \alpha\right) r.$$

Da nach Hilfssatz 2.2 $\alpha \geqslant 0$ und nach Voraussetzung von Hauptsatz 2.1 $\sigma p > 1$ gilt, ist auch $\frac{\alpha}{\sigma p} \geqslant 0$. Dann folgt die Behauptung aus Hilfssatz 2.5. ▨

2.4.2 Biasterm

2.26 Satz: (*i*) *Es gibt eine Konstante $C > 0$, sodass*

$$\|f - E_{j_1+1} f\|_{L^r}^r \leqslant C 2^{-j_1 \sigma' r}. \tag{2.26}$$

(*ii*) *Nach der Wahl von j_1 in* (2.6) *erreicht dieser Term die Konvergenzrate aus* (2.7) *falls $\varepsilon \neq 0$ oder falls $\varepsilon = 0$ und $\frac{r}{2p} \leqslant \frac{1}{q}$ gilt. In den übrigen Fällen ist er gegen die Schranken aus* (2.7) *vernachlässigbar.*

Diese Aussage ist dafür verantwortlich, dass die Behauptung von Hauptsatz 2.1 gegenüber [4, Thm. 3] abgeschwächt wurde. Die rechte Seite von (2.26) erreicht nämlich auch im Fall $\varepsilon > 0$ genau die in Hauptsatz 2.1 behauptete Konvergenzrate, die echt schwächer ist als die Behauptung aus [4, Thm. 3].

Beweis. (*i*) Es gilt zunächst wegen der Zerlegung (1.1) und dann wegen Satz 1.5

$$\begin{aligned}\|f - E_{j_1+1}f\|_{L^r} = \left\|\sum_{j>j_1} D_j f\right\|_{L^r} &= \left\|\sum_{j>j_1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\right\|_{L^r} \\ &\leqslant \sum_{j>j_1}\left\|\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\right\|_{L^r} \\ &\leqslant \sum_{j>j_1} c_2 2^{j(\frac{1}{2}-\frac{1}{r})}\, \|\beta_{j\bullet}\|_{\ell^r}\,.\end{aligned}$$

Aufgrund von Satz 1.6 folgt im Fall $r > p$: $f \in B_{\sigma pq} \subset B_{\sigma' rq} \subset B_{\sigma' r\infty}$ und im Fall $r = p$ wegen $\sigma' = \sigma - \frac{1}{p} + \frac{1}{r} = \sigma$ die Einbettung $f \in B_{\sigma pq} \subset B_{\sigma' r\infty}$. Insbesondere ist

$$\sup_{j\geqslant 0}\left(2^{j(\sigma'+\frac{1}{2}-\frac{1}{r})}\right)\|\beta_{j\bullet}\|_{\ell^r} = C < \infty,$$

da die linke Seite ein Summand der Norm $\|f\|_{\sigma' r\infty}$ ist. Es ergibt sich

$$\begin{aligned}\|f - E_{j_1}f\|_{L^r} &\leqslant \sum_{j>j_1} c_2 2^{j(\frac{1}{2}-\frac{1}{r})} 2^{-j(\sigma'+\frac{1}{2}-\frac{1}{r})} 2^{j(\sigma'+\frac{1}{2}-\frac{1}{r})}\, \|\beta_{j\bullet}\|_{\ell^r} \\ &\leqslant \sum_{j>j_1} c_2 2^{j(\frac{1}{2}-\frac{1}{r})} 2^{-j(\sigma'+\frac{1}{2}-\frac{1}{r})} C \\ &= C \sum_{j>j_1} 2^{-j(-\frac{1}{2}+\frac{1}{r}+\sigma'+\frac{1}{2}-\frac{1}{r})} = C\sum_{j>j_1} 2^{-j\sigma'}.\end{aligned}$$

Mit der geometrischen Reihe folgt schließlich die behauptete Abschätzung:

$$\begin{aligned}\|f - E_{j_1+1}f\|_{L^r}^r \leqslant C\left(\sum_{j>j_1} 2^{-j\sigma'}\right)^r &= C2^{-(j_1-1)\sigma' r}\sum_{j\geqslant 0} 2^{-j\sigma' r} = C2^{-j_1\sigma' r}\frac{1}{1-2^{-\sigma' r}} \\ &= C2^{-j_1\sigma' r}.\end{aligned}$$

(*ii*) Nach der Wahl von j_1 in (2.6) ergibt sich für die asymptotische Schranke des Biasterms

$$2^{-j_1(n)\sigma' r} \asymp \left(\left(\frac{n}{\log n}\right)^{\frac{\alpha}{\sigma'}}\right)^{-\sigma' r} = \left(\frac{\log n}{n}\right)^{\alpha r}.$$

Für $\varepsilon \neq 0$ ist das genau die in (2.7) behauptete Schranke. Das Gleiche gilt, falls $\varepsilon = 0$ und $\frac{r}{2p} \leqslant \frac{1}{q}$ ist, wie man genauso wie beim linearen Term sieht (der Positivteil, der in (2.7) vorkommt, fällt in diesem Fall weg). Ist $\varepsilon = 0$ und $\frac{r}{2p} > \frac{1}{q}$, ist die Schranke in (2.26), wieder wie beim linearen Term, echt besser als die in (2.7) behauptete Schranke.▨

2.4.3 Nichtlinearer Term

Wie bereits in der Diskussion direkt nach Hauptsatz 2.1 angedeutet wurde, ist der Schätzer TW in vielen Fällen tatsächlich nichtlinear, weil $j_0 \leqslant j_1$ ist und daher die Waveletkoeffizienten mit Thresholding nicht in einer leeren Summe stehen. In den Fällen, in denen TW zu einem linearen Schätzer ausartet, sind auch die im Folgenden bewiesenen Konvergenzraten nicht von Interesse, da sie gar nicht benötigt werden. Allgemeine Aussagen zu linearen Schätzern folgen in Kapitel 3.

Für alle Abschätzungen des nichtlinearen Terms gilt daher ohne Einschränkung die Generalvoraussetzung $j_0 \leqslant j_1$. Ist diese Voraussetzung nämlich nicht erfüllt, tritt überhaupt kein nichtlinearer Term in TW auf.

2.27 Satz: *Sei $p < r$ und $\varepsilon \neq 0$. Dann gilt $j_0(n) \leqslant j_1(n)$ für $n \gg 0$.*
Insbesondere kommen in diesen Fällen im Schätzer TW *auch die nichtlinearen Schätzer $\tilde{\beta}_{jk}$ überhaupt vor.*

Beweis. Nach (2.6) ist

$$j_0 \asymp (1-2\alpha)\left(\log n + \frac{r-p}{p}\mathbb{1}_{\{\varepsilon \geqslant 0\}} \log\log n\right), \qquad j_1 \asymp \frac{\alpha}{\sigma'}(\log n - \log\log n)\,.$$

Da der iterierte Logarithmus gegen den einfachen Logarithmus vernachlässigbar ist, ist die Anzahl der Summanden über j im Wesentlichen proportional zu $\log n$.

Sei $\varepsilon > 0$. Nach (2.10) ist $1-2\alpha = \frac{\alpha}{\sigma}$ und damit für $n \gg 0$

$$j_0 < j_1 \iff 1-2\alpha < \frac{\alpha}{\sigma'} \iff \frac{\alpha}{\sigma} < \frac{\alpha}{\sigma'} \iff \sigma' < \sigma \iff \sigma - \frac{1}{p} + \frac{1}{r} < \sigma,$$

was gerade für $p < r$ erfüllt ist.

Sei nun $\varepsilon < 0$. Dann folgt die Behauptung aus Hilfssatz 2.7. ▨

Für die Abschätzung des nichtlinearen Terms werden sechs Mengen definiert:

$$\begin{aligned}
\hat{B}_j &:= \left\{k\colon |\hat{\beta}_{jk}| > K\sqrt{\frac{j}{n}}\right\}, & \hat{S}_j &:= \hat{B}_j^{\complement},\\
B_j &:= \left\{k\colon |\beta_{jk}| > \frac{K}{2}\sqrt{\frac{j}{n}}\right\}, & S_j &:= B_j^{\complement},\\
B'_j &:= \left\{k\colon |\beta_{jk}| > 2K\sqrt{\frac{j}{n}}\right\}, & S'_j &:= B_j'^{\complement}.
\end{aligned}$$

Man beachte, dass K darin noch von (einem noch näher zu bestimmenden) γ abhängt.

Damit gilt nach Wahl von $\tilde{\beta}_{jk}$

$$\begin{aligned}
\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\tilde{\beta}_{jk}\psi_{jk}(x) - D_{j_0j_1}f &= \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\big(\tilde{\beta}_{jk}-\beta_{jk}\big)\psi_{jk} \\
&= \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\big(\tilde{\beta}_{jk}-\beta_{jk}\big)\psi_{jk}\mathbb{1}_{\{k\in\hat{B}_j\}} \\
&\qquad + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\big(0-\beta_{jk}\big)\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\}} \\
&= \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\big(\tilde{\beta}_{jk}-\beta_{jk}\big)\psi_{jk}\mathbb{1}_{\{k\in\hat{B}_j\cap(B_j\cup S_j)\}} \\
&\qquad - \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap(B'_j\cup S'_j)\}} \\
&= \left(\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\big(\tilde{\beta}_{jk}-\beta_{jk}\big)\psi_{jk}\mathbb{1}_{\{k\in\hat{B}_j\cap S_j\}}\right. \\
&\qquad \left. + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\big(\tilde{\beta}_{jk}-\beta_{jk}\big)\psi_{jk}\mathbb{1}_{\{k\in\hat{B}_j\cap B_j\}}\right) \\
&\qquad - \left(\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}}\right. \\
&\qquad \left. + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap S'_j\}}\right) \\
&=: (e_{bs}+e_{bb}) - (e_{sb}+e_{ss}),
\end{aligned}$$

wobei die vier am Schluss neu definierten Terme die nahe liegenden Bezeichnungen tragen. Nun werden also nacheinander die vier Terme e_{bs}, e_{bb}, e_{sb} und e_{ss} abgeschätzt. Wie üblich folgt nämlich

$$\begin{aligned}
E\left\|\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\tilde{\beta}_{jk}\psi_{jk}(x) - D_{j_0j_1}f\right\|_{L^r}^r &\leqslant \\
&4^r\left(E\left\|e_{bs}\right\|_{L^r}^r + E\left\|e_{bb}\right\|_{L^r}^r + E\left\|e_{sb}\right\|_{L^r}^r + E\left\|e_{ss}\right\|_{L^r}^r\right).
\end{aligned}$$

Zunächst zu den „gemischten Termen“ e_{bs} und e_{sb}. Sie beschreiben große Fehler der Schätzer $\tilde{\beta}_{jk}$. Es sei an Satz 2.16 erinnert, der besagt, dass solche großen Fehler des Schätzers „selten“ sind.

2.28 Satz: *(i) Sei $j_0 \leqslant j_1$. Sei weiter $t > 1$ so gewählt, dass $\frac{rt}{t-1} > 2$, und sei $\gamma > \frac{rt}{2}$ genügend groß. Dann gibt es eine Konstante $C > 0$, sodass gilt*

$$E \left\| e_{bs} \right\|_{L^r}^r \leqslant C n^{-\frac{r}{2}} 2^{j_0\left(\frac{r}{2}-\frac{\gamma}{t}\right)}.$$

(ii) Diese Schranke ist vernachlässigbar gegen die rechte Seite von (2.24).

Vergleiche zur Wahl von γ auch die Aussagen von Satz 2.29 und Satz 2.16.

Beweis. (*i*) Setze $\hat{f}_{jk} := \left(\hat{\beta}_{jk} - \beta_{jk}\right) \mathbb{1}_{\left\{k \in \hat{B}_j \cap S_j\right\}}$. Nach Konstruktion gilt offensichtlich

$$\hat{B}_j \cap S_j \subset \left\{ k \colon \left|\hat{\beta}_{jk} - \beta_{jk}\right| > \frac{K}{2}\sqrt{\frac{j}{n}} \right\} =: \Delta_{jk}.$$

Damit und mit der Hölderschen Ungleichung, angewendet auf zueinander konjugierte Exponenten t und t',

$$\begin{aligned}\sum_{k \in \mathbb{Z}} E\left|\hat{f}_{jk}\right|^r = \sum_{k \in \mathbb{Z}} E \left| \left(\hat{\beta}_{jk} - \beta_{jk}\right) \mathbb{1}_{\left\{k \in \hat{B}_j \cap S_j\right\}} \right|^r &\leqslant \sum_{k \in \mathbb{Z}} E \left(\left|\hat{\beta}_{jk} - \beta_{jk}\right|^r \mathbb{1}_{\left\{k \in \Delta_{jk}\right\}} \right) \\ &\leqslant \sum_{k \in \mathbb{Z}} E \left(\left|\hat{\beta}_{jk} - \beta_{jk}\right|^{rt'} \right)^{\frac{1}{t'}} P(\Delta_{jk})^{\frac{1}{t}}.\end{aligned}$$

Wegen der Voraussetzung $\frac{rt}{t-1} = rt' > 2$ und wegen Hilfssatz 2.17 ist (2.14) anwendbar. Zusammen mit (2.15) ergibt dies ((2.15) ist ebenfalls wegen Hilfssatz 2.17 anwendbar)

$$\sum_{k \in \mathbb{Z}} E\left|\hat{f}_{jk}\right|^r \leqslant \sum_{k \in \mathbb{Z}} C n^{-\frac{rt'}{2} \cdot \frac{1}{t'}} P(\Delta_{jk})^{\frac{1}{t}} \leqslant C n^{-\frac{r}{2}} 2^j \sup_{k \in \mathbb{Z}} P(\Delta_{jk})^{\frac{1}{t}} \leqslant C n^{-\frac{r}{2}} 2^{j - \frac{j\gamma}{t}}.$$

Beim Auflösen der Summe ist wieder ein Faktor 2^j dazu gekommen. Vergleiche hierzu den Beweis von Satz 2.25(*i*).

Nun wird eine Fallunterscheidung nach dem Wert von r durchgeführt.

Sei $r > 2$. Aus (2.22) folgt dann für ein beliebiges $\kappa \in \mathbb{R}$ und für das ξ aus (2.19)

$$\begin{aligned}E \left\| e_{bs} \right\|_{L^r}^r &= E \left\| \sum_{j=j_0}^{j_1} \sum_{k \in \mathbb{Z}} \hat{f}_{jk} \psi_{jk} \right\|_{L^r}^r \\ &\leqslant C S(\kappa\xi)^{\left(\frac{r}{2}-1\right)} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)} \sum_{k \in \mathbb{Z}} E\left|\hat{f}_{jk}\right|^r \\ &\leqslant C S(\kappa\xi)^{\left(\frac{r}{2}-1\right)} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)} C n^{-\frac{r}{2}} 2^{j\left(1-\frac{\gamma}{t}\right)} \\ &= C S(\kappa\xi)^{\left(\frac{r}{2}-1\right)} \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-\frac{\kappa r}{2}-\frac{\gamma}{t}\right)} n^{-\frac{r}{2}} \\ &= C S(\kappa\xi)^{\left(\frac{r}{2}-1\right)} S\left((1-\kappa)\frac{r}{2} - \frac{\gamma}{t}\right) n^{-\frac{r}{2}}\end{aligned}$$

Wählt man nun $\kappa < 0$ und $\gamma \geqslant 1$ genügend groß, sodass $(1-\kappa)\frac{r}{2} - \frac{\gamma}{t} < 0$ gilt, dann folgt aus Hilfssatz 2.23

$$E\,\|e_{bs}\|_{L^r}^r \leqslant Cn^{-\frac{r}{2}} 2^{j_0\left(\kappa\frac{r}{2}+(1-\kappa)\frac{r}{2}-\frac{\gamma}{t}\right)} = Cn^{-\frac{r}{2}} 2^{j_0\left(\frac{r}{2}-\frac{\gamma}{t}\right)}.$$

Ist $r \in [1,2]$, folgt aus (2.22) zunächst

$$\begin{aligned}
E\,\|e_{bs}\|_{L^r}^r \leqslant E \left\| \sum_{j=j_0}^{j_1} \sum_{k\in\mathbb{Z}} \hat{f}_{jk}\psi_{jk} \right\|_{L^r}^r &\leqslant C \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)} \sum_{k\in\mathbb{Z}} E\big|\hat{f}_{jk}\big|^r \\
&\leqslant C \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)} 2^{j-\frac{j\gamma}{t}} n^{-\frac{r}{2}} \\
&= Cn^{-\frac{r}{2}} S\left(\frac{r}{2}-\frac{\gamma}{t}\right) \\
&\leqslant Cn^{-\frac{r}{2}} 2^{\max\left\{j_0\left(\frac{r}{2}-\frac{\gamma}{t}\right), j_1\left(\frac{r}{2}-\frac{\gamma}{t}\right)\right\}} \\
&= Cn^{-\frac{r}{2}} 2^{j_0\left(\frac{r}{2}-\frac{\gamma}{t}\right)}.
\end{aligned}$$

Dabei ging die zur Voraussetzung $\gamma > \frac{rt}{2}$ äquivalente Ungleichung $\frac{r}{2} - \frac{\gamma}{t} < 0$ ein, zusammen mit (2.20).

Das war zu zeigen.

(ii) Da $\gamma, t \geqslant 1$ gewählt waren, ist diese Schranke asymptotisch stets echt kleiner als die Schranke aus (2.24) und daher für die Behauptung von Hauptsatz 2.1 irrelevant. ▨

2.29 Satz: *(i) Sei $j_0 \leqslant j_1$. Es gibt eine Konstante $C > 0$, sodass gilt*

$$E\,\|e_{sb}\|_{L^r}^r \leqslant C 2^{-j_0(\gamma+\sigma' r)}. \tag{2.27}$$

(ii) Wählt man $\gamma \geqslant \frac{r\alpha}{1-2\alpha} - \sigma' r$ genügend groß, ist dieser Term gegen die rechte Seite von (2.26) vernachlässigbar.

In dieser Aussage wird eine quantitative Definition für γ angegeben, die auf die behauptete Konvergenzrate zugeschnitten ist und nur vom betrachteten Besovraum $B_{\sigma pq} \ni f$ abhängt, nicht von sonstigen freien Parametern (vgl. Satz 2.16 und Satz 2.28). Bisher war $\gamma \geqslant 1$ stets „genügend groß" gewählt.

Beweis. *(i)* Setze $\hat{f}_{jk} := \beta_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}}$. Wie bei Satz 2.28 gilt auch hier $\hat{S}_j \cap B'_j \subset \Delta_{jk}$. Da die β_{jk} deterministisch sind, folgt nach (2.15) (dies ist anwendbar wegen Hilfssatz 2.17)

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} E\big|\hat{f}_{jk}\big|^r &= \sum_{k\in\mathbb{Z}} E\big|\beta_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}}\big|^r \\
&= \sum_{k\in\mathbb{Z}} |\beta_{jk}|^r\, P\big(\hat{S}_j \cap B'_j\big) \\
&\leqslant C \sum_{k\in\mathbb{Z}} |\beta_{jk}|^r\, 2^{-\gamma j} \\
&= C\,\|\beta_{j\bullet}\|_{\ell^r}^r\, 2^{-\gamma j} 2^{-j(\sigma'+\frac{1}{2}-\frac{1}{r})r} 2^{j(\sigma'+\frac{1}{2}-\frac{1}{r})r} \\
&\leqslant C 2^{-j(\sigma'+\frac{1}{2}-\frac{1}{r})r} \sup_{j\geqslant 0} \left(2^{j(\sigma'+\frac{1}{2}-\frac{1}{r})}\,\|\beta_{j\bullet}\|_{\ell^r}\right)^r 2^{-\gamma j}.
\end{aligned}$$

Das Supremum ist ein Summand in der Norm $\|f\|_{\sigma' r\infty}$ und aufgrund der Einbettung $f \in \mathcal{D}_{\sigma pq} \subset \mathcal{D}_{\sigma' rq} \subset \mathcal{D}_{\sigma' r\infty}$ aus Satz 1.6 endlich. Damit ergibt sich nun

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} E\big|\hat{f}_{jk}\big|^r &\leqslant C2^{-j(\sigma' r+\frac{r}{2}-1+\gamma)}\left(\sup_{j\geqslant 0} 2^{j(\sigma'+\frac{1}{2}-\frac{1}{r})}\,\|\beta_{j\bullet}\|_{\ell^r}\right)^r \\
&\leqslant C2^{-j(\sigma' r+\frac{r}{2}-1+\gamma)}\left(\|\alpha_{0\bullet}\|_{\ell^r} + \sup_{j\geqslant 0} 2^{j(\sigma'+\frac{1}{2}-\frac{1}{r})}\,\|\beta_{j\bullet}\|_{\ell^r}\right)^r \\
&= C2^{-j(\sigma' r+\frac{r}{2}-1+\gamma)}\,\|f\|^r_{\sigma' r\infty} \\
&\leqslant C2^{-j(\sigma' r+\frac{r}{2}-1+\gamma)} M^r.
\end{aligned}$$

Man beachte hier, dass wegen Satz 1.6 $\mathcal{D}_{\sigma pq}(M) \subset \mathcal{D}_{\sigma' rq}(M) \subset \mathcal{D}_{\sigma' r\infty}(M)$ und die Abschätzung $\|f\|_{\sigma' r\infty} \leqslant M$ gilt.

Sei nun zunächst $r > 2$. Für den abzuschätzenden Term e_{sb} folgt dann mit (2.22)

$$\begin{aligned}
E\,\|e_{sb}\|^r_{L^r} &= E\left\|\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}}\right\|^r_{L^r} \\
&\leqslant CS(\kappa\xi)^{\left(\frac{r}{2}-1\right)}\sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\kappa\frac{r}{2}\right)}\sum_{k\in\mathbb{Z}} E\big|\hat{f}_{jk}\big|^r \\
&\leqslant CS(\kappa\xi)^{\left(\frac{r}{2}-1\right)}\sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1-\kappa\frac{r}{2}\right)} 2^{-j(\sigma' r+\frac{r}{2}-1+\gamma)} M^r \\
&= CM^r S(\kappa\xi)^{\left(\frac{r}{2}-1\right)} S\left(-r\left(\frac{\kappa}{2}+\sigma'\right)-\gamma\right).
\end{aligned}$$

Analog zum Beweis von Satz 2.28 können nun $\kappa < 0$ und $\gamma \geqslant 1$ genügend groß gewählt werden, sodass $-r\left(\frac{\kappa}{2}+\sigma'\right)-\gamma < 0$ ist. Dies führt mit (2.21) zu

$$E\,\|e_{sb}\|^r_{L^r} \leqslant CM^r C2^{\left(\frac{\kappa r}{2}+\left(-r\left(\frac{\kappa}{2}+\sigma'\right)\right)-\gamma\right)j_0} = C2^{-(r\sigma'+\gamma)j_0}.$$

Für $r \in [1,2]$ folgt aus (2.22) zunächst

$$\begin{aligned}
E\,\|e_{sb}\|^r_{L^r} &= E\left\|\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap B'_j\}}\right\|^r_{L^r} \\
&\leqslant C\sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)}\sum_{k\in\mathbb{Z}} E\big|\hat{f}_{jk}\big|^r \\
&\leqslant C\sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)} 2^{-j(\sigma' r+\frac{r}{2}-1+\gamma)} M^r \\
&= CM^r S\left(-r\sigma'-\gamma\right) \\
&= CM^r 2^{-j_0(r\sigma'+\gamma)}.
\end{aligned}$$

Dabei ging die Wahl $\gamma \geqslant 1$ ein, da nach Voraussetzung von Hauptsatz 2.1 $r \geqslant 1$, $\sigma' \geqslant 0$ und daher insgesamt $-r\sigma' - \gamma < 0$ gilt, sowie (2.20).

(*ii*) Sei $\gamma \geqslant \frac{\alpha r}{1-2\alpha} - \sigma' r$ und so groß, dass außerdem die oben schon verwendete Ungleichung $-r\left(\frac{\kappa}{2} + \sigma'\right) - \gamma < 0$ gilt. Dann ist für $n \gg 0$ wegen (2.6)

$$
\begin{aligned}
\left(E\left\|e_{sb}\right\|_{L^r}^r\right)^{-1} \geqslant C2^{j_0\left(r\sigma' + \frac{r\alpha}{1-2\alpha} - \sigma' r\right)} &= C2^{j_0\frac{\alpha r}{1-2\alpha}} \\
&\asymp \left(\left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon \geqslant 0\}}}\right)^{1-2\alpha}\right)^{\frac{\alpha r}{1-2\alpha}} \\
&= \left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon \geqslant 0\}}}\right)^{\alpha r} \\
&> \left(n(\log n)^{-1}\right)^{\alpha r} = \left(\frac{n}{\log n}\right)^{\alpha r\frac{\sigma'}{\sigma'}} \asymp 2^{j_1 r\sigma'}.
\end{aligned}
$$

Daher ist $2^{j_1 r\sigma'}$ vernachlässigbar gegen $\left(E\left\|e_{sb}\right\|_{L^r}^r\right)^{-1}$, was durch potenzieren mit -1 die Behauptung liefert. ▨

Nun zur Abschätzung der Hauptterme e_{bb} und e_{ss}, die ein ähnliches Verhalten von Schätzer und wahrem β_{jk} abdecken. Bei e_{bb} nehmen sowohl der Schätzer $\hat{\beta}_{jk}$ als auch β_{jk} selbst betragsmäßig „große“ Werte an.

2.30 Satz: (*i*) *Sei $j_0 \leqslant j_1$. Es gibt eine Konstante $C > 0$, sodass*

$$
E\left\|e_{bb}\right\|_{L^r}^r \leqslant \begin{cases} C2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}} n^{-\frac{r-p}{2}} & \text{für } \varepsilon \neq 0, \\ C\left(\frac{j_1}{n}\right)^{\frac{r-p}{2}} & \text{für } \varepsilon = 0. \end{cases}
$$

(*ii*) *Dieser Term ist vernachlässigbar gegen die in Hauptsatz 2.1 behaupteten Konvergenzraten.*

Beweis. (*i*) Sei $\hat{f}_{jk} := \left(\hat{\beta}_{jk} - \beta_{jk}\right)\mathbb{1}_{\left\{k \in \hat{B}_j \cap B_j\right\}}$. Dann gilt, weil für alle $r \geqslant p \geqslant 1$ trivialerweise $1^p = 1^r$ gilt, und wegen (2.14) und Hilfssatz 2.17

$$
\begin{aligned}
\sum_{k\in\mathbb{Z}} E\left|\hat{f}_{jk}\right|^r = \sum_{k\in\mathbb{Z}} E\left(\left|\hat{\beta}_{jk} - \beta_{jk}\right|\mathbb{1}_{\left\{k\in\hat{B}_j\cap B_j\right\}}\right)^r &\leqslant \sum_{k\in\mathbb{Z}} E\left(\left|\hat{\beta}_{jk} - \beta_{jk}\right|\mathbb{1}_{\{k\in B_j\}}\right)^r \\
&= \sum_{k\in\mathbb{Z}} E\left(\left|\hat{\beta}_{jk} - \beta_{jk}\right|\right)^r \left(\mathbb{1}_{\{k\in B_j\}}\right)^p \\
&\leqslant Cn^{-\frac{r}{2}} \sum_{k\in\mathbb{Z}} \left(\mathbb{1}_{\{k\in B_j\}}\right)^p \\
&< Cn^{-\frac{r}{2}} \sum_{k\in B_j} \left|\frac{2\left|\beta_{jk}\right|}{K}\sqrt{\frac{n}{j}}\right|^p,
\end{aligned}
$$

letzteres nach Definition von B_j. Weiter gilt

$$\begin{aligned}
\sum_{k\in\mathbb{Z}} E\left|\hat{f}_{jk}\right|^r &\leqslant Cn^{-\frac{r}{2}}\frac{n^{\frac{p}{2}}}{j^{\frac{p}{2}}}\sum_{k\in B_j}|\beta_{jk}|^p \\
&\leqslant Cn^{-\frac{r-p}{2}}j^{-\frac{p}{2}}\left\|2^{j(\sigma+\frac{1}{2}-\frac{1}{p})}\beta_{j\bullet}\right\|_{\ell^p}^p 2^{-j(\sigma+\frac{1}{2}-\frac{1}{p})p} \\
&\leqslant Cn^{-\frac{r-p}{2}}j^{-\frac{p}{2}}2^{-j(\sigma+\frac{1}{2}-\frac{1}{p})p}\left(\sup_{j\geqslant 0}\left\|2^{j(\sigma+\frac{1}{2}-\frac{1}{p})}\beta_{j\bullet}\right\|_{\ell^p}+\|\alpha_{0\bullet}\|_{\ell^p}\right)^p \\
&= Cn^{-\frac{r-p}{2}}j^{-\frac{p}{2}}2^{-j(\sigma+\frac{1}{2}-\frac{1}{p})p}\,\|f\|_{\sigma p\infty}^p \\
&\leqslant Cn^{-\frac{r-p}{2}}j^{-\frac{p}{2}}2^{-j(\sigma+\frac{1}{2}-\frac{1}{p})p}M^p.
\end{aligned}$$

Hierbei folgte aus Satz 1.6, dass $\|f\|_{\sigma p\infty}^p \leqslant M^p$ ist.

Ist $\varepsilon = \sigma p - \frac{r-p}{2} > 0$ und $r > 2$, wählt man ein $\kappa < 0$ derart, dass $-\varepsilon - \frac{\kappa}{2} < 0$ erfüllt ist. Daraus folgt dann mit (2.20) und (2.22), und da $j^{-\frac{p}{2}} \leqslant 1$ ist,

$$\begin{aligned}
E\,\|e_{bb}\|_{L^r}^r &\leqslant CS(\kappa\xi)^{\left(\frac{r}{2}-1\right)}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)}\sum_{k\in\mathbb{Z}}E\left|\hat{f}_{jk}\right|^r \\
&\leqslant CS(\kappa\xi)^{\left(\frac{r}{2}-1\right)}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}\right)}Cn^{-\frac{r-p}{2}}j^{-\frac{p}{2}}2^{-j(\sigma+\frac{1}{2}-\frac{1}{p})p}M^p \\
&\leqslant CM^pS(\kappa\xi)^{\left(\frac{r}{2}-1\right)}n^{-\frac{r-p}{2}}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r}{2}-1-\frac{\kappa r}{2}-\sigma p-\frac{p}{2}+1\right)} \\
&= CS(\kappa\xi)^{\left(\frac{r}{2}-1\right)}n^{-\frac{r-p}{2}}\sum_{j=j_0}^{j_1}2^{j\left(\frac{r-p}{2}-\frac{\kappa r}{2}-\sigma p\right)} \\
&= Cn^{-\frac{r-p}{2}}S(\kappa\xi)^{\left(\frac{r}{2}-1\right)}S\left(-\varepsilon-\frac{\kappa r}{2}\right) \\
&\leqslant Cn^{-\frac{r-p}{2}}2^{-j_0\varepsilon} \\
&= Cn^{-\frac{r-p}{2}}2^{\max\{-j_0\varepsilon,-j_1\varepsilon\}}.
\end{aligned} \tag{2.28}$$

Dabei fiel das oben gewählte $\kappa \in \mathbb{R}$ nach Anwendung von (2.21) weg.

Falls $\varepsilon < 0$ und $r > 2$ ist, wählt man ein $\kappa > 0$, sodass $-\varepsilon - \frac{\kappa}{2} > 0$ gilt. Wie in der Abschätzung (2.28) ergibt sich zunächst

$$E\,\|e_{bb}\|_{L^r}^r \leqslant Cn^{-\frac{r-p}{2}}S(\kappa\xi)^{\left(\frac{r}{2}-1\right)}S\left(-\varepsilon-\frac{\kappa r}{2}\right),$$

und wieder mit (2.21)

$$E\,\|e_{bb}\|_{L^r}^r \leqslant Cn^{-\frac{r-p}{2}}2^{-j_1\varepsilon} = Cn^{-\frac{r-p}{2}}2^{\max\{-j_0\varepsilon,-j_1\varepsilon\}}.$$

Für $\varepsilon \neq 0$ und $r \in [1,2]$ gilt nach (2.20) und (2.22), sowie wegen $j^{-\frac{p}{2}} \leqslant 1$,

$$
\begin{aligned}
E\,\|e_{bb}\|_{L^r}^r &\leqslant C \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1\right)} \sum_{k\in\mathbb{Z}} E\big|\hat{f}_{jk}\big|^r \\
&\leqslant C \sum_{j=j_0}^{j_1} 2^{j\left(\frac{r}{2}-1\right)} n^{-\frac{r-p}{2}} j^{-\frac{p}{2}} 2^{-j(\sigma+\frac{1}{2}-\frac{1}{p})p} M^p \\
&\leqslant C n^{-\frac{r-p}{2}} S\left(\frac{r}{2} - 1 - \sigma p - \frac{p}{2} + 1\right) \\
&\leqslant C n^{-\frac{r-p}{2}} S(-\varepsilon) \\
&\leqslant C n^{-\frac{r-p}{2}} 2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}}.
\end{aligned}
$$

Das war zu zeigen.

Ist $\varepsilon = 0$, so folgt wegen der Voraussetzung $\sigma p > 1$ direkt

$$1 < \sigma p = \frac{r-p}{2} \implies 2 + p < r \implies r > 2.$$

Setzt man in (2.22) nun $\kappa = 0$, ergibt sich hier wegen $f \in B_{\sigma pq}(M)$

$$
\begin{aligned}
E\,\|e_{bb}\|_{L^r}^r &\leqslant C M^p S(0)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-1)} C 2^{-j(\sigma+\frac{1}{2}-\frac{1}{p})p} j^{-\frac{p}{2}} n^{-\frac{r-p}{2}} \\
&\leqslant C M^p (j_1 - j_0)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j(\frac{r}{2}-\sigma p-\frac{p}{2})} j^{-\frac{p}{2}} n^{-\frac{r-p}{2}} \\
&= C (j_1 - j_0)^{\frac{r}{2}-1} \sum_{j=j_0}^{j_1} 2^{j\cdot 0} j^{-\frac{p}{2}} n^{-\frac{r-p}{2}} \\
&\leqslant C (j_1 - j_0)^{\frac{r}{2}-1} n^{-\frac{r-p}{2}} \sum_{j=j_0}^{j_1} j_0^{-\frac{p}{2}} \\
&= C (j_1 - j_0)^{\frac{r}{2}} n^{-\frac{r-p}{2}} j_0^{-\frac{p}{2}} \\
&= C \left(\frac{j_1}{j_0} - 1\right)^{\frac{r}{2}} j_0^{\frac{r}{2}} n^{-\frac{r-p}{2}} j_0^{-\frac{p}{2}} \\
&\asymp C \left(\frac{r}{r-2} - 1\right)^{\frac{r}{2}} n^{-\frac{r-p}{2}} j_0^{\frac{r-p}{2}} \leqslant C n^{-\frac{r-p}{2}} j_1^{\frac{r-p}{2}},
\end{aligned}
$$

wobei Hilfssatz 2.10 verwendet wurde. Das war zu zeigen.

(*ii*) Sei zunächst $\varepsilon > 0$. Dann muss nach Voraussetzung $\max\{-j_0\varepsilon, -j_1\varepsilon\} = -j_0\varepsilon$ sein und daher mit (2.9) und (2.6)

$$
\begin{aligned}
n^{-\frac{r-p}{2}} 2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}} &= n^{\varepsilon(1-2\alpha)-\alpha r} 2^{-j_0\varepsilon} \\
&\asymp n^{\varepsilon(1-2\alpha)-\alpha r} n^{-\varepsilon(1-2\alpha)} (\log n)^{-\frac{r-p}{p}\varepsilon(1-2\alpha)} \\
&= n^{-\alpha r} (\log n)^{-\frac{r-p}{p}\varepsilon(1-2\alpha)}.
\end{aligned}
$$

Der $n^{-\alpha r}$-Term ist bereits der gleiche wie in der Behauptung von Hauptsatz 2.1, es ist also nur zu zeigen, dass die Potenz des $(\log n)$-Terms höchstens αr ist. Dies folgt aber aus Hilfssatz 2.6.

Für $\varepsilon < 0$ wird ähnlich verfahren. Dann ist nämlich $\max\{-j_0\varepsilon, -j_1\varepsilon\} = -j_1\varepsilon$ und wieder nach (2.9) und (2.6)

$$\begin{aligned} n^{-\frac{r-p}{2}} 2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}} = n^{\frac{\varepsilon\alpha}{\sigma'}-\alpha r} 2^{-j_1\varepsilon} &\asymp n^{\frac{\varepsilon\alpha}{\sigma'}-\alpha r} n^{-\frac{\varepsilon\alpha}{\sigma'}} (\log n)^{\frac{\alpha\varepsilon}{\sigma'}} \\ &= n^{-\alpha r} (\log n)^{\frac{\alpha\varepsilon}{\sigma'}}. \end{aligned}$$

Wieder erreicht der $n^{-\alpha r}$-Term die in (2.7) behauptete Konvergenzrate; dass der $(\log n)$-Term vernachlässigbar ist, folgt aus der Äquivalenz

$$\frac{\alpha\varepsilon}{\sigma'} < \alpha r \iff \varepsilon < \sigma' r.$$

Es ist nämlich nach Voraussetzung $\varepsilon < 0$ und $\sigma' r > 0$ weil $\sigma p > 1$, $r \geqslant p$ und

$$\sigma' r \geqslant \sigma' p = \sigma p - \frac{p}{p} + \frac{p}{r} > 1 - 1 + \frac{p}{r} \geqslant 0.$$

Sei $\varepsilon = 0$. Die in (i) für diesen Fall gezeigte Schranke genügt der Behauptung aus Hauptsatz 2.1, denn

$$\begin{aligned} \left(\frac{j_1}{n}\right)^{\frac{r-p}{2}} &= n^{-\frac{r-p}{2}} \left(\log_2 \left(\frac{n}{\log n}\right)^{\frac{\alpha}{\sigma'}}\right)^{\frac{r-p}{2}} \\ &= n^{-\frac{r-p}{2}} \left(\frac{\alpha}{\sigma'} (\log_2 n - \log_2 \log n)\right)^{\frac{r-p}{2}} \\ &\leqslant n^{-\frac{r-p}{2}} \left(\frac{\alpha}{\sigma'} \log_2 n\right)^{\frac{r-p}{2}} \\ &= C n^{-\frac{r-p}{2}} (\log n)^{\frac{r-p}{2}} \\ &= C n^{-\alpha r} (\log n)^{\alpha r} \end{aligned}$$

wegen $\varepsilon = 0$ und (2.9). Dies ist entweder genau die behauptete Konvergenzrate oder ist gegen diese vernachlässigbar. ▨

Der letzte abzuschätzende Term ist e_{ss}, der den Fall behandelt, dass sowohl Schätzer als auch wahrer Waveletkoeffizient betragsmäßig „kleine“ Werte annehmen.

2.31 Satz: *Sei $j_0 \leqslant j_1$. Der Term $\|e_{ss}\|_{L^r}$ erreicht höchstens die in (2.7) beschriebene Rate.*

Die Hauptlast des Beweises trägt ein Ergebnis des Artikels [5]. An die Notation und die Voraussetzungen dieser Arbeit angepasst besagt nämlich [5, Thm. 3]:

2.32 Satz: *Sei $r \geqslant p \geqslant 1$, $\|\cdot\|_{0r(r\wedge 2)}$ eine Besovnorm und $B_{\sigma pq}(M)$ ein Besovraum. Sei weiter*

$$\Omega\big(\delta; \|\cdot\|_{0r(r\wedge 2)}, B_{\sigma pq}(M)\big) := \sup\left\{\|f\|_{0r(r\wedge 2)} : f \in B_{\sigma pq}(M),\ |\beta_{jk}| < \delta \text{ für alle } j,k\right\},$$

wobei die β_{jk} die Koeffizienten in der Waveletentwicklung von f sind. Dann gilt für alle genügend kleinen $\delta > 0$:

$$\Omega\big(\delta; \|\cdot\|_{0r(r\wedge 2)}, B_{\sigma pq}(M)\big) \asymp M^{1-2\alpha}\delta^{2\alpha}\left(\log\frac{M}{\delta}\right)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)^{+}}.$$

Beweis (von Satz 2.31). Zunächst ergibt sich mit Satz 1.6

$$\begin{aligned}\|e_{ss}\|_{L^r} = \left\|\sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}}\beta_{jk}\psi_{jk}\mathbb{1}_{\{k\in\hat{S}_j\cap S'_j\}}\right\|_{L^r} &\leqslant \left\|\sum_{j=j_0}^{j_1}\sum_{k\in S'_j}\beta_{jk}\psi_{jk}\right\|_{L^r}\\ &\leqslant \left\|\sum_{j=j_0}^{j_1}\sum_{k\in S'_j}\beta_{jk}\psi_{jk}\right\|_{0r(r\wedge 2)}.\end{aligned}$$

Nach Definition der Menge S'_j gilt für alle β_{jk} mit $k \in S'_j$ die Abschätzung

$$|\beta_{jk}| < 2K\sqrt{\frac{j}{n}} \leqslant 2K\sqrt{\frac{j_1}{n}} =: \delta_n. \tag{2.29}$$

Da es um asymptotische Betrachtungen für $n \to \infty$ geht und da nach (2.6) j_1 sublinear (nämlich im Wesentlichen logarithmisch) wächst, wird schließlich δ_n „genügend klein“ und Satz 2.32 damit anwendbar. Nach obiger Rechnung folgt also, da die β_{jk} deterministisch sind,

$$\begin{aligned}\left(E\,\|e_{ss}\|_{L^r}^r\right)^{\frac{1}{r}} &\leqslant \left(E\left\|\sum_{j=j_0}^{j_1}\sum_{k\in S'_j}\beta_{jk}\psi_{jk}\right\|_{0r(r\wedge 2)}^r\right)^{\frac{1}{r}}\\ &= \left\|\sum_{j=j_0}^{j_1}\sum_{k\in S'_j}\beta_{jk}\psi_{jk}\right\|_{0r(r\wedge 2)}\\ &\leqslant \left\|\sum_{k\in\mathbb{Z}}\alpha_{j_0k}\varphi_{j_0k} + \sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}}\mathbb{1}_{S'_j}(k)\beta_{jk}\psi_{jk}\right\|_{0r(r\wedge 2)}\end{aligned}$$

Die Funktion $\tilde{f} := \sum_{k\in\mathbb{Z}}\alpha_{j_0k}\varphi_{j_0k}+\sum_{j\geqslant j_0}\sum_{k\in S'_j}\beta_{jk}\psi_{jk}$ liegt offensichtlich im Raum $B_{\sigma pq}(M)$ (sie hat bis auf den zusätzlichen Term $\mathbb{1}_{S'_j}(k)$ die gleichen Koeffizienten wie $f \in B_{\sigma pq}(M)$) und nach (2.29) gilt $\left|\mathbb{1}_{S'_j}(k)\beta_{jk}\right| < \delta_n$ für alle $j, k \in \mathbb{Z}$. Damit ist

$$\left(E\,\|e_{ss}\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant \|\tilde{f}\|_{0r(r\wedge 2)} \leqslant \Omega\big(\delta_n; \|\cdot\|_{0r(r\wedge 2)}, B_{\sigma pq}(M)\big).$$

Satz 2.32 liefert nun

$$\left(E\,\|e_{ss}\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant CM^{1-2\alpha}\left(2K\sqrt{\frac{j_1}{n}}\right)^{2\alpha}\left(\log\frac{M\sqrt{n}}{2K\sqrt{j_1}}\right)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)^{+}}.$$

Nun gilt nach (2.6) $2^{j_1} \asymp \left(\frac{n}{\log n}\right)^{\frac{\alpha}{\sigma'}}$, also $j_1 \asymp \log \frac{n}{\log n} = \log n - \log\log n$. Damit folgt

$$\begin{aligned}
(E\,\|e_{ss}\|^r_{L^r})^{\frac{1}{r}} &\leqslant CM^{1-2\alpha}(2K)^{2\alpha}\left(\frac{j_1}{n}\right)^{\alpha}\left(\log\frac{M}{2K}+\log\frac{\sqrt{n}}{\sqrt{j_1}}\right)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)^+}\\
&= C\left(\frac{j_1}{n}\right)^{\alpha}\left(\log\frac{\sqrt{n}}{\sqrt{j_1}}\right)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)^+}\\
&\asymp \left(\frac{\log n-\log\log n}{n}\right)^{\alpha}\left(\log\frac{n}{\log n-\log\log n}\right)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)^+}\\
&\leqslant \left(\frac{\log n}{n}\right)^{\alpha}(\log n)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)^+}.
\end{aligned}$$

Nach Hilfssatz 2.4 ergibt sich schließlich $1-2\alpha = \frac{p}{r}$ für den Fall $\varepsilon = 0$, und damit:

$$(E\,\|e_{ss}\|^r_{L^r})^{\frac{1}{r}} \leqslant \left(\frac{\log n}{n}\right)^{\alpha}(\log n)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{p}{rq}\right)^+},$$

also

$$E\,\|e_{ss}\|^r_{L^r} \leqslant \left(\frac{\log n}{n}\right)^{\alpha r}(\log n)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{r}{2}-\frac{p}{q}\right)^+}.$$

Das ist genau die Konvergenzrate aus der Behauptung. ▨

Damit sind alle auftretenden Terme abgeschätzt und Hauptsatz 2.1 ist bewiesen.

2.5 Verallgemeinerungen und Varianten

Hier sollen Varianten von Hauptsatz 2.1, zum Teil ohne Beweis, dargestellt werden. Zunächst eine triviale Umformung, die in den vorangegangenen Abschnitten vermieden wurde, da sie im Beweis zu mehr Schreibaufwand geführt hätte. In der Literatur ist die Angabe von Konvergenzraten in der folgenden Form üblicher:

2.33 Korollar: *Unter den Voraussetzungen von Hauptsatz 2.1 gilt*

$$\sup_{f\in\mathcal{D}_{\sigma pq}(M,T)} \left(E\,\|\mathrm{TW}-f\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant \begin{cases} C\left(\frac{\log n}{n}\right)^{\alpha}, & \varepsilon \neq 0,\\ C(\log n)^{\left(\frac{1}{2}-\frac{p}{qr}\right)^+}\left(\frac{\log n}{n}\right)^{\alpha}, & \varepsilon = 0.\end{cases}$$

Dies folgt sofort durch Potenzieren beider Seiten mit r. Ohne Beweis sei noch folgende Aussage angegeben:

2.34 Satz: *Die Aussage von Hauptsatz 2.1 gilt auch, falls $r=\infty$ ist. Dann muss die L^r-Norm durch die Supremums-Norm ersetzt werden.*

Man kann den Schätzer TW abwandeln, indem man andere Schwellenwerte für das harte Thresholding zulässt. Wählt man etwa für die Koeffizienten β_{jk} im Level j den Schwellenwert $K\sqrt{\frac{j-j_0}{n}}$, so sind einige der logarithmischen Terme verzichtbar (vgl. [9, Rem. 10.5]). Darauf soll hier nicht näher eingegangen werden.

Ein zentraler Punkt dieses Abschnitts ist die Verallgemeinerung des Hauptsatzes an der Stelle, an der in [4] der Beweis fehlerhaft war. Im Beweis zu Hauptsatz 2.1 wurde dies behoben, indem die Behauptung abgeschwächt wurde; hier soll die Behauptung in der Schärfe gezeigt werden, die in [4] vorkommt, dann allerdings unter Verschärfung der Voraussetzungen. Allerdings geht es hierbei nur um die Verbesserung der Konvergenzrate, nicht um die „neue“ Voraussetzung $r \leqslant \frac{1+2\sigma}{\sigma}$. Diese Bedingung an r ist notwendig für die Gültigkeit einiger der Abschätzungen im Beweis von Hauptsatz 2.1, und kann daher nicht ohne weiteres weggelassen werden. Um sich dieser Bedingung entledigen zu können, bedürfte es eines anderen Beweisansatzes.

2.35 Hauptsatz: *Unter den Voraussetzungen von Hauptsatz 2.1 und*

$$2^{j_0(n)} \asymp \left(n(\log n)^{\frac{r-p}{p}\mathbb{1}_{\{\varepsilon\geqslant 0\}}}\right)^{1-2\alpha}, \qquad 2^{j_1(n)} \asymp \left(\frac{n}{\log n}\right)^{\frac{\alpha}{\sigma'}}(\log n)^{\mathbb{1}_{\{\varepsilon>0\}}\frac{\alpha\varepsilon}{\sigma p\sigma'}} \tag{2.30}$$

gilt

$$\sup_{f\in\mathcal{D}_{\sigma pq}(M,T)} \left(E\,\|\mathrm{TW}-f\|_{L^r}^r\right) \leqslant \begin{cases} C(\log n)^{(1-\frac{\varepsilon}{\sigma p})\alpha r}n^{-\alpha r}, & \varepsilon > 0,\\ C(\log n)^{\left(\frac{r}{2}-\frac{p}{q}\right)^+}\left(\frac{\log n}{n}\right)^{\alpha r}, & \varepsilon = 0\\ C\left(\frac{\log n}{n}\right)^{\alpha r}, & \varepsilon < 0,\end{cases} \tag{2.31}$$

Nur die Wahl von j_1 ist verändert worden, und das auch nur im Fall $\varepsilon > 0$. Nur für diesen Fall wird eine Konvergenzrate behauptet, die sich von der aus Hauptsatz 2.1 unterscheidet. Die hier behauptete Konvergenzrate ist eine strikt bessere, da nach Voraussetzung $\sigma p > 1$

und in diesem Fall $\varepsilon > 0$ gilt; es ist also eine Verbesserung für „ähnliche Werte von r und p“ (wegen (2.5)).

Die Wahl von j_1 wurde so getroffen, dass der Biasterm keine Schwierigkeiten mehr bereitet, vgl. Satz 2.26. Da mehr Detailterme den Bias von TW verringern (es sind dann „mehr lokale Details“ einer Funktion darstellbar), musste die Zahl dieser Detailterme vergrößert werden. Dies ist in (2.30) geschehen.

Beweis. Alle Abschätzungen aus dem Beweis von Hauptsatz 2.1 bleiben gültig. Da sich nur j_1 und die Schranke im Fall $\varepsilon > 0$ verändert hat, bleiben alle Überlegungen zu Vernachlässigbarkeit bestehen, sofern darin nur j_0 vorkommt. Von Interesse sind daher nur die Überlegungen aus den Beweisen zu Satz 2.26, Satz 2.30 und Satz 2.31.

In Satz 2.26 wurde die Abschätzung

$$\left\|f - E_{j_1(n)} f\right\|_{L^r}^r \leqslant C 2^{-j_1(n)\sigma' r}.$$

gezeigt. Mit der Wahl aus (2.30) verhält die rechte Seite dieser Ungleichung sich asymptotisch wie

$$C\left(\left(\frac{n}{\log n}\right)^{-\frac{\alpha}{\sigma'}} (\log n)^{-\mathbb{1}_{\{\varepsilon>0\}}\frac{\alpha\varepsilon}{\sigma p \sigma'}}\right)^{\sigma' r} = \begin{cases} C n^{-\alpha r} (\log n)^{\alpha r - \frac{\alpha \varepsilon r}{\sigma p}}, & \text{für } \varepsilon > 0 \\ C n^{-\alpha r} (\log n)^{\alpha r}, & \text{für } \varepsilon \leqslant 0. \end{cases}$$

Offensichtlich ändert sich im Fall $\varepsilon \leqslant 0$ nichts und die Schranke ist nach wie vor vernachlässigbar gegen die Konvergenzraten in (2.31). Im Fall $\varepsilon > 0$ ist dies jedoch genau die behauptete Konvergenzrate aus Hauptsatz 2.35.

Die Behauptung aus Satz 2.30 war

$$E\left\|e_{bb}\right\|_{L^r}^r \leqslant \begin{cases} C 2^{\max\{-j_0\varepsilon, -j_1\varepsilon\}} n^{-\frac{r-p}{2}} & \text{für } \varepsilon \neq 0 \\ C\left(\frac{j_1}{n}\right)^{\frac{r-p}{2}} & \text{für } \varepsilon = 0 \end{cases}.$$

Hier spielt die Änderung von j_1 allerdings keine Rolle, denn im einzig relevanten Fall $\varepsilon > 0$ ist $\max\{-j_0\varepsilon, -j_1\varepsilon\} = -j_0\varepsilon$ (sofern $j_0 \leqslant j_1$ ist; sonst tritt gar kein nichtlinearer Term auf und die hier angestellten Überlegungen sind unnötig; vgl. Satz 2.27). Folglich ändert sich auch nichts an dieser Konvergenzrate und sie genügt der Behauptung aus Hauptsatz 2.35.

Die Betrachtungen im Beweis zu Satz 2.31 reichen allerdings nicht aus, um auch die Schranke aus Hauptsatz 2.35 zu zeigen. Zunächst wird genauso wie in Satz 2.31 vorgegangen, um

$$\left(E\left\|e_{ss}\right\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant M^{1-2\alpha}\left(2K\sqrt{\frac{j_1}{n}}\right)^{2\alpha}\left(\log\frac{M\sqrt{n}}{2K\sqrt{j_1}}\right)^{\mathbb{1}_{\{\varepsilon=0\}}\left(\frac{1}{2}-\frac{1-2\alpha}{q}\right)^+}$$

zu erhalten. Der letzte Faktor fällt dabei allerdings weg, da sich die Änderungen in Hauptsatz 2.35 nur auf den Fall $\varepsilon > 0$ beziehen. Man bekommt also die Abschätzung

$$\left(E\left\|e_{ss}\right\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant C\left(\frac{j_1}{n}\right)^{\alpha}.$$

Es gilt auch nach der neuen Wahl von j_1 in (2.30)

$$j_1 \asymp \frac{\alpha}{\sigma'}(\log n - \log\log n) + \frac{\alpha\varepsilon}{\sigma p \sigma'} \log\log n = \frac{\alpha}{\sigma'}\left(\log n - \left(1 - \frac{\varepsilon}{\sigma p}\right)\log\log n\right) \leqslant C \log n,$$

wobei Hilfssatz 2.9 benutzt wurde. Das ist die obere Schranke, die auch in Satz 2.31 gezeigt wurde.

Um zu zeigen, dass im Fall $\varepsilon > 0$ auch die schärfere Konvergenzrate aus Hauptsatz 2.35 eingehalten wird, ist eine nähere Untersuchung des Beweises von [5, Thm. 3] nötig. Diese Verschärfung ist, kurz gesagt, deswegen möglich, weil in (2.30) nur solche Levels von j betrachtet werden, die eine gute Auflösung der Funktion f liefern. In [5] wird nicht von einer solchen Wahl von j ausgegangen und daher eine allgemeinere, aber schwächere Abschätzung gezeigt. Für den vorliegenden Spezialfall lässt sich gerade die Behauptung von Hauptsatz 2.35 folgern. $\boxtimes$

Kapitel 3

Andere Waveletschätzer und ihre Eigenschaften

Die in diesem Kapitel vorgestellten Ergebnisse, sollen dazu dienen, Hauptsatz 2.1 zu ergänzen und einzuordnen. Die Sätze werden hier lediglich diskutiert, sowie Beweisskizzen und -ideen angegeben; näheres zu den Beweisdetails findet sich in den aufgeführten Referenzen.

3.1 Lineare Schätzer

Zunächst sollen lineare Schätzer betrachtet werden, bei denen insbesondere kein Thresholding verwendet wird. Die Menge aller linearen Dichteschätzer sei

$$\mathscr{C}_L := \left\{ \hat{f}(X_1, \ldots, X_n, x) = \sum_{i=1}^{n} T_i(X_i, x) \colon T_i(\cdot, \cdot) \text{ messbar} \right\}. \tag{3.1}$$

Der Schätzer TW aus (2.4) ist kein linearer Schätzer, da durch das Thresholding die $\tilde{\beta}_{jk}$ nicht mehr die hier angegebene Gestalt haben. Nur der erste Summand von TW kann als linearer Schätzer aufgefasst werden.

Der folgende Satz gibt eine Schranke für die Verlustfunktion linearer Schätzer an und zeigt, dass der lineare Wavelet-Schätzer, der in [10] angegeben wird, optimal ist. Die Notationen sind die gleichen wie in Kapitel 2.

3.1 Satz ([4, Thm. 1]):

(*i*) *Sei* $1 \leqslant p, q < \infty$, $p \leqslant r$, $2 \leqslant r < \infty$ *und* $\sigma > \frac{1}{p}$. *Sei weiter*

$$R_n^L := \inf_{\hat{f}_n \in \mathscr{C}_L} \sup_{f \in \mathcal{D}_{\sigma pq}(M)} \left(E_f \left\| \hat{f} - f \right\|_r^r \right)^{\frac{1}{r}}.$$

Dann gibt es Konstanten C_1, $C_2 \in \mathbb{R}^+$ *mit*

$$C_1 n^{-\frac{\sigma'}{1+2\sigma'}} \leqslant R_n^L \leqslant C_2 n^{-\frac{\sigma'}{1+2\sigma'}}, \qquad \sigma' := \sigma - \frac{1}{p} + \frac{1}{r}.$$

(*ii*) *Für* $1 \leqslant r < 2$ *sei* $\omega \in L^{\frac{r}{2}}(\mathbb{R})$ *eine gerade, auf* $\mathbb{R}^+$ *nichtfallende Funktion und*

$$\mathscr{N}_\omega := \{f\colon \exists a \in \mathbb{R} \text{ mit } f(x-a) \leqslant \omega(x) \text{ für alle } x \in \mathbb{R}\}\,.$$

Dann gilt die gleiche Aussage wie oben mit

$$R_n^L := \inf_{\hat{f}_n \in \mathscr{C}_L} \sup_{f \in \mathcal{D}_{\sigma pq}(M) \cap \mathscr{N}_r} \left(E_f \left\|\hat{f} - f\right\|_r^r\right)^{\frac{1}{r}}.$$

(*iii*) *Die Aussagen bleiben gültig, wenn* $r = \infty$ *ist, falls* $\mathcal{D}_{\sigma pq}(M)$ *durch* $\mathcal{D}_{\sigma pq}(M,T)$ *und* n *durch* $\frac{n}{\log n}$ *ersetzt wird.*

Das Minimaxrisiko linearer Schätzer erreicht also genau die Rate $n^{-\frac{\sigma'}{1+2\sigma'}}$. Diese Rate ist im Fall $r \neq p$, echt schlechter als die Konvergenzgeschwindigkeit von TW. Dort wurde die Rate $n^{-\frac{\sigma}{1+2\sigma}}(\log n)^\alpha$ erreicht (wobei die logarithmischen Terme die Konvergenz zwar verschlechtern, aber wegen der schlechteren Exponenten nicht wesentlich).

In [10, Eq. (3.1)] wird ein Schätzer angegeben, der die hier angeführte Rate erreicht; er entspricht im Wesentlichen dem ersten Summanden des Schätzers TW, wobei $j_0(n) \asymp n^{\frac{\alpha}{\sigma}}$ gewählt wird. Aus dem Beweis, dass der angegebene Schätzer tatsächlich diese Konvergenzrate einhalten kann, kommt auch die Bedingung aus (*ii*), dass es eine f dominierende $L^{\frac{r}{2}}$-Funktion gibt.

Der Beweis, dass der Schätzer aus [10] tatsächlich eine Minimaxschranke erreicht, dass also die angegebene Konvergenzrate optimal für lineare Schätzer ist, soll hier nicht geführt werden. Das Argument ist in [4, Sect. A.2] zu finden. Dort wird eine Teilmenge von Dichten in $\mathcal{D}_{\sigma pq}(M)$ betrachtet, die in dem Sinne „am schwersten“ zu schätzen ist, dass die Verlustfunktion über diese Funktionen mindestens die in der Behauptung angegebene Rate hat. Da eine schlechtere Konvergenzrate als diese unmöglich ist, ist die Rate über *alle* Funktionen in $\mathcal{D}_{\sigma pq}(M)$ gerade die aus der Behauptung.

Der Fall $r = \infty$ wird auf ähnliche Weise behandelt. Dabei sind allerdings andere technische Abschätzungen notwendig, die besondere Aufmerksamkeit erfordern. Die Beweisidee bleibt die gleiche.

3.2 Untere Schranken für Waveletschätzer

3.2 Satz ([4, Thm. 2]): *Sei* $1 \leqslant p, q \leqslant \infty$, $r \geqslant p$, $\sigma > \frac{1}{p}$ *und*

$$R_n := \inf_{\hat{f}} \sup_{f \in \mathcal{D}_{\sigma pq}(M)} \left(E_f \left\| \hat{f}_n - f \right\|_r^r \right)^{\frac{1}{r}},$$

wobei das Infimum über alle Schätzer $\hat{f}$ *betrachtet wird, die Werte in* $V \supset \mathcal{D}_{\sigma pq}(M)$ *annehmen. Dann gibt es eine Konstante* $C > 0$*, sodass gilt*

$$R_n \geqslant \begin{cases} C \left(\frac{\log n}{n} \right)^{\alpha}, & \varepsilon \leqslant 0, \\ C n^{-\alpha}, & \varepsilon > 0. \end{cases}$$

Es kann also keinen Schätzer für Besovdichten geben, der besser als mit der angegebenen Rate konvergiert. Ein Vergleich mit Hauptsatz 2.1 zeigt, dass diese untere Schranke tatsächlich in einigen Fällen angenommen wird. Insbesondere kann der Schätzer TW im Fall $\varepsilon < 0$ und im Fall $\varepsilon = 0$, $\frac{1}{2} - \frac{pq}{r} \leqslant 0$ nicht verbessert werden, da er bereits die optimale Rate erreicht. Für $\varepsilon > 0$ ist TW noch nicht optimal, da ein logarithmischer Term dazu kommt (auch in der Verallgemeinerung aus Hauptsatz 2.35); vgl. dazu den nächsten Abschnitt.

Vergleicht man diesen Satz mit Satz 3.1, so wird klar, dass die im vorigen Abschnitt betrachteten linearen Schätzer ebenfalls nicht optimal sein können, falls $r \neq p$ ist (die linearen Schätzer sind gerade für $r = p$ und $\varepsilon > 0$ optimal). Zur Optimierung in den übrigen Fällen müssen daher, wie es bei TW geschehen ist, nichtlineare Terme hinzugefügt werden.

Man bezeichnet diejenigen Fälle, für die $\varepsilon \leqslant 0$ ist, als „*spärlichen Bereich*“ („sparse zone“), die übrigen als „*regulären Bereich*“ („regular zone“). Diese Namen sind aus dem Verhalten derjenigen Funktionen entstanden, die in den entsprechenden Fällen am schwersten abzuschätzen sind.

Im so genannten spärlichen Bereich sind dies Funktionen, die an den meisten Stellen glatt sind und nur in kleinen Bereichen irreguläres Verhalten haben. Aus diesem Grund sind in der zugehörigen Waveletentwicklung nur wenige der Detailkoeffizienten β_{jk} von Null verschieden, die Entwicklung ist in diesem Sinne „spärlich“ oder „sparse“.

Beim regulären Bereich sind die am schwersten abzuschätzenden Funktionen solche mit starken Oszillationen, die sich gleichmäßig entlang der reellen Achse verteilen. Der Begriff „regulär“ steht also weniger für eine Glattheitseigenschaft als vielmehr für eine gleichmäßige Wiederholung des selben Effekts.

Der Beweis von Satz 3.2 ist in [9, S. 155-163] ausführlich dargestellt.

Die Grundidee des Beweises ist, eine Teilklasse von Funktionen anzugeben, die die schlechteste Konvergenzrate hervorbringt. Dies sind, abhängig von $\operatorname{sgn}\varepsilon$ gerade die oben beschriebenen Funktionen. Für diese Teilklassen sind untere Schranken zu bestimmen, was für den spärlichen Bereich mit dem Lemma von Fano (vgl. [13, Lem. 2.8]) und für den regulären Bereich mit dem Lemma von Assouad (vgl. [13, Lem. 2.10]) erreicht wird. Diese Aussagen garantieren bereits die Existenz gewisser unterer Schranken für die Konvergenzrate, die technische Arbeit besteht danach noch darin, die Voraussetzungen zu überprüfen und die Konvergenzrate konkret anzugeben.

3.3 Optimalitätseigenschaften bei einer L^2-Verlustfunktion

Falls $r = 2$ ist, können Schätzer angegeben werden, die die optimale Konvergenzrate aus Satz 3.2 zumindest für $\varepsilon > 0$ erreichen. Dafür wird im Gegensatz zu den Ausführungen in Kapitel 2 kein hartes, sondern weiches Thresholding verwendet (vgl. (2.3)).

Der zentrale Unterschied zu hartem Thresholding ist, dass die Waveletkoeffizienten im hier vorliegenden Fall unterhalb vom Schwellenwert λ nicht mehr auf Null gesetzt werden, sondern nur betragsmäßig verringert werden. Die Information aus diesen Koeffizienten geht daher nicht verloren, sondern wird nur weniger stark gewichtet. Das führt dazu, dass die Approximation an die „wahre" Funktion f grundsätzlich besser ist, denn auch geringe Schwankungen können berücksichtigt werden. Allerdings werden kleine Störungen (etwa Rauschen, siehe unten) nicht mehr so stark unterdrückt wie in den früher behandelten Fällen.

3.3 Satz ([4, Thm. 4]): *Sei entweder $p \geqslant 1$ und $\sigma p > 1$ oder $\sigma = \frac{1}{p}$ und $p > 1$. Betrachte den Schätzer*

$$\hat{f}(x) = \sum_{k\in\mathbb{Z}} \hat{\alpha}_{j_1 k}\varphi_{j_1 k}(x) + \sum_{j=j_0}^{j_1}\sum_{k\in\mathbb{Z}} \delta_s(\beta_{jk}, \lambda)\psi_{jk}(x),$$

mit geeigneten $j_0, j_1 \in \mathbb{Z}$ und $(\lambda_j)_j \subset \mathbb{Z}$. Dann gibt es eine Konstante $C > 0$, so dass für $n \gg 0$ und $\varepsilon > 0$ gilt

$$\inf_{\hat{f}_n} \sup_{f\in\mathcal{D}_{\sigma pq(M,T)}} E\left\|\hat{f}_n - f\right\|_{L^2}^2 \leqslant Cn^{-\alpha}.$$

Dieser tief liegende Satz verwendet im Beweis Ergebnisse aus der Schrift [3]. Dort werden allgemeine Eigenschaften der L^2-Verlustfunktion hergeleitet und diskutiert, die dann in einem Modell von Gaußschem Weißen Rauschen angewendet werden:

$$Y_{jk} = \vartheta_{jk} + \varepsilon Z_{jk}, \qquad j \geqslant 0,\ k \in \{0, 1, \dots, 2^{j-1}\},$$

wobei $Z_{jk} \sim \mathcal{N}(0,1)$ i.i.d. Zufallsvariablen, $\varepsilon > 0$ und ϑ_{jk} unbekannt sind. Die ϑ_{jk} sollen geschätzt werden, wobei man die Y_{jk} als Beobachtungen auffasst, die durch normalverteilte Zufallsvariable gestört („verrauscht") sind.

Verschiedene Abschätzungen, unter anderem Satz 2.12 und ähnliche Ungleichungen, führen dann zur Behauptung. Der Beweis ist in [4, Sect. 6] ausführlicher dargestellt.

Die Aussage von Satz 3.3 ist in dem Sinne eine Ergänzung zu Hauptsatz 2.1 und Satz 3.2, als ein optimaler Schätzer für den Fall $\varepsilon > 0$ angegeben wird. Man ist allerdings auf die Wahl $r = 2$ festgelegt, wenn diese Optimalität erreicht werden soll. Der Schätzer TW wurde bereits als optimal für den Fall $\varepsilon < 0$ und gewisse Wahlen der Parameter bei $\varepsilon = 0$ herausgestellt. Nur für den Fall $\varepsilon = 0$, $\frac{1}{2} - \frac{pq}{r} > 0$ ist kein optimaler Schätzer bekannt geworden.

Als der Artikel [4] veröffentlicht wurde, gab es Hinweise darauf, dass die hierbei verwendete Technik auch für den Fall $r \neq 2$ zu verallgemeinern wäre. Allerdings hat sich kein entsprechendes Ergebnis in der Literatur finden lassen.

3.4 Adaptive Waveletschätzer

Ein zentrales Problem der Dichteschätzung ist, dass im Vorfeld nicht bekannt ist, in welchem Besovraum die zu schätzende Dichte überhaupt liegt. Welche Konvergenzgeschwindigkeit der Schätzer TW also tatsächlich erreicht, kann nicht im Vorhinein bestimmt werden, da σ, p und q unbekannt sind. Insbesondere kann nicht unterschieden werden, ob ε positiv oder negativ ist und folglich, ob der Schätzer TW optimal ist oder $\hat{f}$ gewählt werden sollte. Allerdings kann TW schon nicht konstruiert werden, da die Wahl von j_0 und j_1 in (2.6) direkt von den unbekannten Parametern abhängt.

Um Abhilfe zu schaffen, kann TW so abgeändert werden, dass er *adaptiv* wird, also unabhängig von σ, p und q konstruiert werden kann. Die Konvergenzrate, die dieser adaptive Schätzer erreicht, entspricht der, die in Hauptsatz 2.1 angegeben wurde, sie ist allerdings schlechter als die Rate aus Hauptsatz 2.35.

Sei $s \in \mathbb{N}$ und definiere eine Menge

$$\mathcal{S} := \left\{ (\sigma, p, q, T) \in \mathbb{R}^4 \colon \frac{1}{p} < \sigma \leqslant s,\ 1 \leqslant p, q \leqslant \infty,\ 0 < T < \infty \right\}.$$

Der adaptive Schätzer, der mit ATW bezeichnet wird, soll für alle Wahlen von Parametern aus $\mathcal{S}$ anwendbar sein. Für diese Konstruktion ist also nur noch eine schwache Vorannahme notwendig, nämlich eine Wahl von s. Die Multiskalenanalyse wird dann als $(s+1)$-regulär gefordert und die Struktur des Schätzers TW beibehalten. Nur j_0 und j_1 werden nun, anders als in (2.6), gewählt mit

$$2^{j_0} \asymp n^{\frac{1}{1+2s}}, \qquad 2^{j_1} \asymp \frac{n}{\log n}.$$

Für die Konstante K aus Hauptsatz 2.1 gilt nun $K := c(M, \psi) sr$ (vgl. auch Satz 2.16).

Nach dieser Konstruktion kann man den folgenden Satz zeigen:

3.4 Satz ([4, Thm. 5]): *Es seien $X_1, \dots, X_n$ i.i.d. Zufallsvariablen mit einer Dichtefunktion $f \in \mathcal{D}_{\sigma pq}(M, T)$. Sei $r \geqslant 1$ und $(\sigma, p, q, T) \in \mathcal{S}$. Dann gibt es eine Konstante $C > 0$, sodass*

$$\left(E \left\| \mathrm{ATW} - f \right\|_{L^r}^r\right)^{\frac{1}{r}} \leqslant \begin{cases} C \left(\frac{\log n}{n}\right)^{\alpha}, & \text{für } \varepsilon \neq 0, \\ C (\log n)^{\left(\frac{1}{2} - \frac{p}{qr}\right)^+} \left(\frac{\log n}{n}\right)^{\alpha}, & \text{für } \varepsilon = 0. \end{cases}$$

Der Schätzer ATW ist nach wie vor nicht völlig adaptiv. Die Unbekannte M und der Parameter s kommen noch immer vor. Außerdem muss im Vorfeld bekannt sein, dass die Dichte f kompakten Träger hat. Es gibt jedoch im Modell des Gaußschen Weißen Rauschens auch völlig adaptive Schätzer (vgl. [6], [3]). Auch die Konvergenzrate aus Hauptsatz 2.35 kann von einem adaptiven Schätzer erreicht werden.

Zum Beweis von Satz 3.4 werden die Argumente des Beweises von Hauptsatz 2.1 verallgemeinert. Hierauf soll in diesem Rahmen nicht näher eingegangen werden.

Literaturverzeichnis

[1] Bergh, Jöran ; Löfström, Jörgen: *Interpolation spaces - An Introduction.* New York : Springer, 1976

[2] Daubechies, Ingrid: *Ten Lectures on Wavelets.* Philadelphia : SIAM, 1992

[3] Donoho, David ; Johnstone, Iain: Minimax Estimation via Wavelet Shrinkage / Stanford University. 1996. – Forschungsbericht

[4] Donoho, David ; Johnstone, Iain ; Kerkycharian, Gerard ; Picard, Dominique: Density Estimation by Wavelet Thresholding. In: *Annals of Statistics* 24 (1996), S. 508–539

[5] Donoho, David ; Johnstone, Iain ; Kerkycharian, Gerard ; Picard, Dominique: Universal near minimaxity of Wavelet shrinkage. In: *Festschrift for L. LeCam* (1997), S. 183–218

[6] Donoho, David ; Johnstone, Iain ; Kerkycharian, Gerard ; Picard, Dominique: Wavelet shrinkage: Asymptotia? In: *Journal of the Royal Statistical Society, Series B* 57 (1997), S. 301–369

[7] Elstrodt, Jürgen: *Maß- und Integrationstheorie.* Berlin : Springer Verlag, 2007

[8] Hernandes, Eugenio ; Weiss, Guido: *A first Course on Wavelets.* Upper Saddle River, New Jersey : Prentice Hall, 1996

[9] Härdle, Wolfgang ; Kerkyacharian, Gérard ; Picard, Dominique ; Tsybakov, Alexander: *Wavelets, Approximation and statistical Applications.* New York : Springer, 1998

[10] Kerkyacharian, Gerard ; Picard, Dominique: Density estimation in Besov spaces. In: *Statistics and Probability letters* 13 (1992), S. 15–24

[11] Kerkycharian, Gerard ; Picard, Dominique: Density estimation by kernel and wavelet methods: Optimality of Besov spaces. In: *Statistics and Probability Letters* 18 (1993), S. 327–336

[12] Meyer, Yves: *Ondelettes et opérateurs.* Paris : Hermann, 1990

[13] Tsybakov, Alexander: *Introduction à l'estimation non-parametrique.* Berlin : Springer, 2004