Appl. Comput. Harmon. Anal. 36 (2014) 434–460

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

www.elsevier.com/locate/acha

Extreme value analysis of empirical frame coefficients and implications for denoising by soft-thresholding

Markus Haltmeier [a,*], Axel Munk [b,c]

[a] Department of Mathematics, University of Innsbruck, Technikerstraße 21a, A-6020 Innsbruck, Austria
[b] Statistical Inverse Problems in Biophysics, Max Planck Institute for Biophysical Chemistry, Am Faßberg 11, D-37077 Göttingen, Germany
[c] Institute for Mathematical Stochastics, University of Göttingen, Goldschmidtstraße 7, D-37077 Göttingen, Germany

ARTICLE INFO

Article history:
Received 9 May 2012
Received in revised form 9 July 2013
Accepted 27 July 2013
Available online 2 August 2013
Communicated by Richard Gundy

Keywords:
Denoising
Thresholding estimation
Extreme value analysis
Gumbel distribution
Berman's inequality
Wavelet thresholding
Curvelet thresholding
Translation invariant wavelets
Frames
Redundant dictionaries

ABSTRACT

Denoising by frame thresholding is one of the most basic and efficient methods for recovering a discrete signal or image from data that are corrupted by additive Gaussian white noise. The basic idea is to select a frame of analyzing elements that separates the data in few large coefficients due to the signal and many small coefficients mainly due to the noise ϵ_n. Removing all data coefficients being in magnitude below a certain threshold yields a reconstruction of the original signal. In order to properly balance the amount of noise to be removed and the relevant signal features to be kept, a precise understanding of the statistical properties of thresholding is important. For that purpose we derive the asymptotic distribution of $\max_{\omega\in\Omega_n}|\langle\phi_\omega^n,\epsilon_n\rangle|$ for a wide class of redundant frames $(\phi_\omega^n\colon\ \omega\in\Omega_n)$. Based on our theoretical results we give a rationale for universal extreme value thresholding techniques yielding asymptotically sharp confidence regions and smoothness estimates corresponding to prescribed significance levels. The results cover many frames used in imaging and signal recovery applications, such as redundant wavelet systems, curvelet frames, or unions of bases. We show that 'generically' a standard Gumbel law results as it is known from the case of orthonormal wavelet bases. However, for specific highly redundant frames other limiting laws may occur. We indeed verify that the translation invariant wavelet transform shows a different asymptotic behaviour.

1. Introduction

We consider the problem of estimating a d-signal or image u_n from noisy observations

$$V_n(k)=u_n(k)+\epsilon_n(k),\quad \text{for } k\in I_n:=\{0,\dots,n-1\}^d \text{ with } d\in\mathbb{N}. \tag{1.1}$$

Here $\epsilon_n(k)\sim N(0,\sigma^2)$ are independent normally distributed random variables (the noise), n is the level of discretization, and σ^2 is the variance of the data (the noise level). The signal u_n is assumed to be a discrete approximation of some underlying continuous domain signal obtained by discretizing a function $u:[0,1]^d\to\mathbb{R}$. One may think of the entries of u_n as point samples $u_n(k)=u(k/n)$ on an equidistant grid. However, in some situations it may be more realistic to consider other discretization models. Area samples, for example, are more appropriate in many imaging applications. In this paper we will not pursue this topic further, because most of the presented results do not crucially depend on the particular discretization model as long as u_n can be associated with a function $u_n^*:[0,1]^d\to\mathbb{R}$ (some kind of abstract interpolation) which, in a suitable way, tends to u as $n\to\infty$.

* Corresponding author.
E-mail address: markus.haltmeier@uibk.ac.at (M. Haltmeier).

1063-5203/$ – see front matter
http://dx.doi.org/10.1016/j.acha.2013.07.004

The aim of denoising is to estimate the unknown signal $u_n := (u_n(k)\colon k \in I_n) \in \mathbb{R}^{I_n}$ from the data $V_n := (V_n(k)\colon k \in I_n) \in \mathbb{R}^{I_n}$. The particular estimation procedure we will analyze in detail is soft-thresholding in frames and overcomplete dictionaries. We stress, however, that a similar analysis also applies to different thresholding methods, such as block-thresholding techniques (as considered, for example, in [1–4]).

1.1. Wavelet soft-thresholding

In order to motivate our results for thresholding for general frames we start by one dimensional wavelet soft-thresholding. For that purpose, let $(\psi^n_{j,k}\colon (j,k) \in \Omega_n)$ denote an orthonormal wavelet basis of $\mathbb{R}^n$, where $n = 2^J$ is the number of data points and

$$\Omega_n := \bigl\{(j,k)\colon j \in \{0,\dots,J-1\} \text{ and } k \in \bigl\{0,\dots,2^j-1\bigr\}\bigr\}$$

the index set of the wavelet basis. Wavelet soft-thresholding is by now a standard method for signal and image denoising (see, for example, [5–13] for surveys and some original references). It consists of the following three basic steps:

(1) Compute all empirical wavelet coefficients $Y_n(j,k) = \langle \psi^n_{j,k}, V_n\rangle$ of the given noisy data with respect to the considered orthonormal wavelet basis.
(2) For some threshold $T_n \in (0,\infty)$, depending on the noise level and the number of data points, apply the nonlinear soft-thresholding function

$$S(\cdot,T_n)\colon \mathbb{R} \to \mathbb{R}\colon y \mapsto S(y,T_n) := \begin{cases} y + T_n & \text{if } y \leqslant -T_n \\ y - T_n & \text{if } y \geqslant T_n \\ 0 & \text{otherwise} \end{cases} \tag{1.2}$$

to each wavelet coefficient of the data. The resulting thresholded coefficients $S(Y_n(j,k),T)$ are then considered as estimates for the wavelet coefficients of u_n. Notice, that the soft-thresholding function can be written in the compact form $S(y,T_n) = \operatorname{sign}(y)(|y| - T_n)_+$. Further, it sets all coefficients being in magnitude smaller than T_n to zero and shrinks the remaining coefficients towards zero by the value T_n.
(3) The desired estimate for the signal u_n is then defined by the wavelet series of the thresholded empirical coefficients $S(Y_n(j,k),T)$,

$$\hat{u}_n = \sum_{j=0}^{J} \sum_{k=0}^{2^j-1} S\bigl(Y_n(j,k),T_n\bigr)\psi^n_{j,k}. \tag{1.3}$$

Every step in the above procedure can be computed in $\mathcal{O}(n)$ operation counts and hence the overall procedure of wavelet soft-thresholding is linear in the number n of unknown parameters (see [8,12,14]). It is thus not only conceptually simple but also allows for fast numerical implementation. Even simple linear spectral denoising techniques using the FFT algorithm have a numerical complexity of $\mathcal{O}(n \log n)$ floating point operations. Besides these practical advantages, wavelet soft-thresholding also obeys certain theoretical optimality properties. It yields to an almost optimal mean square error simultaneously over a wide range of function spaces (including Sobolev and Besov spaces) and, at the same time, has a smoothing effect with respect to any of the norms in these spaces. Hence soft-thresholding automatically adapts to the unknown smoothness of the desired signal [7,15].

Any particular choice of the thresholding parameter T_n is a tradeoff between signal approximation and noise reduction: A large threshold removes much of the noise but also removes parts of the signal. Hence a reasonable threshold choice should be as small as possible under the side constrained that a significant amount of the noise is removed. The smaller the actual threshold is taken, the more emphasis is given on signal representation and the less emphasis on noise reduction. A commonly used threshold is the so called *universal threshold* $T_n = \sigma\sqrt{2\log n}$ as proposed in the seminal work [7], where the following result is shown.

Theorem 1.1 *(Denoising property of wavelet soft-thresholding). (See [7].) Suppose that $\mathcal{D}_n$ are consistent with an underlying orthonormal wavelet basis $\mathcal{D}$ on $[0,1]$ having m times continuously differentiable elements and m vanishing moments, that $u_n(k) = u(k/n)$, for $k = 0,\dots,n-1$, denote point samples of a function $u\colon[0,1] \to \mathbb{R}$ and that $\hat{u}_n$ are constructed by (1.3) with the universal threshold $T_n = \sigma\sqrt{2\log n}$. Then, there exists a special smooth interpolation of $\hat{u}_n$ producing a function $u^*_n\colon[0,1] \to \mathbb{R}$. Further, there are universal constants $(\pi_n)_n \subset (0,1)$ with $\pi_n \to 1$ as $n = 2^J \to \infty$, such that for any Besov space $\mathcal{B}^r_{p,q}$ which embeds continuously into $C[0,1]$ (hence $r > 1/p$) and for which $\mathcal{D}$ is an unconditional basis (hence $r < m$),*

$$\mathbf{P}\bigl\{\bigl\|u^*_n\bigr\|_{\mathcal{B}^r_{p,q}} \leqslant c\bigl(\mathcal{B}^r_{p,q},\mathcal{D}\bigr)\|u\|_{\mathcal{B}^r_{p,q}};\ \forall u \in \mathcal{B}^r_{p,q}\bigr\} \geqslant \pi_n, \tag{1.4}$$

for constants $c(\mathcal{B}^r_{p,q},\mathcal{D})$ depending on $\mathcal{B}^r_{p,q}$ and $\mathcal{D}$ but neither on u nor on n.

Theorem 1.1 states that the estimate u_n^* is, with probability tending to one, simultaneously as smooth as u for all smoothness spaces $\mathcal{B}_{p,q}^r$. This result can be derived from the denoising property (see [7,11,16] and also Section 3.2)

$$\mathbf{P}\Big\{\max_{(j,k)\in\Omega_n}\big|\langle\psi_{j,k}^n,\epsilon_n\rangle\big|\leqslant\sigma\sqrt{2\log n}\Big\}\geqslant\pi_n\to 1\quad\text{as } n\to\infty. \tag{1.5}$$

For an orthonormal basis, the noise coefficients $\langle\psi_{j,k}^n,\epsilon_n\rangle\sim N(0,\sigma^2)$ are independently distributed. Hence Eq. (1.5) is a consequence from standard extreme value results for independent normally distributed random vectors [17,18]. Extreme value theory also implies the limiting Gumbel law

$$\lim_{n\to\infty}\mathbf{P}\bigg\{\max_{\omega\in\Omega_n}\big|\langle\phi_\omega^n,\epsilon_n\rangle\big|\leqslant\sigma\sqrt{2\log n}+\sigma\frac{2z-\log\log n-\log\pi}{2\sqrt{2\log n}}\bigg\}=\exp\big(-e^{-z}\big), \tag{1.6}$$

uniformly in $z\in\mathbb{R}$. This even allows to exactly characterize all thresholding sequences T_n yielding a denoising property like (1.5) with T_n in place of $\sigma\sqrt{2\log n}$.

In the case that a redundant frame is considered instead of an orthonormal wavelet basis, then the empirical coefficients are no more linear independent and limiting result like (1.6) are much harder to come up with. In this paper we verify that a similar distributional result as in (1.6) holds for a wide class of redundant frames with n replaced by the number of frame elements. This class is shown to include non-orthogonal wavelets, curvelet frames and unions of bases (see Theorems 4.4, 4.7 and 4.12). Roughly speaking, the reason is, that the redundancy is of these frames weak enough that it asymptotically vanishes in a statistical sense and the system behaves as an independent system. However, we also an important example (in the form of the translational wavelet system; see Theorem 4.9) which shows that highly redundant systems may show a different asymptotic behaviour.

Our work is motivated by the well known observation that the universal threshold sigma $\sigma\sqrt{2\log n}$ often is found to be too large in applications, hence including too few coefficients into the final estimator (see [12,15,19]). This recently has initiated further research on refined thresholding methods and we would like to shed some light on this phenomenon for a large class of frame systems by providing a refined asymptotics as in (1.6) in addition to results of the type (1.5). We also provide a selective review on current thresholding methodology where we focus on the link between statistical extreme value theory and thresholding techniques.

1.2. Frame soft-thresholding: Main results

For any $n\in\mathbb{N}$, let $\mathcal{D}_n=(\phi_\omega^n\colon\ \omega\in\Omega_n)$ denote a frame of $\mathbb{R}^{I_n}$, where Ω_n is a finite index set, that consists of normalized frame elements (that is, $\|\phi_\omega^n\|=1$ holds for all $\omega\in\Omega_n$) and has frame bounds $a_n\leqslant b_n$ (compare Section 2.1). Our main results concerning thresholding estimation in the frame $\mathcal{D}_n$ will hold for asymptotically stable frames, which are defined as follows.

Definition 1.2 *(Asymptotically stable frames).* We say that a family of frames $(\mathcal{D}_n)_{n\in\mathbb{N}}$ with normalized frame elements is *asymptotically stable*, if the following assertions hold true:

(i) For some $\rho\in(0,1)$, we have $|\{(\omega,\omega')\in\Omega_n^2\colon\ |\langle\phi_\omega^n,\phi_{\omega'}^n\rangle|\geqslant\rho\}|=o(\frac{|\Omega_n|}{\sqrt{\log|\Omega_n|}})$ as $n\to\infty$.

(ii) The upper frame bounds b_n are uniformly bounded, that is, $B:=\sup\{b_n\colon\ n\in\mathbb{N}\}<\infty$.

The following Theorem 1.3 is the key to most results of this paper. It states, that after proper normalization the distribution of $\max_{\omega\in\Omega_n}|\langle\phi_\omega^n,\epsilon_n\rangle|$ converges to the Gumbel distribution – provided that the frames are asymptotically stable.

Theorem 1.3 *(Limiting distribution for asymptotically stable frames). Assume that $(\mathcal{D}_n)_{n\in\mathbb{N}}$ is an asymptotically stable family of frames, and let $(\epsilon_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^{I_n}$ with independent $N(0,\sigma^2)$-distributed entries. Then, for every $z\in\mathbb{R}$,*

$$\lim_{n\to\infty}\mathbf{P}\bigg\{\max_{\omega\in\Omega_n}\big|\langle\phi_\omega^n,\epsilon_n\rangle\big|\leqslant\sigma\sqrt{2\log|\Omega_n|}+\sigma\frac{2z-\log\log|\Omega_n|-\log\pi}{2\sqrt{2\log|\Omega_n|}}\bigg\}=\exp\big(-e^{-z}\big). \tag{1.7}$$

The function $z\mapsto\exp(-e^{-z})$ is known as the Gumbel distribution.

Proof. See Section 3.1. □

In the case that $\mathcal{D}_n$ are orthonormal bases, results similar to the one of Theorem 1.3 follow from standard extreme value results (see, for example, [17,18]) and are well known in the wavelet community (see, for example, [7,11,12]). However, neither the convergence of the maxima including absolute values (which is the actually relevant case) nor the use of redundant systems are covered by these results. In Section 4 we shall verify that many redundant frames, such as non-orthogonal wavelet frames, curvelets and unions of bases, are asymptotically stable and hence the limiting Gumbel law of Theorem 1.3

can be applied. Based on this limiting extreme value distribution we provide an exact characterisation of all thresholds T_n satisfying a denoising property similar to the one of (1.5) for general frame thresholding; see Section 3 for details.

Suppose that $(\alpha_n)_{n\in\mathbb{N}}$ is a sequence in $(0,1)$ converging to $\alpha \in [0,1)$, let z_n satisfy $\exp(-e^{-z_n}) = \alpha_n$ and let u_n denote the wavelet soft-thresholding estimate with the threshold

$$T_n = \sigma\sqrt{2\log|\Omega_n|} + \sigma\frac{2z_n - \log\log|\Omega_n| - \log\pi}{2\sqrt{2\log|\Omega_n|}}. \tag{1.8}$$

According to Theorem 1.3, the probability that u_n is contained in $R_n = \{\bar{u}_n\colon \max_{\omega\in\Omega_n}|\langle\phi_{j,k}, V_n - \bar{u}_n\rangle|\}$ tends to $1-\alpha$ as $n\to\infty$. Hence the sets R_n define asymptotically sharp confidence regions around the given data for any significance level α; see Section 3.3 for details.

The proof of Theorem 1.3 relies on new extreme value results for dependent chi-square distributed random variables (with one degree of freedom) which we establish in Appendix A. In the field of statistical extreme value theory, the following definition is common.

Definition 1.4 *(Gumbel type).* A sequence $(M_n)_{n\in\mathbb{N}}$ of real valued random variables is said to be of *Gumbel type* (or to be of extreme value type I), if there are real valued normalizing sequences $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$, such that the limit $\mathbf{P}\{M_n \leqslant a_n z + b_n\} \to \exp(-e^{-z})$ as $n\to\infty$ holds pointwise for all $z\in\mathbb{R}$ (and therefore uniformly).

Using the notion just introduced, Theorem 1.3 states that $\max_{\omega\in\Omega_n}|\langle\phi_\omega^n, \epsilon_n\rangle|$ is of Gumbel type, with normalizing sequences $\sigma a(\chi, |\Omega_n|)$ and $\sigma b(\chi, |\Omega_n|)$, where

$$a(\chi, |\Omega_n|) := \frac{1}{\sqrt{2\log|\Omega_n|}}, \tag{1.9}$$

$$b(\chi, |\Omega_n|) := \sqrt{2\log|\Omega_n|} - \frac{\log\log|\Omega_n| + \log\pi}{2\sqrt{2\log|\Omega_n|}}. \tag{1.10}$$

As shown in Theorem 3.3, the maxima of $\langle\phi_\omega^n, \epsilon_n\rangle$ without taking absolute values are also of Gumbel type. We emphasize, however, that the corresponding normalizing sequences differ from those required for the maxima with absolute values. Indeed, $\max_{\omega\in\Omega_n}|\langle\phi_\omega^n, \epsilon_n\rangle|$ behaves as the maximum of $2|\Omega_n|$ (opposed to $|\Omega_n|$) independent standard normally distributed random variables; compare with Remark A.6. The different fluctuation behaviour of the maxima with and without absolute values is not resembled by Eq. (1.5), which is exactly the same for the maxima with and without absolute values. Only in a refined distributional limit (1.7) this difference becomes visible. Moreover, in the case that the frames $\mathcal{D}_n$ are redundant, no result similar to Theorem 1.3 is known at all.

Asymptotical stability typically fails for frames without an underlying infinite dimensional frame. A prototype for such a family is the dyadic translation invariant wavelet transform (see Section 4.1.4). In this case, the redundancy of the translation invariant wavelet system increases boundlessly with increasing n, which implies that the corresponding upper frame bounds tend to infinity as $n\to\infty$. We indeed prove the following counterexample if condition (ii) in Definition 1.2 fails to hold.

Theorem 1.5 *(Tighter bound for translation invariant wavelet systems). Suppose that $(\psi_\omega^n)_{\omega\in\Omega_n}$ is a discrete translation invariant wavelet system with unit norm elements generated by a mother wavelet ψ that is continuously differentiable, and let $(\epsilon_n)_{n\in\mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^{I_n}$ with independent $N(0,\sigma^2)$-distributed entries. Then, for some constant $c>0$ and all $z\in\mathbb{R}$ we have*

$$\liminf_{n\to\infty}\mathbf{P}\left\{\max_{\omega\in\Omega_n}\left|\left\langle\psi_\omega^n, \epsilon_n\right\rangle\right| \leqslant \sqrt{2\log n} + \frac{z+\log(c/\pi)}{\sqrt{2\log n}}\right\} \geqslant \exp\left(-e^{-z}\right).$$

Proof. This follows from Theorem 4.9, that we proof in Appendix B.2. □

Theorem 1.5 shows that the maximum of the translation invariant wavelet coefficients is strictly smaller (in a distributional sense; see Section 4.1.4) than the maximum of an asymptotically stable frame with $|\Omega_n| = n\log n$ elements and therefore the result of Theorem 1.3 does not hold for a translation invariant wavelet system. Moreover, Theorem 1.5 shows that there exists a thresholding sequence being strictly smaller than $\sqrt{2\log|\Omega_n|}$ yields asymptotic smoothness; see Section 4.1.4 for details. This also reveals the necessity of a detailed extreme value analysis of the empirical noise coefficients in the case of redundant frames.

1.3. Outline

In the following Section 2 we introduce some notation used throughout this paper. In particular, we define the soft-thresholding estimator in redundant frames. The core part of this paper is Section 3, where we proof the asymptotic distribution of the frame coefficients claimed in Theorem 1.3. This result is then applied to define extreme value based thresholding rules and corresponding sharp confidence regions. Moreover, in this section we show that the resulting thresholding estimators satisfy both, oracle inequalities for the mean square error and smoothness estimates for a wide class of smoothness measures. Our proofs require new facts from statistical extreme value theory for the maxima of absolute values of dependent normal random variables that we derive in Appendix A. Finally, in Section 4 we discuss in detail several examples, including, non-orthogonal frames, biorthogonal wavelets, curvelets and unions of bases in detail.

2. Thresholding in redundant frames

For the following recall the model (1.1) and write $\epsilon_n = (\epsilon_n(k)\colon\ k \in I_n) \in \mathbb{R}^{I_n}$ for the noise vector in (1.1). We assume throughout that the variance σ^2 of the noise is given. Fast and efficient methods for estimating the variance are well known (see, for example, [20–23] for $d = 1$ and [24] for $d \geqslant 2$).

Throughout this paper all estimates for the signal u_n are based on thresholding the coefficients of the given data V_n with respect to prescribed frames of analyzing elements.

2.1. Frames

In the sequel $\mathcal{D}_n := (\phi^n_\omega\colon\ \omega \in \Omega_n) \subset \mathbb{R}^{I_n}$ denotes a frame of $\mathbb{R}^{I_n}$, with Ω_n being a finite index set. Hence there exist constants $0 < a_n \leqslant b_n < \infty$, such that

$$\left(\forall u_n \in \mathbb{R}^{I_n}\right) \quad a_n \|u_n\|^2 \leqslant \sum_{\omega \in \Omega_n} \left|\left\langle \phi^n_\omega, u_n \right\rangle\right|^2 \leqslant b_n \|u_n\|^2. \tag{2.1}$$

(Here $\|\cdot\|$ is the Euclidean norm on $\mathbb{R}^{I_n}$ and $\langle\cdot,\cdot\rangle$ the corresponding inner product.) The largest and smallest numbers a_n and b_n, respectively, that satisfy (2.1) are referred to as frame bounds. Notice that in a finite dimensional setting any family that spans the whole space is a frame.

We further denote by $\boldsymbol{\Phi}_n : \mathbb{R}^{I_n} \to \mathbb{R}^{\Omega_n}$ the operator that maps the signal $u_n \in \mathbb{R}^{I_n}$ to the analyzing coefficients with respect to the given frame,

$$(\forall \omega \in \Omega_n) \quad (\boldsymbol{\Phi}_n u_n)(\omega) := \left\langle \phi^n_\omega, u_n \right\rangle.$$

The mapping $\boldsymbol{\Phi}_n$ is named the *analysis operator*, its adjoint $\boldsymbol{\Phi}_n^*$ the *synthesis operator*, and $\boldsymbol{\Phi}_n^*\boldsymbol{\Phi}_n$ the *frame operator* corresponding to $\mathcal{D}_n$.

The frame property (2.1) implies that the frame operator $\boldsymbol{\Phi}_n^*\boldsymbol{\Phi}_n : \mathbb{R}^{I_n} \to \mathbb{R}^{I_n}$ is an invertible linear mapping. Hence, for any $\omega \in \Omega_n$, the elements

$$\tilde{\phi}^n_\omega := \left(\boldsymbol{\Phi}_n^*\boldsymbol{\Phi}_n\right)^{-1} \phi^n_\omega$$

are well defined and the family $(\tilde{\phi}^n_\omega\colon\ \omega \in \Omega_n)$ is again a frame of $\mathbb{R}^{I_n}$. It is called the *dual frame* and has frame bounds $1/b_n \leqslant 1/a_n$.

Finally, we denote by $\boldsymbol{\Phi}_n^+ := (\boldsymbol{\Phi}_n^*\boldsymbol{\Phi}_n)^{-1}\boldsymbol{\Phi}_n^*$ the pseudoinverse of the analysis operator $\boldsymbol{\Phi}_n$. Due to linearity and the definitions of the pseudoinverse and the dual frame elements, we have the identities

$$\left(\forall u_n \in \mathbb{R}^{I_n}\right) \quad u_n = \boldsymbol{\Phi}_n^+ \boldsymbol{\Phi}_n u_n = \sum_{\omega \in \Omega_n} \left\langle \phi^n_\omega, u_n \right\rangle \tilde{\phi}^n_\omega. \tag{2.2}$$

In particular, the mapping $\boldsymbol{\Phi}_n^+$ is the synthesis operator corresponding to the dual frame. Eq. (2.2) provides a simple representation of the given signal in terms of its analyzing coefficients. This serves as basis of thresholding estimators defined and studied in the following subsection. For further details on frames see, for example, [12,25].

Remark 2.1 *(Thresholding in a subspace).* It is not essential at all, that $\mathcal{D}_n$ is a frame of the whole image space $\mathbb{R}^{I_n}$. In fact, in typical thresholding applications, such as in wavelet denoising, the space $\mathbb{R}^{I_n}$ naturally decomposes into a low resolution space having small fixed dimension and a detail space having large dimension that increases with n. The soft-thresholding procedure is then only applied to the signal part in the detail space and hence it is sufficient to assume that $\mathcal{D}_n$ is a frame therein. In order to avoid unessential technical complication we present our results for the case of frames of the whole image space. In the concrete applications presented in Section 4 the thresholding will indeed only be performed in some subspace; all results carry over to such a situation in a straightforward manner.

2.2. Thresholding estimation

By applying $\boldsymbol{\Phi}_n$ to both sides of (1.1), the original denoising problem in the signal space $\mathbb{R}^{I_n}$ is transferred into the denoising problem

$$Y_n(\omega) = x_n(\omega) + (\boldsymbol{\Phi}_n \epsilon_n)(\omega), \quad \text{for } \omega \in \Omega_n, \tag{2.3}$$

in the possibly higher dimensional coefficient space $\mathbb{R}^{\Omega_n}$. Here and in the following we denote by

$$Y_n(\omega) := \langle \phi_\omega^n, V_n \rangle \quad \text{and} \quad x_n(\omega) := \langle \phi_\omega^n, u_n \rangle, \tag{2.4}$$

the coefficients of the data V_n and the signal u_n with respect to the given frame. The following elementary Lemma 2.2 states that the noise term in (2.3) is again a centred normal vector but has possibly non-vanishing covariances. Indeed it implies that the entries of $\boldsymbol{\Phi}_n \epsilon_n$ are not uncorrelated and hence not independent, unless $\mathcal{D}_n$ is an orthogonal basis.

Lemma 2.2 *(Covariance matrix). Let ϵ_n be a random vector in the image space $\mathbb{R}^{I_n}$ with independent $N(0, \sigma^2)$-distributed entries. Then $\boldsymbol{\Phi}_n \epsilon_n$ is a centred normal vector in $\mathbb{R}^{\Omega_n}$ and the covariance matrix of $\boldsymbol{\Phi}_n \epsilon_n$ has entries $\mathbf{Cov}(\boldsymbol{\Phi}_n \epsilon_n(\omega), \boldsymbol{\Phi}_n \epsilon_n(\omega')) = \sigma^2 \langle \phi_\omega^n, \phi_{\omega'}^n \rangle$.*

Proof. As the sum of normal random variables with zero mean, the random variables $(\boldsymbol{\Phi}_n \epsilon_n)(\omega) = \sum_{k \in I_n} \phi_\omega^n(k) \epsilon_n(k)$ are again normally distributed with zero mean. In particular, we have $\mathbf{Cov}(\boldsymbol{\Phi}_n \epsilon_n(\omega), \boldsymbol{\Phi}_n \epsilon_n(\omega')) = \mathbf{E}(\boldsymbol{\Phi}_n \epsilon_n(\omega) \boldsymbol{\Phi}_n \epsilon_n(\omega'))$. Hence the claim follows from the linearity of the expectation value and the independence of $\epsilon_n(k)$. □

Recall the soft-thresholding function $S(y, T_n) = \operatorname{sign}(y)(|y| - T_n)_+$ defined by Eq. (1.2). The thresholding estimators we consider apply $S(\cdot, T_n)$ to each coefficient of Y_n in (2.3) to define an estimator for the parameter x_n. In order to get an estimate for the signal u_n one must map the coefficient estimate back to the original signal domain. This is usually implemented by applying the dual synthesis operator (compare with Eq. (2.2)).

Definition 2.3 *(Frame thresholding).* Consider the data models (1.1) and (2.3) and let $T_n > 0$ be a given thresholding parameter.

(a) The soft-thresholding estimator for $x_n \in \mathbb{R}^{\Omega_n}$ using the threshold T_n is defined by

$$\hat{x}_n = \mathbf{S}(Y_n, T_n) := \big(S\big(Y_n(\omega), T_n\big) \colon \omega \in \Omega_n\big) \in \mathbb{R}^{\Omega_n}. \tag{2.5}$$

(b) The soft-thresholding estimator for u_n with respect to the frame $\mathcal{D}_n$ using the threshold T_n is defined by

$$\hat{u}_n = \boldsymbol{\Phi}_n^+ \circ \mathbf{S}(\boldsymbol{\Phi}_n V_n, T_n) = \sum_{\omega \in \Omega_n} S\big(\langle \phi_\omega^n, V_n \rangle, T_n\big) \tilde{\phi}_\omega^n. \tag{2.6}$$

Hence the frame soft-thresholding estimator $\hat{u}_n$ is simply the composition of analysis with $\boldsymbol{\Phi}_n$, component-wise thresholding, and dual synthesis with $\boldsymbol{\Phi}_n^+$.

If $\mathcal{D}_n$ is an overcomplete frame, then $\boldsymbol{\Phi}_n$ has infinitely many left inverses, and the pseudoinverse used in Definition 2.3 is a particular one. In principle one could use other left inverses for defining the soft-thresholding estimator (2.6). Since, in general, $\mathbf{S}(Y_n, T_n) \notin \operatorname{Ran}(\boldsymbol{\Phi}_n)$ is outside the range of $\boldsymbol{\Phi}_n$, the use of a different left inverse will result in a different estimator. The special choice $\boldsymbol{\Phi}_n^+$ has the advantage that for many frames used in practical applications, the dual synthesis operator is known explicitly and, more importantly, that fast algorithms are available for its computation (typically algorithms using only $\mathcal{O}(|I_n| \log |I_n|)$ or even $\mathcal{O}(|I_n|)$ floating point operations [12]).

Remark 2.4 *(Thresholding variations).* Instead of the *soft* thresholding function $S(\cdot, T_n)$ several other nonlinear thresholding methods have been proposed and used. Prominent examples are the hard thresholding function $z \mapsto z\chi_{\{|z| \geqslant T_n\}}$ and the nonnegative garrote $z \mapsto z \max\{1 - T_n^2/z^2, 0\}$ of [26,27]. Strictly taken, the smoothness estimates derived in Section 3.5 only hold for thresholding functions $F(\cdot, T_n)$ satisfying the shrinkage property $|F(y \pm T_n, T_n)| \leqslant |y|$ for all $y \in \mathbb{R}$. This property is, for example, not satisfied by the nonnegative garrote. In this case, however, similar estimates may be derived under additional assumptions on the signals of interest. Other prominent denoising techniques are based on block-thresholding (see, for example, [1,2,28,29]). In this case, the derivation of sharp smoothness estimates requires extreme value results for dependent χ^2-distributed random variables (with more than one degree of freedom). Such an extreme value analysis seems possible but is beyond the scope of this paper. Our given results can be seen as the first step towards a more general theory.

Remark 2.5 *(Multiple selections).* Be aware, that we allow certain elements ϕ_ω^n to be contained more than once in the frame $\mathcal{D}_n$. Hence we may have $|\{\phi_\omega^n \colon \omega \in \Omega_n\}| < |\Omega_n|$. Such multiple selections often arises naturally for frames that are the union of several bases having some elements in common. A standard example is the wavelet cycle spinning procedure of [6], where the underlying frame is the union of several shifted orthonormal wavelet bases (see Section 4.1.3). Multiple selections of frame elements also affect the pseudoinverse and finally the soft-thresholding estimator. Hence, if $(\phi_\omega^n \colon \omega \in \Omega_n)$ and $(\psi_\lambda^n \colon \lambda \in \Lambda_n)$ denote two frames composed by the same frame elements, $\{\phi_\omega^n \colon \omega \in \Omega_n\} = \{\psi_\lambda^n \colon \lambda \in \Lambda_n\}$, but having different cardinalities $|\Omega_n| \neq |\Lambda_n|$, then the soft-thresholding estimators corresponding to these frames differ from each other.

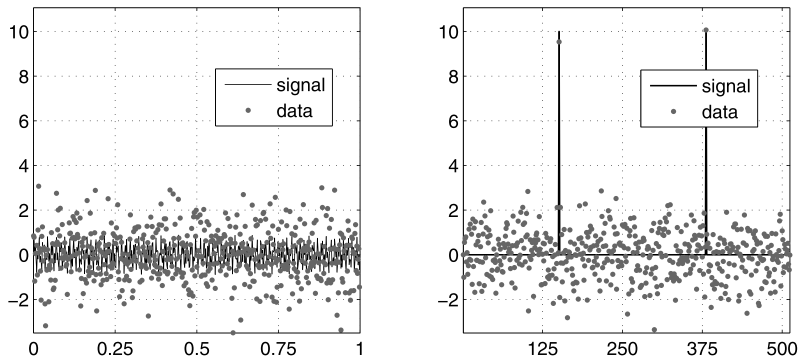

Fig. 1. Left: Signal u_n (superposition of two sine waves) and data $V_n = u_n + \epsilon_n$ from Example 2.6. Right: Coefficients of the signal and the data with respect to the sine basis.

2.3. Rationale behind thresholding estimation

We conclude this section by commenting on the underlying rationale behind thresholding estimation and situations where it is expected to produce good results.

The basic assumption underlying thresholding estimation is that the frame operator separates the data into large coefficients due to the signal and small coefficients mainly due to the noise. For additive noise models $V_n = u_n + \epsilon_n$ both issues can be studied separately. In this case, one requires that for some threshold T_n (which finally judges between signal and noise) the following two conditions are satisfied:

(1) *Coherence between signal and frame:* The signal u_n is well represented by few large coefficients having $|\langle \phi^n_\omega, u_n\rangle| > T_n$.
(2) *Incoherence between noise and frame:* With high probability, all noise coefficients with respect to the frame $\mathcal{D}_n$ satisfy $|\langle \phi^n_\omega, \epsilon_n\rangle| \leqslant T_n$.

In the following sections we shall see that item (2) can be analyzed in a unified way for asymptotically stable frames. Item (1), however, is more an approximation issue rather than an estimation issue. Given a frame, it, of course, cannot be satisfied for every $u_n \in \mathbb{R}^n$. The choice of a 'good frame' depends on the specific application at hand and in particular on the type of signals that are expected. The better the signals of interest are represented by a few but large frame coefficients, the better the denoising result will be. The richer the analyzing family is, the more signals can be expected to be recovered properly. The price to pay must be, of course, a higher computational cost.

The following two simple examples demonstrate how the use of redundant frames may significantly improve the performance of the thresholding estimation.

Example 2.6 *(Thresholding in the sine basis).* We consider the discrete signal $u_n \in \mathbb{R}^n$ defined by $u_n(k) = 5\sqrt{2}/16 \sin(\pi\omega_1 k/n) + 5\sqrt{2}/16 \sin(\pi\omega_2 k/n)$, which is a superposition of two sine waves having frequencies $\omega_1 = 150$ and $\omega_2 = 380$, respectively, and amplitudes $5\sqrt{2}/16 \simeq 0.45$. The left image in Fig. 1 shows the signal u_n and the noisy data $V_n = u_n + \epsilon_n$ obtained by adding Gaussian white noise of variance equal to one to the signal. Apparently, there seems little hope to recover u_n from the data V_n in the original signal domain. Almost like a miracle, the situation changes drastically after computing the coefficients with respect to the sine basis $(n^{-1/2}\sin(\pi\omega k/n)\colon\ \omega = 1, \ldots, n)$. Now, the signal and the noise are clearly separated as can be seen from the right image in Fig. 1. Obviously we will get an almost perfect reconstruction by simply removing all coefficients below a proper threshold.

Example 2.7 *(Thresholding in a redundant sine frame).* The signal in Example 2.6 is a combination of sine waves with integer frequencies covered by the sine frame. However, in practical application the signal may also have non-integer frequencies. In order to investigate this issue, we now consider the signal $u'_n(k) = 5\sqrt{2}/16 \sin(\pi\omega'_1 k/n) + 5\sqrt{2}/16 \sin(\pi\omega_2 k/n)$ having frequencies $\omega'_1 = 150.5$ and $\omega_2 = 380$ (hence ω'_1 is a slight perturbation of the frequency ω_1 considered in Example 2.6). The new signal u'_n is not a sparse linear combination of elements of the sine basis. As a matter of fact, the energy of the first sine wave is spread over many coefficients and thus submerges in the noise. Indeed, as can be seen from the left image in Fig. 2, the low frequency coefficient disappears. However, by taking the two times redundant frame $(n^{-1/2}\sin(\pi\omega k/n)\colon\ \omega = \{1/2, 1, \ldots, n\})$ instead of the sine basis, the coefficient due to frequency ω'_1 appears again in the transformed domain. Moreover, as can be seen from Fig. 3 the reconstruction by thresholding the coefficients with respect to the overcomplete sine frame is almost perfect, whereas the reconstruction by thresholding the basis coefficients is useless.

In Examples 2.6 and 2.7 the threshold choice is not a very delicate issue since the signal and the noise are separated very clearly in the transformed domain. Indeed as can be seen from the right plots in Figs. 1 and 2 there is a quite wide

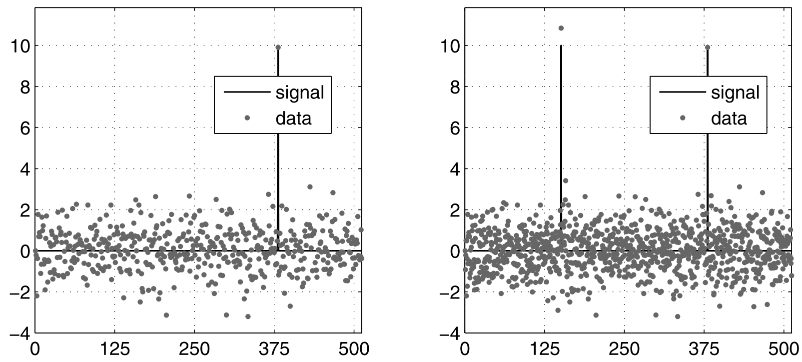

Fig. 2. Left: Coefficients of the signal u_n' and the data $V_n' = u_n' + \epsilon_n$ from Example 2.7 with respect the sine basis. Right: Coefficients of the same signal and data with respect to the two times oversampled sine frame.

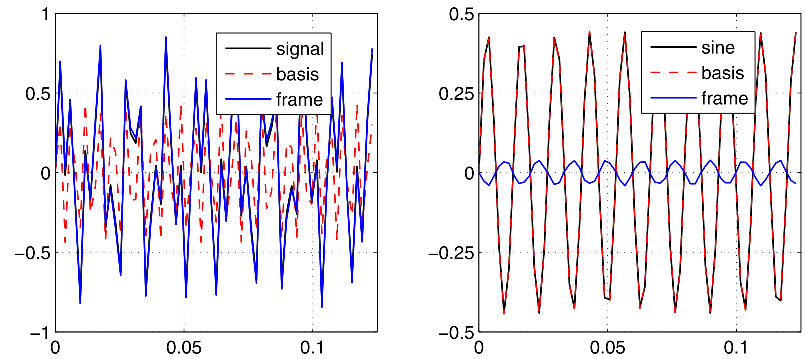

Fig. 3. Left: Signal u_n', and the reconstructions from the data $V_n' = u_n' + \epsilon_n$ by soft-thresholding in the sine basis and in the overcomplete sine frame, respectively. Only the first 64 components are plotted. Right: Sine wave $5\sqrt{2}/16 \sin(\pi \omega_1' k/n)$ and residuals of the two reconstructions. As can be seen, thresholding in the sine frame almost perfectly recovers the signal u_n, whereas the result of thresholding in the sine basis is useless (the residual is almost equal to the displayed sine wave of frequency ω_1').

range of thresholds that would yield an almost noise free estimate close to the original signal. However, if the signal also contains important coefficients of moderate size, then the choice of a good threshold is crucial and difficult. This is typically the case for image denoising using wavelets or curvelet frames: Natural images are approximately sparse in these frames but almost never strictly sparse. The particular threshold choice now will always be a tradeoff between noise removal and signal representation and becomes a delicate issue. In order to develop rationale threshold choices, a precise understanding of the distribution of $|\langle \phi_\omega^n, \epsilon_n \rangle|$ is helpful. This is the subject of our following considerations.

3. Extreme value analysis of frame thresholding

Now we turn back to the denoising problem (1.1). After application of the analysis operator $\boldsymbol{\Phi}_n$ corresponding to the normalized frame $\mathcal{D} = (\phi_\omega^n\colon\ \omega \in \Omega_n)$ our aim is to estimate the vector $x_n \in \mathbb{R}^{\Omega_n}$ from given noisy coefficients (compare with Eq. (2.3))

$$Y_n(\omega) = x_n(\omega) + (\boldsymbol{\Phi}_n \epsilon_n)(\omega), \quad \text{for } \omega \in \Omega_n.$$

Here $\boldsymbol{\Phi}_n \epsilon_n$ is the transformed noise vector which is normally distributed, has zero mean and covariance matrix $\kappa_n(\omega, \omega') = \sigma^2 \langle \phi_\omega^n, \phi_{\omega'}^n \rangle$; see Lemma 2.2. In this section we shall analyze in detail the component-wise soft-thresholding estimator $\hat{x}_n = \mathbf{S}(Y_n, T_n)$ defined by (2.5). We will start by computing the extreme value distribution of $\boldsymbol{\Phi}_n \epsilon_n$ claimed in Theorem 1.3. Based on the limiting law will then introduce extreme value thresholding techniques that will be shown to provide asymptotically sharp confidence regions.

3.1. Proof of Theorem 1.3

The main aim of this subsection is to verify Theorem 1.3, which states that the distribution of the maxima of the noise coefficients $\boldsymbol{\Phi}_n \epsilon_n(\omega)$ are of Gumbel type with explicitly given normalization constants. The proof of Theorem 1.3 will be a consequence of Lemmas 3.1 and 3.2 to be derived in the following. The main Lemma 3.1 relies itself on a new extreme value result that we establish in Appendix A.

Lemma 3.1. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of normal random vectors in $\mathbb{R}^{\Omega_n}$ with covariance matrices κ_n having ones in the diagonal. Assume additionally, that the following hold:*

(i) *For every $\delta \in (0,1)$, $|\{(\omega,\omega') \in \Omega_n^2\colon |\kappa_n(\omega,\omega')| \geqslant \delta\}| = \mathcal{O}(|\Omega_n|)$ as $n \to \infty$.*
(ii) *For some $\rho \in (0,1)$, $|\{(\omega,\omega') \in \Omega_n^2\colon |\kappa_n(\omega,\omega')| \geqslant \rho\}| = o(|\Omega_n|/\sqrt{\log|\Omega_n|})$ as $n \to \infty$.*
(iii) $B := \sup\{\sum_{\omega'\in\Omega_n} |\kappa_n(\omega,\omega')|^2\colon n \in \mathbb{N}$ *and* $\omega \in \Omega_n\} < \infty$.

Then, $\|\xi_n\|_\infty$ is of Gumbel type (see Definition 1.4) with normalization constants $a(\chi, |\Omega_n|)$ and $b(\chi, |\Omega_n|)$ defined by (1.9) and (1.10).

Proof. Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of normal random vectors satisfying conditions (i)–(iii). According to Theorem A.8 it is sufficient to show that

$$R_n := \sum_{\omega\neq\omega'} \bigl|\kappa_n(\omega,\omega')\bigr| \left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{1/(1+|\kappa_n(\omega,\omega')|)} \to 0 \quad \text{as } n \to \infty.$$

This will be done by splitting the sum R_n into three parts and showing that each of them tends to zero as $n \to \infty$. For that purpose, let $\delta \in (0, 1/3)$ be any small number, let $\rho \in (0,1)$ be as in condition (ii) and define

$$\begin{aligned}
\Lambda_n(1) &:= \bigl\{(\omega,\omega') \in \Omega_n^2\colon\ \omega \neq \omega' \text{ and } \bigl|\kappa_n(\omega,\omega')\bigr| \geqslant \rho\bigr\},\\
\Lambda_n(2) &:= \bigl\{(\omega,\omega') \in \Omega_n^2\colon\ \delta \leqslant \bigl|\kappa_n(\omega,\omega')\bigr| < \rho\bigr\},\\
\Lambda_n(3) &:= \bigl\{(\omega,\omega') \in \Omega_n^2\colon\ \bigl|\kappa_n(\omega,\omega')\bigr| < \delta\bigr\}.
\end{aligned}$$

We further write $R_n = R_n(1) + R_n(2) + R_n(3)$ with

$$R_n(i) := \sum_{(\omega,\omega')\in\Lambda_n(i)} \bigl|\kappa_n(\omega,\omega')\bigr| \left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{1/(1+|\kappa_n(\omega,\omega')|)} \qquad \text{for } i \in \{1,2,3\}.$$

It remains to verify that any of the terms $R_n(i)$ converges to zero as $n \to \infty$.

- Since any ξ_n is a normal random vector with zero mean and unit variance, we have $|\kappa_n(\omega,\omega')| \leqslant 1$ for any index pair $(\omega,\omega') \in \Omega_n^2$, which yields the inequality $R_n(1) \leqslant |\Lambda_n(1)|\sqrt{\log|\Omega_n|}/|\Omega_n|$. By condition (ii) we have $|\Lambda_n(1)| = o(|\Omega_n|/\sqrt{\log|\Omega_n|})$ which shows that $R_n(1) \to 0$ as $n \to \infty$.
- To estimate the second sum $R_n(2)$, recall that by definition of the set $\Lambda_n(2)$, we have $|\kappa_n(\omega,\omega')| \leqslant \rho$ for any pair of indices $(\omega,\omega') \in \Lambda_n(2)$. Moreover, recall that by condition (i) we further have $|\Lambda_n(2)| = \mathcal{O}(|\Omega_n|)$. Hence we obtain

$$R_n(2) \leqslant \bigl|\Lambda_n(2)\bigr| \left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{1/(1+\rho)} = \bigl(\log|\Omega_n|\bigr)^{1/(1+\rho)} \mathcal{O}\bigl(|\Omega_n|^{1-2/(1+\rho)}\bigr).$$

  Since by assumption $\rho < 1$, the inequality $1 - 2/(1+\rho) < 0$ holds which implies that we have $R_n(2) \to 0$ as $n \to \infty$.
- It remains to estimate the final sum $R_n(3)$. The Cauchy–Schwarz inequality, condition (iii), and the estimate $|\kappa_n(\omega,\omega')| \leqslant \delta$ yield

$$\begin{aligned}
R_n(3)^2 &\leqslant \sum_{(\omega,\omega')\in\Lambda_n(3)} \bigl|\kappa_n(\omega,\omega')\bigr|^2 \sum_{(\omega,\omega')\in\Lambda_n(3)} \left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{2/(1+\delta)}\\
&\leqslant B|\Omega_n| \left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{2/(1+\delta)} |\Omega_n|^2 = \bigl(\log|\Omega_n|\bigr)^{2/(1+\delta)} \mathcal{O}\bigl(|\Omega_n|^{3-4/(1+\delta)}\bigr).
\end{aligned}$$

  Now, by assumption the inequality $\delta < 1/3$ holds and hence we have $4/(1+\delta) > 3$. This implies that also $R_n(3)$ tends to zero as $n \to \infty$.

In summary, we have verified that $R_n(i) \to 0$ as $n \to \infty$ for every $i \in \{1,2,3\}$. Hence their sum R_n converges to zero, too. The claimed distributional convergence results now follows from Theorem A.8 and concludes the proof. □

We next state a simple auxiliary lemma that bounds the number of inner products $\langle\phi_\omega^n, \phi_{\omega'}^n\rangle$ being bounded away from zero.

Lemma 3.2. *For any n let $(\phi_\omega^n\colon \omega \in \Omega_n)$ be a family of normalized vectors in $\mathbb{R}^{I_n}$, such that the upper frame bounds b_n are uniformly bounded. Then, for every $\delta > 0$, we have*

$$\bigl|\bigl\{(\omega,\omega') \in \Omega_n^2\colon \bigl|\langle\phi_\omega^n, \phi_{\omega'}^n\rangle\bigr| \geqslant \delta\bigr\}\bigr| = \mathcal{O}\bigl(|\Omega_n|\bigr). \tag{3.1}$$

Proof. To verify (3.1) it is sufficient to find, for every given $\delta > 0$, some constant $K \in \mathbb{N}$ such that

$$(\forall n \in \mathbb{N})(\forall \omega \in \Omega_n) \quad \left|\left\{\omega' \in \Omega_n\colon \left|\langle \phi_\omega^n, \phi_{\omega'}^n \rangle\right| \geqslant \delta\right\}\right| \leqslant K. \tag{3.2}$$

Indeed, if (3.2) holds then summing over all $\omega \in \Omega_n$ yields (3.1).

To show (3.2) we assume to the contrary that there is some $\delta > 0$ such that for all $m \in \mathbb{N}$ there exists some $n(m) \in \mathbb{N}$ and some $\omega \in \Omega_{n(m)}$ such that the set $\Lambda_m = \{\omega' \in \Omega_{n(m)}\colon |\langle \phi_\omega^{n(m)}, \phi_{\omega'}^{n(m)} \rangle| \geqslant \delta\}$ contains more then m elements. By assumption we have the equality $\|\phi_\omega^{n(m)}\| = 1$ for all $\omega \in \Omega_n$. Together with assumption (ii) this implies

$$B = B\left\|\phi_\omega^{n(m)}\right\|^2 \geqslant \sum_{\omega' \in \Omega_{n(m)}} \left|\langle \phi_\omega^{n(m)}, \phi_{\omega'}^{n(m)} \rangle\right|^2 \geqslant \sum_{\omega' \in \Lambda_M} \left|\langle \phi_\omega^{n(m)}, \phi_{\omega'}^{n(m)} \rangle\right|^2 \geqslant m\delta.$$

Since the last estimate should hold for all $m \in \mathbb{N}$ and we have $B < \infty$ by assumption, this obviously gives is a contradiction. □

Proof of Theorem 1.3. Theorem 1.3 is now an immediate consequence of the above results: Lemma 2.2 and Lemma 3.2 show that the sequence of normalized frame coefficients $(\boldsymbol{\Phi}_n \epsilon_n / \sigma)_{n \in \mathbb{N}}$ satisfies conditions (i)–(iii) of Lemma 3.1. Hence Lemma 3.1 applied to the random vectors $\xi_n = \boldsymbol{\Phi}_n \epsilon_n / \sigma$ shows the assertion. □

We conclude this subsection by stating an extreme value result for the maximum without the absolute values. Although we do not need this result further, the distributional limit is interesting in its own.

Theorem 3.3 *(Limiting Gumbel law without absolute values). Assume that the frames $\mathcal{D}_n$ are asymptotically stable and let $(\epsilon_n)_{n \in \mathbb{N}}$ be a sequence of random vectors in $\mathbb{R}^{I_n}$ having independent $N(0,1)$-distributed entries. Then, the random sequence of the maxima* $\max(\boldsymbol{\Phi}_n \epsilon_n) := \max\{\langle \phi_\omega^n, \epsilon_n \rangle\colon \omega \in \Omega_n\}$ *is of Gumbel type with normalization constants*

$$a\big(N, |\Omega_n|\big) := \frac{1}{\sqrt{2\log|\Omega_n|}}, \tag{3.3}$$

$$b\big(N, |\Omega_n|\big) := \sqrt{2\log|\Omega_n|} - \frac{\log\log|\Omega_n| + \log(4\pi)}{2\sqrt{2\log|\Omega_n|}}. \tag{3.4}$$

Proof. The proof is analogous to the proof of Theorem 1.3 and uses the extreme value result of Theorem A.4 for dependent normal random vectors instead of the one of Theorem A.8 for absolute values of dependent normal random vectors. □

3.2. Universal threshold: Qualitative denoising property

In the case that the family $\mathcal{D}_n = (\phi_\omega^n\colon \omega \in \Omega_n)$ is an orthonormal basis it is well known that the thresholding sequence $T_n = \sigma\sqrt{2\log|\Omega_n|}$ satisfies the asymptotic denoising property (see, for example, [7,11,12] and also Section 1.1 in the introduction), that is,

$$\lim_{n\to\infty} \mathbf{P}\big\{\|\boldsymbol{\Phi}_n \epsilon_n\|_\infty \leqslant T_n\big\} = 1. \tag{3.5}$$

Eq. (3.5) implies that the estimates obtained with the threshold $\sigma\sqrt{2\log|\Omega_n|}$ are, with probabilities tending to one, at least as smooth as u_n. Hence the relation (3.5) is often used as theoretical justification for using the universal threshold choice $\sigma\sqrt{2\log|\Omega_n|}$ originally proposed by Donoho and Johnstone (see [7,8]). The following Proposition 3.4 states that the same denoising property indeed holds true for any normalized frame. Actually it proves much more: First, we verify (3.5) for a wide class of thresholds including the Donoho–Johnstone threshold. Second, we show that this class in fact includes all thresholds that satisfy the denoising property (3.5) – provided that the frames are asymptotically stable. Our results can be seen as a generalization and a refinement of [7, Theorem 4.1] from the basis case to the possibly redundant frame case.

Proposition 3.4 *(Thresholds yielding the denoising property). Assume that $\mathcal{D}_n$ are frames of $\mathbb{R}^{I_n}$ having normalized frame elements and analysis operators $\boldsymbol{\Phi}_n$, and let $(\epsilon_n)_{n\in\mathbb{N}}$ be a sequence of noise vectors in $\mathbb{R}^{I_n}$ with independent $N(0,\sigma^2)$-distributed entries.*

(a) *If $(\mathcal{D}_n)_{n\in\mathbb{N}}$ is asymptotically stable, then a sequence $(T_n)_{n\in\mathbb{N}}$ of thresholds satisfies (3.5) if and only if it has the form*

$$T_n := \sigma\sqrt{2\log|\Omega_n|} + \sigma\frac{2z_n - \log\log|\Omega_n| - \log\pi}{2\sqrt{2\log|\Omega_n|}} \quad \text{with } \lim_{n\to\infty} z_n = \infty. \tag{3.6}$$

(b) *If $(\mathcal{D}_n)_{n\in\mathbb{N}}$ is not necessarily asymptotically stable, then still any sequence $(T_n)_{n\in\mathbb{N}} \subset (0,\infty)$ of the form (3.6) satisfies the asymptotic denoising property (3.5).*

Proof. (a) Theorem 1.3 immediately implies that a sequence $(T_n)_{n\in\mathbb{N}}$ satisfies (3.5) if and only if it has the form

$$T_n = \sigma\sqrt{2\log|\Omega_n|} + \sigma\frac{2z_n - \log\log|\Omega_n| - \log\pi}{2\sqrt{2\log|\Omega_n|}}$$

for some sequence $(z_n)_{n\in\mathbb{N}}$ with $z_n \to \infty$.

(b) Now let $\mathcal{D}_n$ be any sequence of frames that is not necessarily asymptotically stable. Further, let η_n be a sequence of random vectors with independent $N(0,\sigma^2)$-distributed entries. Since $\boldsymbol{\Phi}_n\epsilon_n$ is a random vector with possibly dependent $N(0,\sigma^2)$-distributed entries, Lemma A.9 implies that

$$\mathbf{P}\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant T_n\} \geqslant \mathbf{P}\{\|\eta_n\|_\infty \leqslant T_n\}.$$

By item (a) we already know that $\mathbf{P}\{\|\eta_n\|_\infty \leqslant T_n\} \to 1$ as $n\to\infty$, for any sequence of thresholds satisfying (3.6), and hence the same must hold true for $\mathbf{P}\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant T_n\}$. □

According to Proposition 3.4, any sequence $(z_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} z_n = \infty$ defines a sequence of thresholds (3.6) that satisfies the asymptotic denoising property. In particular, by taking $2z_n = \log\log|\Omega_n| + \log\pi$ the thresholds in (3.6) reduce to the universal threshold $\sigma\sqrt{2\log|\Omega_n|}$ of Donoho and Johnstone. Proposition 3.4 further shows that the asymptotic relation $T_n \sim \sigma\sqrt{2\log|\Omega_n|}$ alone is not sufficient for the denoising property (3.5) to hold and that second order approximations have to be considered. One may call a thresholding sequence $(T_n)_n$ smaller than $(T'_n)_n$, if $(T'_n - T_n)T_n \to \infty$ for $n\to\infty$. The smaller the thresholding sequence is taken, the slower the convergence of $\mathbf{P}\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant T_n\}$ will be, and hence this just yields a different compromise between noise reduction and signal approximation.

3.3. Extreme value threshold: Sharp confidence regions

For the following notice that the soft-thresholding estimate $\hat{x}_n = \mathbf{S}(Y_n, T_n)$ with thresholding parameter T_n is an element of the $\|\cdot\|_\infty$-ball

$$R(Y_n, T_n) := \{\bar{x}_n \in \mathbb{R}^{\Omega_n}\colon \|\bar{x}_n - Y_n\|_\infty \leqslant T_n\} \tag{3.7}$$

around the given data Y_n. Our aim is to select the thresholding value T_n in such a way, that $R(Y_n, T_n)$ is an asymptotically sharp confidence region corresponding to some prescribed significance level α, in the sense that the probability that we have $x_n \in R(Y_n, T_n)$ tends to $1-\alpha$ as $n\to\infty$. By definition, $x_n \in R(Y_n, T_n)$ if and only if $\|x_n - Y_n\|_\infty \leqslant T_n$. The data model $Y_n = x_n + \boldsymbol{\Phi}_n\epsilon_n$ thus implies that

$$\mathbf{P}\{x_n \in R(Y_n, T_n);\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\} = \mathbf{P}\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant T_n\}. \tag{3.8}$$

Here and in similar situations, $\mathbf{P}\{x_n \in R(Y_n, T_n);\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\}$ denotes the probability of the intersection of all the events $\{x_n \in R(Y_n, T_n)\}$ taken over all $x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)$.

Now assume that the frames are asymptotically stable. Then Theorem 1.3 states that the probabilities in Eq. (3.8) with $T_n = \sigma a(\chi, |\Omega_n|)z + \sigma b(\chi, |\Omega_n|)$ tend to the Gumbel distribution $\exp(-\exp(-z))$. This suggests the following threshold choice based on the quantiles of the limiting Gumbel distribution.

Definition 3.5 *(Extreme value threshold).* Let $(\alpha_n)_{n\in\mathbb{N}} \in (0,1)$ be any sequence of significance levels, denote by $z(\alpha_n) = -\log\log(1/(1-\alpha_n))$ the α_n-quantile of the Gumbel distribution, and set

$$T(\alpha_n, |\Omega_n|) := \sigma\sqrt{2\log|\Omega_n|} + \sigma\frac{2z(\alpha_n) - \log\log|\Omega_n| - \log\pi}{2\sqrt{2\log|\Omega_n|}}. \tag{3.9}$$

We then name $T(\alpha_n, |\Omega_n|)$ the sequence of extreme value threshold (EVT) corresponding to the significance levels α_n.

The following Theorem 3.6 states that the EVTs defined by Eq. (3.9) indeed define asymptotically sharp confidence regions. Actually it is mere a corollary of the extreme value result derived in Theorem 1.3.

Theorem 3.6 *(Asymptotically sharp confidence regions). Let $(\mathcal{D}_n)_{n\in\mathbb{N}}$ be an asymptotically stable family of frames in $\mathbb{R}^{I_n}$ and let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence of numbers in $(0,1)$ converging to some $\alpha \in [0,1)$. Then, with the extreme value thresholds $T(\alpha_n, |\Omega_n|)$ defined in Eq. (3.9), we have*

$$\lim_{n\to\infty} \mathbf{P}\{x_n \in R(Y_n, T(\alpha_n, |\Omega_n|));\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\} = 1-\alpha. \tag{3.10}$$

Hence, the sets $R(Y_n, T(\alpha_n, |\Omega_n|))$ defined in (3.7) are asymptotically sharp confidence regions with significance level α.

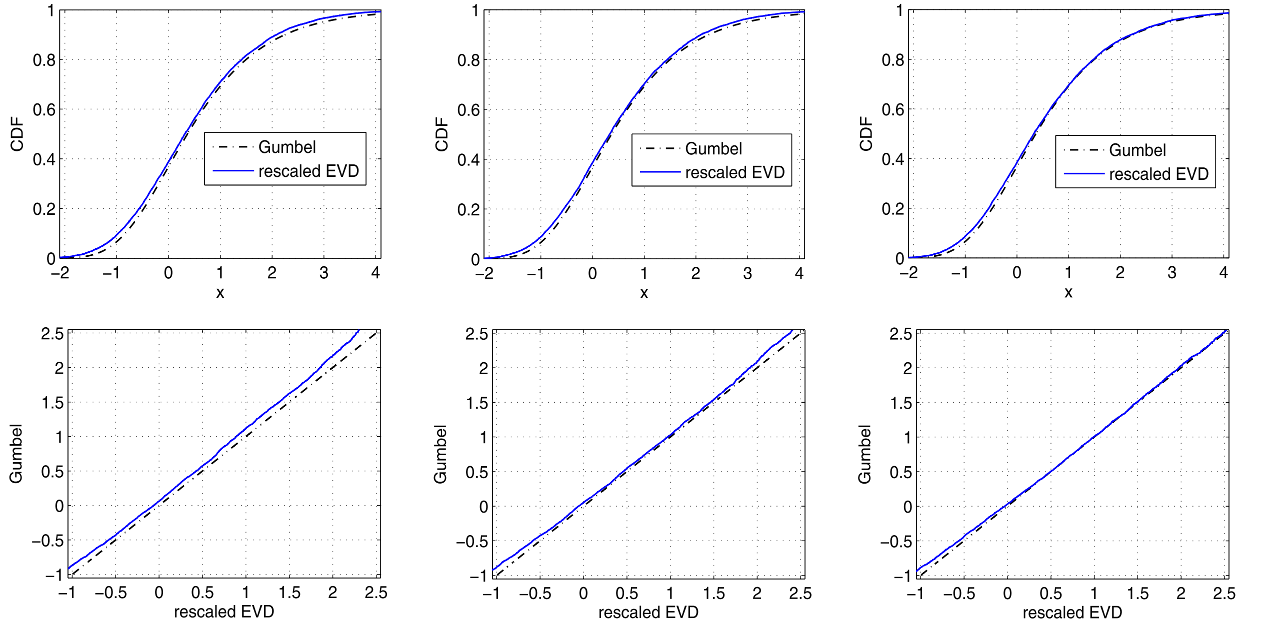

Fig. 4. Top: Rescaled distribution of $\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty$ and the Gumbel distribution for $n = 128$ (left), $n = 512$ (middle) and $n = 1024$ (right). Bottom: Q–Q plot of those distributions.

Proof. According to (3.8) it is sufficient to show that $\mathbf{P}\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant T(\alpha_n, |\Omega_n|)\} \to 1-\alpha$ as $n \to \infty$. Theorem 1.3 and the definition of the thresholds in (3.9) imply that the probability of the event $\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant T(\alpha_n, |\Omega_n|)\}$ converges to $\exp(-\exp(-z(\alpha)))$ as $n \to \infty$. Since the quantile $z(\alpha)$ is defined as the solution of $\exp(-\exp(-z)) = 1-\alpha$ this yields Eq. (3.10). □

Corollary 3.7. *Let $(\mathcal{D}_n)_{n\in\mathbb{N}}$ be any family of frames (not necessarily asymptotic stable) having normalized elements, and consider the data model $Y_n = x_n + \boldsymbol{\Phi}_n\epsilon_n$ with noise vectors ϵ_n having possibly dependent $N(0,\sigma^2)$-distributed entries. Then, it still holds that*

$$\liminf_{n\to\infty} \mathbf{P}\bigl\{x_n \in R\bigl(Y_n, T\bigl(\alpha_n, |\Omega_n|\bigr)\bigr);\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\bigr\} \geqslant 1-\alpha. \tag{3.11}$$

Proof. This follows from Theorem 3.6 and Lemma A.9. □

Notice, that in Corollary 3.7 the sets $R(Y_n, T(\alpha_n, |\Omega_n|))$ are not necessarily asymptotically sharp confidence regions, in the sense that inequality (3.11) may be strict. Actually, we believe that asymptotical stability of the frames $\mathcal{D}_n$ is close to being necessary for the sets $R(Y_n, T(\alpha_n, |\Omega_n|))$ defining asymptotically sharp confidence regions. For specific highly redundant dictionaries where asymptotic stability fails to hold (such as the translation invariant wavelet frame; see Section 4.1.4) we expect that $\mathbf{P}\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant \sigma a_n z + \sigma b_n\}$ still converges to the Gumbel distribution – however with normalization sequences a_n and b_n being strictly smaller than $\sigma a(\chi, |\Omega_n|)$ and $\sigma b(\chi, |\Omega_n|)$. If this is the case, then the smaller thresholds $T_n = \sigma a_n z(\alpha_n) + \sigma b_n$ again define sharp confidence regions. Surprisingly, results on the distributional convergence of $\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty$ or even of $\max(\boldsymbol{\Phi}_n\epsilon_n)$ for redundant frames are almost nonexistent.

3.4. Rate of approximation

Strictly taken, Theorem 3.6 only claims that the $\|\cdot\|_\infty$-balls $R(Y_n, T(\alpha_n, |\Omega_n|))$ turn into confidence regions in the limit $n \to \infty$, but it does not directly give any result for finite n. Sometimes it is argued that, even in the independent case without taking absolute values, the rate of convergence of $\mathbf{P}\{\max(\boldsymbol{\Phi}_n\epsilon_n) \leqslant T\}$ to the Gumbel distribution is known to be rather slow (see, for example, [18, Section 2.4]). Another option could be to derive non-asymptotic coverage probabilities along the lines of [30], however at the price of typically quite conservative confidence bands.

Nevertheless, numerical simulations clearly demonstrate, that even for moderate n, the approximation of $\mathbf{P}\{\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty \leqslant \sigma a(\chi, |\Omega_n|)z + \sigma b(\chi, |\Omega_n|)\}$ with the limiting Gumbel distribution is quite good. This even holds true for redundant frames as can be seen from Fig. 4, where the distribution functions of the rescaled maxima of the coefficients with respect to the two times oversampled sine frame of Example 2.7 are compared with the limiting Gumbel distribution. The top line in Fig. 4 displays the normalized empirical distributions of $\|\boldsymbol{\Phi}_n\epsilon_n\|_\infty$ for signal lengths of $n = 128$, $n = 512$ and $n = 1024$ (computed from 10 000 realizations in each case) and the limiting Gumbel distribution. As can be seen, there is only a small difference between those functions. The bottom line in Fig. 4 shows a Q–Q plot (quantile against quantile) of those distributions and again indicates that the quantiles of the rescaled maximum for finite n are quite well approximated by the ones of the limiting Gumbel distribution.

3.5. Smoothness estimates

We have just seen that the x_n is contained in the confidence regions $R(Y_n, T(\alpha_n, |\Omega_n|))$ around the data Y_n with probability tending to $1-\alpha$. Moreover, by definition, the soft-thresholding estimate $\hat{x}_n = \mathbf{S}(Y_n, T(\alpha_n, |\Omega_n|))$ is contained in $R(Y_n, T(\alpha_n, |\Omega_n|))$, too. The following theorem shows that the soft-thresholding estimate is indeed the smoothest element in this confidence region, with smoothness measured in terms of a wide class of functionals.

Theorem 3.8 *(Smoothness estimates). Let $(\mathcal{J}_n)_{n\in\mathbb{N}}$ be a family of functionals $\mathcal{J}_n : \mathbb{R}^{\Omega_n} \to \mathbb{R} \cup \{\infty\}$ having the property that*

$$\mathcal{J}_n(x_n) \leqslant \mathcal{J}_n(\bar{x}_n) \quad \text{whenever } \bigl|x_n(\omega)\bigr| \leqslant \bigl|\bar{x}_n(\omega)\bigr| \text{ for all } \omega \in \Omega_n. \tag{3.12}$$

Moreover, consider the data model $Y_n = x_n + \boldsymbol{\Phi}_n \epsilon_n$, where $(\epsilon_n)_{n\in\mathbb{N}}$ is a sequence of random vectors with $N(0, \sigma^2)$-distributed entries, let $(\alpha_n)_{n\in\mathbb{N}}$ be a sequence in $(0,1)$ converging to some $\alpha \in [0,1)$, and denote $\hat{x}_n := \mathbf{S}(Y_n, T(\alpha_n, |\Omega_n|))$. Then,

$$\liminf_{n\to\infty} \mathbf{P}\bigl\{\mathcal{J}_n(\hat{x}_n) \leqslant \mathcal{J}_n(x_n);\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\bigr\} \geqslant 1-\alpha. \tag{3.13}$$

Hence, the soft-thresholding estimate $\hat{x}_n$ is at least as smooth as the original parameter x_n, with probability tending to $1-\alpha$ as $n \to \infty$, where smoothness is measured in terms of any family of functionals $\mathcal{J}_n$ satisfying (3.12).

Proof. The definition of the soft-thresholding function implies that $\hat{x}_n$ is an element of the confidence region $R(Y_n, T(\alpha_n, |\Omega_n|))$ and that for every other element $\bar{x}_n$ contained in this confidence region we have $|\hat{x}_n(\omega)| \leqslant |\bar{x}_n(\omega)|$ for all $\omega \in \Omega_n$. By Corollary 3.7 the true parameter x_n is contained in $R(Y_n, T(\alpha_n, |\Omega_n|))$, too, with a probability tending to $1-\alpha$. We conclude that

$$\liminf_{n\to\infty} \mathbf{P}\bigl\{\bigl|\hat{x}_n(\omega)\bigr| \leqslant \bigl|x_n(\omega)\bigr|;\ \omega \in \Omega_n;\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\bigr\} \geqslant 1-\alpha. \tag{3.14}$$

Assumption (3.12) on component-wise monotonicity of the functionals $\mathcal{J}_n$ now implies that the event $\{|\hat{x}_n(\omega)| \leqslant |x_n(\omega)|;\ \forall \omega \in \Omega_n;\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\}$ is contained in the event $\{\mathcal{J}_n(\hat{x}_n) \leqslant \mathcal{J}_n(x_n);\ \forall x_n \in \operatorname{Ran}(\boldsymbol{\Phi}_n)\}$. Together with (3.14) this yields (3.13). □

Remark 3.9 *(Shrinkage property).* The proof of Theorem 3.8 uses two main ingredients: First, soft-thresholding selects that element in $R(Y_n, T(\alpha_n, |\Omega_n|))$ which has minimal component-wise magnitudes and second, the true coefficient $\hat{x}_n$ is contained in the set $R(Y_n, T(\alpha_n, |\Omega_n|))$ with probability tending to $1-\alpha$. The former property is often referred to as the *shrinkage property* of soft-thresholding and has already been used in [7] for deriving smoothness estimates for orthogonal wavelet soft-thresholding. The second property relies on our extreme value result derived in Theorem 1.3. Notice, that the weaker result $\mathbf{P}\{\mathcal{J}_n(\hat{x}_n) \leqslant \mathcal{J}_n(x_n)\} \to 0$ using the threshold $\sigma\sqrt{2\log|\Omega_n|}$ is well known; compare [16]. However, the proof of Theorem 3.8 reveals that for asymptotically stable frames the considered thresholds $T(\alpha_n, |\Omega_n|)$ are close to being the smallest ones yielding smoothness estimates of the form (3.13). For strongly redundant frames, however, where asymptotic stability fails to hold, smaller thresholds yielding the same smoothness bounds can exist. In Theorem 4.9 we show that this is indeed the case for the dyadic discrete translation invariant wavelet system.

Basic but important examples for functionals satisfying the component-wise monotonicity property (3.12) are powers of weighted ℓ^2-norms,

$$\|x_n\|_2 := \sqrt{\sum_{\omega\in\Omega_n} c(\omega)\bigl|x_n(\omega)\bigr|^2} \quad \text{for some } c(\cdot) > 0.$$

In the case of wavelet and Fourier frames, these norms of the coefficients provide norm equivalents to Sobolev norms in the original signal domain (assuming an appropriate discretization model $u \mapsto u_n$). Sobolev norms are definitely the most basic smoothness measures of functions. More general and also practically relevant classes of smoothness measures are Besov norms. Assume for the moment that $\mathcal{D}_n$ is a wavelet frame where the index set has the multiresolution form $\Omega_n = \{(\lambda, k):\ \lambda \in \Lambda_n \text{ and } k \in D_\lambda\}$ for some index sets Λ_n and D_λ corresponding to scale/resolution and scale dependent location, respectively. In this case one takes the functional $\mathcal{J}_n$ as one of the weighted $\ell^{p,q}$-norms

$$\|x_n\|_{p,q} := \sqrt[q]{\sum_{\lambda\in\Lambda_n} c(\lambda)\bigl\|x_n(\lambda,\cdot)\bigr\|_p^q} \quad \text{for some } c(\cdot) > 0.$$

These norms again satisfy the monotonicity property (3.12) and moreover yield to norm equivalents of Besov norms for properly chosen weights $c(\lambda)$; see Section 4.1. Such weighted (p,q)-norms are also reasonable in combination with other multiresolution systems, such as the curvelet frame (see Section 4.2).

3.6. Risk estimates

Although the main focus in this work is on confidence regions and smoothness estimates, in the following Proposition 3.10 we shall verify that using the EVTs of Definition 3.5 yields risk estimates similar to the oracle inequalities of [8]. The following result is non-standard regarding two aspects: First, it allows arbitrary frames instead of orthonormal bases. Second, and more importantly, it considers our more general class of extreme value thresholds instead of the universal threshold $\sigma\sqrt{2\log|\Omega_n|}$.

Proposition 3.10 *(Oracle inequality). Let $\mathcal{D}_n = (\phi_\omega^n\colon\ \omega \in \Omega_n)$ be a frame in $\mathbb{R}^{I_n}$ with corresponding analysis operator $\boldsymbol{\Phi}_n$. Moreover, let $\hat{u}_n = \boldsymbol{\Phi}_n^+ \circ \mathbf{S}(\boldsymbol{\Phi}_n(V_n), T)$ denote the soft-thresholding estimator in (2.6) corresponding to the extreme value thresholds $T_n = T(\alpha_n, |\Omega_n|)$ defined by Eq. (3.9), and assume for simplicity that $T(\alpha_n, |\Omega_n|) \leqslant \sigma\sqrt{2\log|\Omega_n|}$. Then, we have*

$$\mathbf{E}\big(\|u_n - \hat{u}_n\|^2\big) \leqslant \frac{\sigma^2}{a_n}\left(\log\big(1/(1-\alpha_n)\big)\sqrt{\pi\log|\Omega_n|} + \big(1 + 2\log|\Omega_n|\big)\sum_{\omega\in\Omega_n}\min\left\{1, \frac{|\langle\phi_\omega^n, u_n\rangle|^2}{\sigma^2}\right\}\right). \tag{3.15}$$

Here a_n is the lower frame bound of $\mathcal{D}_N$; see Eq. (2.1).

Proof. See Appendix B.1. □

4. Examples from signal and image denoising

In this section we verify that many important frames used for thresholding in signal and image processing are asymptotically stable and thus covered by the results of the previous section. These examples include redundant wavelet systems and curvelet frames. We also consider an important example, where our basic asymptotic stability fails to hold; namely the discrete translation invariant wavelet frame. Actually, we show that not even the result of Theorem 1.3 (and thus all of its implications) holds in this case. This indicates that the stated conditions are close to being necessary for the asymptotical distributional law of Theorem 1.3. Further, we derive confidence regions and smoothness estimates for the translation invariant wavelet transform that significantly improve over simple application of Proposition 3.4, item (b) (and also the main result of [31]).

4.1. Redundant and non-redundant wavelet denoising

In the following we consider one dimensional wavelet denoising. The generalization to higher dimensional wavelet denoising is straightforward. We shall discuss thresholding in biorthogonal wavelet bases, certain overcomplete wavelet frames (using the so called cycle spinning procedure), and fully translation invariant wavelet systems. Before considering those particular examples, we collect some notation and present basic facts about biorthogonal wavelets (which include the orthogonal ones) that we need for the application of our general results.

4.1.1. Biorthogonal wavelet bases

One dimensional wavelets are generated by dilating and translating a single function, the so called mother wavelet. The distinguished feature of wavelet systems is that various classical smoothness measures (Triebel, Sobolev and Besov norms) can be characterized by simple norms in the wavelet domain. In the following, for the sake of simplicity, we only consider real valued periodic wavelets on the interval $[0, 1]$. Moreover, we restrict ourselves to compactly supported biorthogonal wavelets that arise from a multiresolution decomposition.

Denote by Ω the set of all index pairs of the form (j,k) with $j \in \mathbb{N}$ and $k \in \{0, \ldots, 2^j - 1\}$. The index j is refereed to as resolution or scale index and k to as the discrete location index. Moreover, let $\phi, \psi \in L^2(\mathbb{R})$ denote the father and mother wavelet, respectively, which are assumed to be compactly supported and to have unit norm with respect to $\|\cdot\|_2$, the Euclidean norm on $L^2(0,1)$. For any $(j,k) \in \Omega$ one then defines (periodic) wavelets $\psi_{j,k}$ and (periodic) scaling functions $\phi_{j,k}$ on $[0, 1]$ by

$$\big(\forall t \in [0,1]\big) \quad \psi_{j,k}(t) = 2^{j/2}\sum_{m\in\mathbb{Z}}\psi\big(2^j(t-m) - k\big), \quad \phi_{j,k}(t) = 2^{j/2}\sum_{m\in\mathbb{Z}}\phi\big(2^j(t-m) - k\big).$$

The wavelet and the scaling coefficients of some signal $u \in L^2(0,1)$ are then simply the inner products of u with the wavelets $\psi_{j,k}$ and the scaling functions $\phi_{j,k}$, respectively. We further write $\mathbf{W}, \mathbf{V}\colon L^2(0,1) \to \ell^2(\Omega)$ for the mappings that take the signal $u \in L^2(0,1)$ to the inner products $(\mathbf{W}u)(j,k) := \langle\psi_{j,k}, u\rangle$ and $(\mathbf{V}u)(j,k) := \langle\phi_{j,k}, u\rangle$, respectively.

In order to get a (biorthogonal) wavelet basis one has to impose some completeness condition and some connections between the wavelets and the scaling functions. Such assumptions are most naturally formulated in the multiresolution framework (below already adapted to the periodic setting). Hence, in the following we assume the existence of subspaces $\mathcal{V}_j$ and $\mathcal{W}_j$ of $L^2(0,1)$, referred to as scaling and detail spaces, respectively, meeting the following requirements:

- For every $j \in \mathbb{N}$, the following mappings are bijections:

$$\mathcal{V}_j \to \mathbb{R}^{2^j}\colon u \mapsto \left(\langle \phi_{j,k}, u\rangle\colon k \in \left\{0, \ldots, 2^j - 1\right\}\right),$$
$$\mathcal{W}_j \to \mathbb{R}^{2^j}\colon u \mapsto \left(\langle \psi_{j,k}, u\rangle\colon k \in \left\{0, \ldots, 2^j - 1\right\}\right).$$

- For every $j \in \mathbb{N}$, we have the multiresolution decomposition $\mathcal{V}_j = \mathcal{V}_{j-1} \oplus \mathcal{W}_{j-1}$.
- The union $\bigcup_{j\in\mathbb{N}} \mathcal{V}_j$ is dense in $L^2(0, 1)$.

Repeated application of the multiresolution decomposition yields the decomposition of the signal space into the sum of the scaling space $\mathcal{V}_0$ and the wavelet space $\mathcal{W} := \bigoplus_{j\geqslant 0} \mathcal{W}_j$. Moreover, the above conditions imply that there is a stable one to one correspondence between any element in $\mathcal{W}$ and its inner product with respect to $\mathcal{D} := (\psi_{j,k}\colon\ (j, k) \in \Omega)$. Moreover, the multiresolution decomposition serves as the basis of both, discretization and fast implementation. Notice that the construction of compactly supported orthogonal and biorthogonal wavelets is non-trivial and such systems have been constructed for the first time in [32,33]. By now such wavelet systems are well known; a detailed construction of orthogonal and biorthogonal wavelet systems together with many interesting details may be found in [12,14,34].

Remark 4.1 *(Biorthogonal basis).* If the spaces $\mathcal{V}_j$ and $\mathcal{W}_j$ are orthogonal to each other, then $\mathcal{D}$ is an orthonormal wavelet basis and $\mathcal{V}_j$ and $\mathcal{W}_j$ are spanned by the scaling and wavelet functions, respectively. However, we do not require orthogonality in the following. In this more general case, the scaling and wavelet spaces are spanned by certain dual systems (or biorthogonal bases; thus the name). Biorthogonal wavelets are often preferred to strictly orthogonal ones since they allow more freedom to adapt them to a particular application in mind. Especially, opposed to orthogonal wavelets, biorthogonal wavelets can at the same time be smooth, symmetric and compactly supported.

Remark 4.2 *(Computing the wavelet transform).* The multiresolution decomposition $\mathcal{V}_j = \mathcal{V}_{j-1} \oplus \mathcal{W}_{j-1}$ is the basis for fast computation of the wavelet transform. Given the scaling coefficients at some scale $L > 0$, the scaling and wavelet coefficients at scale $L - 1$ can be computed by cyclic convolution of the given scaling coefficients with a certain discrete filter pair. Repeated application of this procedure eventually yields all scaling and all wavelet coefficients at scales below L. In the case of biorthogonal wavelets, the multiresolution decomposition can be inverted again by repeated application of convolution with a different pair of reconstruction filters.

Throughout the following we assume that a discrete signal $u_n \in \mathbb{R}^n$ is given, where the discretization number $n = 2^J$ is an integer power of some maximal level of resolution. One then interprets the components of the discrete signal as the scaling coefficients of some underlying continuous domain signal, that is,

$$\left(\forall k \in \{0, \ldots, n-1\}\right) \quad u_n(k) = (\mathbf{V}u)(J, k) = \langle \phi_{J,k}, u\rangle.$$

Obviously there are infinitely many continuous domain signals yielding to the same scaling coefficients. However, according to the made assumptions, there exists a unique element in the scaling space $\mathcal{V}_J$ having scaling coefficients u_n. This element will be denoted as $u_n^* \in \mathcal{V}_J$.

The wavelet coefficients of the discrete signal are then simply defined as the wavelet coefficients of the continuous domain signal with indices in

$$\Omega_n := \left\{(j, k)\colon\ j \in \{0, \ldots, J-1\} \text{ and } k \in \left\{0, \ldots, 2^j - 1\right\}\right\}.$$

According to the multiresolution decomposition, these coefficients depend only on the discrete signal and can also be written as discrete inner products

$$\left(\forall u_n \in \mathbb{R}^n\right)\left(\forall (j, k) \in \Omega_n\right) \quad \left\langle \psi_{j,k}^n, u_n\right\rangle := \langle \psi_{j,k}, u\rangle.$$

This serves as definition of both, the discrete wavelets $\psi_{j,k}^n \in \mathbb{R}^n$ and the wavelet coefficients of u_n. Finally we define $\mathcal{D}_n$ as the family of all discrete wavelets $\psi_{j,k}^n$ and denote by $\mathbf{W}_n : \mathbb{R}^n \to \mathbb{R}^{\Omega_n}$ the corresponding analysis operator, which we refer to as the *discrete wavelet transform*.

Remark 4.3 *(Numerical computation).* The discrete wavelet transform is computed by repeated application of the multiresolution decomposition, yielding to all discrete wavelet coefficients and the scaling coefficient $u_1 = \langle \phi_{0,0}, u\rangle$; see Remark 4.2. Since the discrete filters usually have small support, the wavelet transform can be computed using only $\mathcal{O}(n)$ operation counts and the same holds true for recovering u_n from those coefficients. Notice that the discrete wavelets are never computed explicitly in the multiresolution algorithm. We defined them in order to verify our general framework. Finally, we stress again that the wavelet coefficients of u_n coincide with the one of u up to scale $\log(n/2)$.

4.1.2. Biorthogonal basis denoising

Now consider the denoising problem (1.1), which simple reads $V_n = u_n + \epsilon_n$. The wavelet soft-thresholding procedure is usually only applied to coefficients above some scale; compare with Remark 2.1. For simplicity we shall consider the case where all wavelet coefficients are thresholded but not the scaling coefficient. Hence, the wavelet soft-thresholding estimator (for the wavelet part of u_n) is defined by

$$\hat{u}_n = \mathbf{W}_n^{-1} \circ \mathbf{S}(\mathbf{W}_n V_n, T).$$

Thanks to the multiresolution algorithm, the wavelet soft-thresholding estimator can be computed with only $\mathcal{O}(n)$ operation counts.

We measure smoothness of the considered estimates in terms of Besov norms. To that end, assume that the mother wavelet has sufficiently many vanishing moments and is sufficiently smooth. Then, for given norm parameters $p, q \geqslant 1$ and given smoothness parameter $r \geqslant 0$ one defines

$$\|x\|_{p,q,r} = \sqrt[q]{\sum_{j\in\mathbb{N}} 2^{jsq} \|x(j,\cdot)\|_p^q} \quad \text{with } s = r + \frac{1}{2} - \frac{1}{p}$$

for any $x \in \ell^2(\Omega)$ and with $\|\cdot\|_p$ denoting the usual ℓ^p-norm taken for fixed scale $j \in \mathbb{N}$. It is then clear that any of these norms satisfies the component-wise monotony property (3.12). We further write $\|u\|_{\mathcal{B}^r_{p,q}} := \|\mathbf{W}u\|_{p,q,r}$ for the corresponding norm of some $u \in L^2(0,1)$ and finally denote by $\mathcal{B}^r_{p,q}$ the set of all signals having finite norm $\|u\|_{\mathcal{B}^r_{p,q}} < \infty$. The pair $(\mathcal{B}^r_{p,q}, \|\cdot\|_{\mathcal{B}^r_{p,q}})$ is a Banach space and referred to as Besov space. The given definitions provide norm equivalents of $\|\cdot\|_{\mathcal{B}^r_{p,q}}$ to the definition of Besov norms in classical analysis, as long as the mother wavelet has $m > r$ vanishing moments and is m times continuously differentiable.

Theorem 4.4 *(Soft-thresholding in wavelet bases). The discrete wavelet bases $\mathcal{D}_n = (\psi^n_{j,k}\colon (j,k) \in \Omega_n)$ are asymptotically stable. In particular, the following hold*:

(a) Distribution: *Let ϵ_n be a sequence of noise vectors in $\mathbb{R}^n$ with independent $N(0,\sigma^2)$-distributed entries. Then, the sequence $\|\mathbf{W}_n \epsilon_n\|_\infty$ is of Gumbel type with normalization constants $\sigma a(\chi, n)$, $\sigma b(\chi, n)$ defined by (1.9), (1.10).*

(b) Confidence regions: *Let $(\alpha_n)_{n\in\mathbb{N}} \subset (0,1)$ be a sequence of significance levels converging to some $\alpha \in [0,1)$ and let $T(\alpha_n, n)$ denote the corresponding EVTs defined in (3.9). Then,*

$$\lim_{n\to\infty} \mathbf{P}\{\|\mathbf{W}_n(u_n - V_n)\|_\infty \leqslant T(\alpha_n, n);\ \forall u_n \in \mathbb{R}^{I_n}\} = 1 - \alpha.$$

(c) Smoothness: *Let $\hat{u}^*_n$ denote the soft-thresholding estimator using the extreme value thresholds $T(\alpha_n, n)$. If the considered mother wavelet has $m > r$ vanishing moments and is m times continuously differentiable, then*

$$\liminf_{n\to\infty} \mathbf{P}\{\|\hat{u}^*_n\|_{\mathcal{B}^r_{p,q}} \leqslant \|u\|_{\mathcal{B}^r_{p,q}};\ \forall u \in \mathcal{B}^r_{p,q}\} \geqslant 1 - \alpha.$$

Proof. By definition, for any $n \in \mathbb{N}$ and pair of any indices $(j,k), (j',k')$, the inner products $\langle \psi^n_{j,k}, \psi^n_{j',k'}\rangle$ of the discrete wavelets coincide with the inner product $\langle \psi_{j,k}, \psi_{j',k'}\rangle$ of the continuous domain wavelets. Since the family $(\psi_{j,k}\colon (j,k) \in \Omega)$ is a Riesz basis with normalized elements this immediately yields condition (ii) required in Definition 1.2 for asymptotically stable frames.

Condition (i) of Definition 1.2 is satisfied since all $|\langle \psi_{j,k}, \psi_{j',k'}\rangle|$ are bounded away from one. To see that this holds true, it is sufficient to consider the case where $\psi(2^j t - k)$ and $\psi(2^{j'} t - k')$ are both supported in the interval $(0,1)$ and satisfy $j' \leqslant j$. Application of the substitution rule yields

$$\begin{aligned} |\langle \psi_{j,k}, \psi_{j',k'}\rangle| &= 2^{j/2+j'/2} \left| \int_{\mathbb{R}} \psi(2^j t - k) \psi(2^{j'} t - k')\, dt \right| \\ &= 2^{(j-j')/2} \left| \int_{\mathbb{R}} \psi(2^{j-j'} t - k + 2^{j-j'} k) \psi(t)\, dt \right| = |\langle \psi_{j-j',k-2^{j-j'}k}, \psi\rangle|. \end{aligned}$$

Because all wavelets have unit norm, the Cauchy–Schwarz inequality shows $|\langle \psi_{j-j',k-2^{j-j'}k}, \psi\rangle| < 1$. The upper frame bound implies that $\sum_{(j,k)\in\Omega_n} |\langle \psi_{j,k}, \psi\rangle|^2 < \infty$, and hence the sequence $(\langle \psi_{j,k}, \psi\rangle\colon (j,k) \in \Omega_n)$ in particular converges to zero. As a consequence, the numbers $|\langle \psi_{j,k}, \psi_{j',k'}\rangle|$ are uniformly bounded by some constant $\rho < 1$.

The other claims in items (a)–(c) then follow from the asymptotic stability of the frames $\mathcal{D}_n$ and the general results derived the previous section. Actually, the first two items are just restatements of Theorems 1.3 and 3.6 adapted to the wavelet setting. For item (c) note that the norms $\|\cdot\|_{p,q,r}$ satisfy the component-wise monotony property (3.12) and therefore Theorem 3.8 yields

$$\liminf_{n\to\infty} \mathbf{P}\left\{\|\hat{x}_n\|_{p,q,r} \leqslant \|\mathbf{W}_n u_n\|_{p,q,r};\ \forall u \in \mathcal{B}^r_{p,q}\right\} \geqslant 1-\alpha \quad \text{with } \hat{x}_n := \mathbf{S}\big(V_n, T(\alpha_n, |\Omega_n|)\big).$$

By definition we have $\|\hat{x}_n\|_{p,q,r} = \|\hat{u}^*_n\|_{\mathcal{B}^r_{p,q}}$ and the inequality $\|\mathbf{W}_n u_n\|_{p,q,r} \leqslant \|u\|_{\mathcal{B}^r_{p,q}}$ for all n which finally yields item (c) and concludes the proof. □

4.1.3. Cycle spinning

A mayor drawback of thresholding in a wavelet basis is its missing translation invariance. This typically causes visually disturbing Gibbs-like artifacts near discontinuities at non-dyadic locations. One way to significantly reduce these artifacts is via so called cycle spinning (see [6]). The idea there is to reduce the artifacts by averaging several estimates obtained by denoising shifted copies of the noisy data.

Let $\mathcal{D}_n = (\psi^n_{j,k}\colon (j,k) \in \Omega_n)$ be an orthonormal wavelet basis and denote by $\mathbf{T}_m : \mathbb{R}^n \to \mathbb{R}^n$ the cyclic translation operator, defined by $(\mathbf{T}_m u_n)(k) = u_n(k-m)$ for $u_n \in \mathbb{R}^n$ and all $k,m \in \{0,\dots,n-1\}$. Cycle spinning then averages the wavelet soft-thresholding estimates of the translated data $\mathbf{T}_m u_n$ over all shifts $m = 0,\dots,M-1$, where M is some prescribed number of considered translations. Hence, one defines

$$\hat{u}_{n,M} := \frac{1}{M}\sum_{m=0}^{M-1} \mathbf{T}_{-m}\mathbf{W}^*_n \circ \mathbf{S}(\mathbf{W}_n\mathbf{T}_m V_n, T). \tag{4.1}$$

The following elementary Lemma 4.5 states that the cycle spinning estimator (4.1) is equal to the soft-thresholding estimator defined by Eq. (2.6) corresponding to the overcomplete wavelet frame

$$\mathcal{D}_{n,M} := \big(\mathbf{T}_{-m}\psi_{j,k}\colon (j,k,m) \in \Omega_{n,M}\big) \quad \text{with } \Omega_{n,M} := \Omega_n \times \{0,\dots,M-1\}. \tag{4.2}$$

Hence wavelet cycle spinning fits into the general framework of soft-thresholding introduced in Section 2.

Lemma 4.5. *Let $\mathcal{D}_{n,M}$ be the overcomplete wavelet frame defined in (4.2) and denote by $\mathbf{W}_{n,M} : \mathbb{R}^n \to \mathbb{R}^{nM}$ the corresponding analysis operator. Then, the cycle spinning estimator (4.1) has the representations*

$$\hat{u}_{n,M} = \frac{1}{M}\mathbf{W}^*_{n,M} \circ \mathbf{S}(\mathbf{W}_{n,M}V_n, T) = \mathbf{W}^+_{n,M} \circ \mathbf{S}(\mathbf{W}_{n,M}V_n, T). \tag{4.3}$$

Hence the cycle spinning estimator equals the soft-thresholding estimator corresponding to the redundant wavelet frame $\mathcal{D}_{n,M}$.

Proof. The first identity in (4.3) immediately follows from (4.2) and (4.1). Next we verify the second equality. Since the decimated wavelet transform $\mathbf{W}_n$ and the translation operators $\mathbf{T}_m$ are isometries, we have

$$\|\mathbf{W}_{n,M}u_n\|^2 = \sum_{m=0}^{M-1} \|\mathbf{W}_n\mathbf{T}_m u_n\|^2 = M\|u_n\|^2$$

whenever u_n are the scaling coefficients of a member u of the wavelet space $\mathcal{W}$. Hence $\mathcal{D}_{n,M}$ is a tight frame with frame bound equals M. This implies that the dual synthesis operator $\mathbf{W}^+_{n,M}$ corresponding to the cycle spinning frame is simply given by $\mathbf{W}^+_{n,M} = \mathbf{W}^*_{n,M}/M$, which yields the second equality in (4.3). □

In the following we will show that redundant cycle spinning frame is asymptotic stable and thus allows the application of our general results. Strictly taken, these conditions do not hold for the frame $\mathcal{D}_{n,M}$, since some of the elements $\mathbf{T}_m\psi_{j,k}$ occur more than once in $\mathcal{D}_{n,M}$. In particular, the cardinality of the set $\{\mathbf{T}_m\psi_{j,k}\colon (j,k,m) \in \Omega_{n,M}\}$ is strictly less than $|\Omega_{n,M}| = nM$; the exact number of different frame elements is computed in the following Lemma 4.6. Asymptotic stability will then be satisfied for the frame that contains every element $\mathbf{T}_m\psi_{j,k}$ exactly once.

Lemma 4.6. *For any $M \leqslant n$, the number of different elements of the frame $\mathcal{D}_{n,M}$ defined in (4.2) is given by*

$$\big|\{\mathbf{T}_m\psi_{j,k}\colon (j,k,m) \in \Omega_{n,M}\}\big| = n\lfloor\log_2 M\rfloor + M\big(2^{\lceil\log_2 n/M\rceil} - 1\big). \tag{4.4}$$

Proof. The definition of the wavelet basis implies that $\psi_{j,k} = T_{n2^{-j}k}\psi_{j,0}$ for very $(j,k) \in \Omega_n$ and hence the periodicity of $\psi_{j,0}$ implies that

$$\mathbf{T}_m\psi_{j,k} = \mathbf{T}_{m+n2^{-j}k}\psi_{j,0} = \psi_{j,k+m2^j/n}$$

whenever $m2^j/n$ is an integer number. This shows that for every given scale index $j \in \{0,\dots,\log_2 n-1\}$, there are exactly $\min\{n, M2^j\}$ different wavelets. One concludes that

$$\begin{aligned}\left|\left\{\mathbf{T}_m\psi_{j,k}\colon (j,k,m)\in\Omega_{n,M}\right\}\right| &= n\left|\left\{j\colon n/M\leqslant 2^j\leqslant n/2\right\}\right| + M\sum_{M2^j<n}2^j\\ &= n\left(\log_2 n-\left\lceil\log_2\frac{n}{M}\right\rceil\right)+M\sum_{j=0}^{\lceil\log_2 n/M\rceil-1}2^j.\end{aligned}$$

This shows Eq. (4.4). □

In the following we shall for simplicity assume that M, the number of shifts in the cycle spin procedure, is an integer power of two. In this case, Eq. (4.4) simplifies to

$$\left|\left\{\mathbf{T}_m\psi_{j,k}\colon (j,k,m)\in\Omega_{n,M}\right\}\right| = n\log_2(M)+n-M. \tag{4.5}$$

Note that this is significantly smaller (at least for large M) than the naive bound Mn given by the cardinality of $\Omega_{n,M}$.

Theorem 4.7 *(Soft-thresholding using cycle spinning). Let M be any fixed integer power of two, denote by $\mathcal{D}_{n,M}=(\mathbf{T}_m\psi_{j,k}\colon (j,k,m)\in\Omega_{n,M})$ the overcomplete wavelet cycle spinning frame and by $\mathbf{W}_{n,M}$ the corresponding analysis operator. Then the following assertions hold true*:

(a) Distribution: *Let $(\epsilon_n)_{n\in\mathbb{N}}$ be a sequence of noise vectors in $\mathbb{R}^n$ with independent $N(0,\sigma^2)$-distributed entries. Then, the sequence $\|\mathbf{W}_{n,M}\epsilon_n\|_\infty$ is of Gumbel type with normalization constants $\sigma a(\chi,n)$ (defined by (1.9)) and $\sigma b_M(\chi,n)$, where*

$$b_M(\chi,n) := \sqrt{2\log n}+\frac{-\log\log n-\log\pi+2\log\log_2(M)}{2\sqrt{2\log n}}.$$

(b) Confidence regions: *Let $\alpha_n\in(0,1)$ be a sequence of significance levels converging to some $\alpha\in[0,1)$ and let $T(\alpha_n,|\mathcal{D}_{n,M}|)$ denote the corresponding EVTs defined in (3.9). Then,*

$$\lim_{n\to\infty}\mathbf{P}\left\{\left\|\mathbf{W}_{n,M}(u_n-V_n)\right\|_\infty\leqslant T\left(\alpha_n,|\mathcal{D}_{n,M}|\right);\ \forall u_n\in\mathbb{R}^{I_n}\right\}=1-\alpha.$$

(Here by some abuse of notation $|\mathcal{D}_{n,M}|$ denotes the number of different elements in that frame, see (4.5).) The same holds true if we replace $T(\alpha_n,|\mathcal{D}_{n,M}|)$ by

$$T_M(\alpha_n,n) := -\sigma a(\chi,n)\log\log\left(1/(1-\alpha_n)\right)+\sigma b_M(\chi,n). \tag{4.6}$$

(c) Smoothness: *Let $\hat{u}^*_{n,M}$ denote the soft-thresholding estimator using either $T(\alpha_n,|\mathcal{D}_n|)$ or $T_{n,M}(\alpha_n)$ as threshold. If the considered mother wavelet has $m>r$ vanishing moments and is m times continuously differentiable, then*

$$\liminf_{n\to\infty}\mathbf{P}\left\{\left\|\hat{u}^*_{n,M}\right\|_{\mathcal{B}^r_{p,q}}\leqslant\|u\|_{\mathcal{B}^r_{p,q}};\ \forall u\in\mathcal{B}^r_{p,q}\right\}\geqslant 1-\alpha.$$

Proof. By using the characterization of the cycle spinning estimator in Lemma 4.5 and the cardinality computed in Lemma 4.6, the proof follows the lines of the proof of Theorem 4.4. Again, one simply uses the fact that the discrete inner products coincide with continuous ones of functions forming an infinite dimensional frame. However, notice the change of the normalization sequences in item (a) which is also used for the threshold $T_{n,M}(\alpha_n)$. As easy to verify we have the asymptotic relation

$$\begin{aligned}a(\chi,n)z+b_M(\chi,n) &= \sqrt{2\log n}+\frac{2z-\log\log n-\log\pi+2\log\log_2(M)}{2\sqrt{2\log n}}\\ &= a\left(\chi,|\mathcal{D}_{n,M}|\right)z+b\left(\chi,|\mathcal{D}_{n,M}|\right)+o\left(1/\sqrt{2\log n}\right).\end{aligned}$$

From basic extreme value theory it follows that we can replace the sequence $a(\chi,|\mathcal{D}_{n,M}|)z+b(\chi,|\mathcal{D}_{n,M}|)$ by the one considered in item (a) and for the threshold $T_{n,M}(\alpha_n)$. □

Remark 4.8. This alternative form (4.6) for the EVTs for cycle spinning denoising has been introduced to allow a better comparison with the EVTs used in the basis case. It fact, it can be seen that the extreme value thresholds $T_M(\alpha_n,n)$ for the redundant wavelet frame $\mathcal{D}_{n,M}$ simply increase by the additive constant $(\log\log_2 M)/\sqrt{2\log n}$ when compared to the extreme value threshold $T(\alpha_n,n)$ for the non-redundant wavelet frame.

The sharp confidence regions of Theorem 4.7 require M to be a fixed number. In the fully translation invariant transform, to be discussed next, one takes $M=n$ dependent on the discretization level. This effects a strong dependence of large scale coefficients and that the distributional limit of item (a) in Theorem 4.7 will not longer hold true.

4.1.4. Fully translation invariant denoising

Translation invariant wavelet denoising introduced in [6,35–37] is similar to cycle spinning denoising. However, now one takes the whole range of $M = n$ integer shifts instead of taking it as a fixed number independent of n. That is, the translation invariant wavelet estimator is defined by

$$\hat{u}_{n,n} := \frac{1}{n}\sum_{m=0}^{n-1} \mathbf{T}_{-m}\mathbf{W}_n^* \circ \mathbf{S}(\mathbf{W}_n\mathbf{T}_m V_n, T). \tag{4.7}$$

Lemma 4.5 implies that $\hat{u}_{n,n}$ equals the soft-thresholding estimator $\mathbf{W}_{n,n}^{+}\mathbf{S}(\mathbf{W}_{n,n}V_n, T)$ where $\mathbf{W}_{n,n}$ denotes the analysis operator corresponding to the translation invariant wavelet frame

$$\mathcal{D}_{n,n} = \left(\mathbf{T}_m\psi_{j,k}^n\colon\ (j,k) \in \Omega_n \text{ and } m \in \{0,\dots,n-1\}\right).$$

Eq. (4.5) shows that the translation invariant wavelet frame contains $n\log_2 n$ different elements. Further, from the proof of this lemma it follows that $\mathcal{D}_{n,n}$ is a tight frame with frame bound equals n. After removing multiple elements in $\mathcal{D}_{n,n}$, the resulting frame is non-tight but still has upper frame bounds $b_n = n$ tending to infinity as $n \to \infty$. One concludes that condition (ii) fails to hold for the translation invariant wavelet transform. The increasing frame bounds somehow reflect the increasing redundancy and dependency of the coarse scale wavelets with increasing n.

One might conjecture that still the distribution result of Theorem 4.7 holds true with M replaced by n. However, we shall show that this is not the case. Intuitively, the increasing correlation of the coarse scale wavelets with increasing n causes the maximum $\|\mathbf{W}_{n,n}\epsilon_n\|_\infty$ to be in probability smaller than the maximum of $n\log_2 n$ independent coefficients. Although the sets if Theorem 4.7 are still confidence regions (as follows from Sidak's Lemma A.9), they are no longer sharp and the considered thresholds are unnecessarily large. The following theorem gives a much smaller radius for these confidence regions; in particular this significantly improves [31, Theorem 4.4].

Theorem 4.9 *(Translation invariant soft-thresholding). Assume that the mother wavelet ψ is continuously differentiable, which implies that $(\bar{\psi} * \psi)(t) = 1 - c^2t^2/2 + o(t^2)$ for some constant c. (Here $*$ denotes the circular convolution and $\bar{\psi}(s) := \psi(-s)$.) Further denote by $\mathbf{W}_{n,n}$ the corresponding discrete translation invariant wavelet transform.*

Then, the following assertions hold true:

(a) Distribution: *Let $(\epsilon_n)_{n\in\mathbb{N}}$ be a sequence of noise vectors in $\mathbb{R}^{I_n}$ with independent $N(0,\sigma^2)$-distributed entries. Then, for every $z \in \mathbb{R}$,*

$$\liminf_{n\to\infty} \mathbf{P}\left\{\left\|\mathbf{W}_{n,n}\epsilon_n\right\|_\infty \leqslant \sqrt{2\log n} + \frac{z + \log(c/\pi)}{\sqrt{2\log n}}\right\} \geqslant \exp\left(-e^{-z}\right). \tag{4.8}$$

(b) Confidence regions: *Let $\alpha_n \in (0,1)$ be a sequence converging to some $\alpha \in [0,1)$. Then, we have*

$$\liminf_{n\to\infty} \mathbf{P}\left\{\left\|\mathbf{W}_{n,n}(u_n - V_n)\right\|_\infty \leqslant T_n(\alpha_n);\ \forall u_n \in \mathbb{R}^{I_n}\right\} \geqslant 1-\alpha,$$

when using the thresholds $T_n(\alpha_n) := \sigma\sqrt{2\log n} + \sigma(2\log n)^{-1/2}\log\log(1/(1-\alpha_n)) + \log(c/\pi)$.

(c) Smoothness: *Let $\hat{u}_{n,n}^*$ denote the soft-thresholding estimator using the threshold $T_n(\alpha_n)$ defined in item* (b). *If the considered mother wavelet has $m > r$ vanishing moments and is m times continuously differentiable, then*

$$\liminf_{n\to\infty} \mathbf{P}\left\{\left\|\hat{u}_{n,n}^*\right\|_{\mathcal{B}_{p,q}^r} \leqslant \|u\|_{\mathcal{B}_{p,q}^r};\ \forall u \in \mathcal{B}_{p,q}^r\right\} \geqslant 1-\alpha.$$

Proof. The key to all results is the distribution bound given item (a). Its proof is somehow technical and is presented in Appendix B.2. The other claims follow from item (a) combined with the results of the previous sections (namely Theorems 1.3, 3.6 and 3.8), and are verified as the corresponding statements in the proof of Theorem 4.4. □

Remark 4.10. Consider a sequence of standardized normal vectors η_n each of them having $Mn(\log n)^r$ independent entries, where M is some fixed integer and $r \geqslant 0$ some fixed nonnegative number. From Proposition A.5 we know that $\|\eta_n\|_\infty$ is of Gumbel type with normalization sequences $a(\chi, Mn(\log n)^r)$ and $b(\chi, Mn(\log n)^r)$. One easily verifies that

$$a\left(\chi, Mn(\log n)^r\right)z + b\left(\chi, Mn(\log n)^r\right) = \sqrt{2\log n} + \frac{(-1/2+r)\log\log n + \log(M/\sqrt{\pi})}{\sqrt{2\log n}} + o(1/\sqrt{2\log n}). \tag{4.9}$$

This allows to compare the bound in (4.8) with the asymptotic distribution of a certain number of independent random variables. Indeed, comparing (4.8) with (4.9) we can conclude, that $\mathbf{W}_{n,n}\epsilon_n$ less or equal in probability than the maximum of $Mn\sqrt{\log n}$ independent normally distributed random variables with $M := [c\sqrt{\pi}]$. Hence (4.8) improves the primitive bound obtained from the distribution of $n\log n$ independent coefficients by a factor $\sqrt{\log n}/c$.

Remark 4.11. It is a difficult task to compute the asymptotic distribution of the translation invariant wavelet coefficients exactly. This is due to the fact that for coarse scales the coefficients get increasingly correlated, whereas on the fine scales the correlations remain bounded away from σ^2. No appropriate tools for asymptotic extreme value analysis of such mixed type random fields seem to exist. Nevertheless, we believe that the maxima of the translation invariant wavelet coefficients are of Gumbel type but with even smaller normalization constants than the ones used in (4.8). In particular, it may even turn out that the threshold $\sigma\sqrt{2\log n}$ provides the denoising property for the translation invariant system.

4.2. Curvelet thresholding

Second generation curvelets (introduced in [38–40]) are functions $\psi_{j,\ell,k}$ in $\mathbb{R}^2$ depending on a scale index $j \in \mathbb{N}$, an orientation parameter $\ell \in \{0,\dots,4\cdot 2^{\lceil j/2\rceil}-1\}$ and a location parameter $k \in \mathbb{Z}^2$. They are known to provide an almost optimal sparse approximation of piecewise C^2 functions with piecewise C^2 boundaries (as shown in [38]); this class of functions usually serves as accurate cartoon model for natural images. The main curvelet property yielding this approximation result is the increasing anisotropy at finer scales. This feature also distinguishes them from standard wavelets in higher dimension.

There exists other related function systems with similar properties. The cone adapted shearlet frame (introduced in [41–43]) is very similar to the curvelet frame and shares its optimality when approximating piecewise C^2 images with piecewise C^2 boundaries, see [44]. Yet another closely related function system are the contourlets introduced by Do and Vetterli [45,46]. For simplicity we focus on the curvelets; similar statements could be made for the shearlet and contourlet frames.

4.2.1. Discrete curvelet frames

The discrete curvelet transform computes inner products of $u_n \in \mathbb{R}^{n\times n}$ with discrete curvelets $\psi^n_{j,\ell,k} \in \mathbb{R}^{n\times n}$. As for the wavelet transform, the elements $\psi^n_{j,\ell,k}$ are not computed explicitly and defined implicitly by the transform algorithm. Different implementations of the continuous curvelet transform give rise to different discrete frame elements $\psi^n_{j,\ell,k}$. Current implementations of the curvelet transform are computed in the Fourier domain. Below we shall focus on the wrapping based implementation of the curvelet transform introduced in [47]. This transform is an isometry which makes the computation of its pseudoinverse particularly simple.

Let $n = 2^J$ be an integer power of two with J denoting the maximal scale index. The discrete curvelets and the discrete curvelet transform are composed of the following ingredients:

- First, define Λ_n as the set of all pairs (j,ℓ) satisfying $j \in \{0,\dots,\log_2 n - 2\}$ and $\ell \in \{0,\dots,4\cdot 2^{\lceil j/2\rceil}-1\}$. The index sets of the discrete curvelets is defined by

  $$\Omega_n := \big((j,\ell,k)\colon\ (j,\ell)\in\Lambda_n \text{ and } k \in D_{j,\ell}\big),$$

  where $D_{j,\ell} = \{0,\dots,K_{j,\ell;1}-1\}\times\{0,\dots,K_{j,\ell;2}-1\}$ for certain given numbers $K_{j,\ell;1}\sim 2^j$ and $K_{j,\ell;2}\sim 2^{j/2}$. One refers to j as scale index, ℓ as the orientation index, and k the location index.
- Next, one constructs smooth nonnegative window functions $w_{j,\ell}:\mathbb{R}^2\to\mathbb{R}$ satisfying the identity

  $$\big(\forall z\in\mathbb{R}^2\big)\quad \sum_{j=0}^{J-2}\sum_{\ell=0}^{4\cdot 2^{\lceil j/2\rceil}-1}\big|w_{j,\ell}(z)\big|^2 = 1.$$

  The functions $w_{j,\ell}$ are essentially obtained by anisotropic scaling and shearing a single window function; see [47] for a detailed construction.
- For any index triple $(j,\ell,k)\in\Omega_n$ the discrete curvelet at scale j, having orientation $\ell/2^{\lceil j/2\rceil}$ and location $k = (k_1/K_{j,\ell;1}, k_2/K_{j,\ell;2})$ is defined by its Fourier representation

  $$\big(\mathbf{F}_n\psi^n_{j,\ell,k}\big)(m) = \frac{w_{j,\ell}(m)}{c_{j,\ell}}e^{-2\pi i(m_1k_1/K_{j,\ell;1}-m_2k_2/K_{j,\ell;2})}. \tag{4.10}$$

  Here the coefficients $c_{j,\ell}$ are chosen in such a way that $\|\psi^n_{j,\ell,k}\| = 1$ and $\mathbf{F}_n$ denotes the discrete Fourier transform.
- Finally, one defines the curvelet frame $\mathcal{D}_n = (\psi^n_{j,\ell,k}\colon (j,\ell,k)\in\Omega_n)$ and denotes by $\mathbf{C}_n:\mathbb{R}^{n\times n}\to\mathbb{R}^{\Omega_n}$ the corresponding analysis operator, which has been named digital curvelet transform via wrapping in [47]. In the following we will refer to $\mathbf{C}_n$ simply as *discrete curvelet transform*. We emphasize again that the implementation of the discrete curvelet transform does not require to compute the curvelets $\psi^n_{j,\ell,k}$ explicitly.

Implementations of the discrete curvelet transform and its pseudoinverse using $\mathcal{O}(n^2\log n)$ operation counts are freely available at http://curvelet.org. Although this implementation does not use normalized frame elements, the constants $c_{j,\ell}$ can easily be computed after the actual curvelet transformation and applied for normalizing the curvelet coefficients prior to denoising. The denoising demo `fdct_wrapping_demo_denoise.m` included in the curvelet software package in fact computes the norms of the discrete curvelets and uses them for proper scaling of the chosen thresholds.

4.2.2. Curvelet denoising

We now consider our denoising problem (1.1), which, after taking the discrete curvelet transform, simply reads $Y_n = x_n + \mathbf{C}_n \epsilon_n$. As usual, the estimator we consider is soft-thresholding $\hat{x}_n = \mathbf{S}(Y_n, T)$ of the curvelet coefficients.

Similar to the wavelet case, we measure smoothness in terms of the weighted (p, q)-norms, depending on certain norm parameters $p, q \geqslant 1$ and a parameter $r \geqslant 0$ describing the degree of smoothness. More precisely, we define

$$\|x_n\|_{p,q,r} := \sqrt[q]{\sum_{j,\ell} 2^{jsq} \|x_n(j,\ell,\cdot)\|_p^q} \quad \text{with } s = r + \frac{3}{2}\left(\frac{1}{2} - \frac{1}{p}\right).$$

These types of norms applied to the continuous domain curvelet coefficients have been defined and studied in [48]. In that paper also relations between these norms and classical Besov norms have been derived.

Theorem 4.12 *(Curvelet soft-thresholding). The discrete curvelet frames $\mathcal{D}_n = (\psi^n_{j,\ell,k}\colon (j,\ell,k) \in \Omega_n)$ defined by Eq. (4.10) are asymptotically stable. In particular, the following assertions hold true*:

(a) Distribution: *Let ϵ_n be a sequence of noise vectors in $\mathbb{R}^{I_n}$ with independent $N(0,\sigma^2)$-distributed entries. Then, for every $z \in \mathbb{R}$,*

$$\lim_{n\to\infty} \mathbf{P}\left\{\|\mathbf{C}_n \epsilon_n\|_\infty \leqslant \sigma\sqrt{2\log|\Omega_n|} + \sigma \frac{2z - \log\log|\Omega_n| - \log\pi}{2\sqrt{2\log|\Omega_n|}}\right\} = \exp\left(-e^{-z}\right).$$

(b) Confidence regions: *Let $\alpha_n \in (0,1)$ be a sequence of significance levels converging to some $\alpha \in [0,1)$ and let $T(\alpha_n, |\Omega_n|)$ denote the corresponding EVTs defined in (3.9). Then,*

$$\lim_{n\to\infty} \mathbf{P}\left\{\left\|\mathbf{C}_n(u_n - V_n)\right\|_\infty \leqslant T\left(\alpha_n, |\Omega_n|\right);\ u_n \in \mathbb{R}^{I_n}\right\} = 1 - \alpha.$$

(c) Smoothness: *Let $\hat{x}_n$ denote the soft-thresholding estimate using the extreme value threshold $T(\alpha_n, |\Omega_n|)$. Then, with any of the norms defined above, we have*

$$\liminf_{n\to\infty} \mathbf{P}\left\{\|\hat{x}_n\|_{p,q,r} \leqslant \|\mathbf{C}_n u_n\|_{p,q,r};\ u_n \in \mathbb{R}^{I_n}\right\} \geqslant 1 - \alpha.$$

Proof. All frame elements are normalized due to the chosen scaling. Moreover, as shown in [47, Proposition 6.1], the discrete curvelet frame $\mathcal{D}_n$ is faithful to an underlying infinite dimensional curvelet frame obtained by periodizing the curvelets on the continuous domain $\mathbb{R}^2$. This immediately yields condition (ii). Moreover, along the lines of [39] (which uses a slightly different curvelet system) one easily shows that the inner products satisfy

$$\left\langle \psi^n_{j,\ell,k}, \psi^n_{j',\ell',k'} \right\rangle \leqslant \rho < 1$$

for some constant $\rho < 1$ independent of n and all indices. This obviously implies condition (i). All claims in items (a)–(c) then follow from Theorems 1.3, 3.6 and 3.8 derived in the previous section. □

Acknowledgments

Both authors want to thank Daphne Böcker, Helene Dallmann, Housen Li and Andrea Kraijna for useful discussions and comments on the manuscript. The work of A.M. has been supported by the Volkswagen Foundation and the German Science Foundation (DFG) within the projects FOR 916, CRC 755, 803 and RTN 1023.

Appendix A. Extremes of normal random vectors

Let $(\Omega_n)_{n\in\mathbb{N}}$ be a sequence of finite index sets with monotonically increasing cardinalities $|\Omega_n|$ satisfying $\lim_{n\to\infty}|\Omega_n| = \infty$. Moreover, for every $n \in \mathbb{N}$, let $\xi_n := (\xi_n(\omega)\colon \omega \in \Omega_n)$ be given standardized normal random vectors, which means that $\xi_n(\omega) \sim N(0,1)$ for every $n \in \mathbb{N}$ and $\omega \in \Omega_n$. We are mainly interested in random vectors with dependent entries, in which case $\mathbf{Cov}(\xi_n(\omega), \xi_n(\omega')) \neq 0$ for at least some pairs $(\omega, \omega') \in \Omega_n^2$ with $\omega \neq \omega'$.

As the main result of this section we derive the asymptotic distribution of $\|\xi_n\|_\infty$ for a sequence ξ_n of dependent normal vectors whose covariances satisfy a certain summability condition (see Theorem A.8). Whereas similar results are known for $\max(\xi_n)$, to the best of our knowledge, such kind of results are new for $\|\xi_n\|_\infty$. Since $|\xi_n(\omega)|^2 \sim \chi^2$ is chi-squared distributed with one degree of freedom, our results can also be interpreted as new results for the asymptotic extreme value theory of dependent χ^2-distributed random vectors.

A.1. Maxima of normal vectors

We will start by reviewing and slightly refining the main results from statistical extreme value theory for maxima of normal vectors as we require them in this paper.

The most basic extreme value result deals with the case where the components of ξ_n are independent. In this case it is well known, that, after rescaling, $\max(\xi_n)$ converges to the Gumbel distribution as $n \to \infty$.

Proposition A.1. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of standardized normal random vectors in $\mathbb{R}^{\Omega_n}$ with independent entries. Then the maxima $\max(\xi_n)$ are of Gumbel type (see Definition 1.4) with normalization sequences $a(|\Omega_n|, n)$, $b(|\Omega_n|, n)$ defined by (3.3), (3.4).*

Proof. See [18, Theorem 1.5.3]. □

If the entries of ξ_n are dependent, then the result of Proposition A.1 does not necessarily hold true. There is, however, a simple and sufficient criterion on the covariances $\mathbf{Cov}(\xi_n(\omega), \xi_n(\omega'))$ of a sequence of dependent normal vectors such that the maxima still are of Gumbel type with the same normalization sequences. This criterion is an immediate consequence of the so called normal comparison lemma or Berman's inequality (see [18, Theorem 4.2.1]). For later purpose, where we study $\|\xi_n\|_\infty$ instead of $\max(\xi_n)$, we require a quite recent improvement of this important inequality which is due to Li and Shao [49]. The standard form of the normal comparison lemma [18, Theorem 4.2.1] has already been applied for redundant wavelet systems in [31,50], which however, only yields results for maxima of ξ_n without taking absolute values. We stress again, that taking absolute values slightly change the constants in contrast to relations (1.5).

Lemma A.2. *Let η_n, ξ_n be standardized normal random vectors in $\mathbb{R}^{\Omega_n}$, denote its covariances by $\kappa_{\eta_n}(\omega, \omega') := \mathbf{Cov}(\eta_n(\omega), \eta_n(\omega'))$, $\kappa_{\xi_n}(\omega, \omega') := \mathbf{Cov}(\xi_n(\omega), \xi_n(\omega'))$, and set $\rho_n(\omega, \omega') := \max\{|\kappa_{\eta_n}(\omega, \omega')|, |\kappa_{\xi_n}(\omega, \omega')|\}$. Then, for all $T_n \in \mathbb{R}$,*

$$\begin{aligned}&\mathbf{P}\{\max(\eta_n) \leqslant T_n\} - \mathbf{P}\{\max(\xi_n) \leqslant T_n\}\\ &\quad\leqslant \frac{1}{4\pi}\sum_{\omega\neq\omega'}\bigl(\arcsin(\kappa_{\eta_n}(\omega, \omega')) - \arcsin(\kappa_{\xi_n}(\omega, \omega'))\bigr)_+ \exp\left(\frac{-T_n^2}{1+\rho_n(\omega, \omega')}\right).\end{aligned}$$

Here $z_+ = \max\{z, 0\}$ denotes the positive part of some real number $z \in \mathbb{R}$ and arcsin *denotes the inverse mapping of* $\sin : [-\pi/2, \pi/2] \to [-1, 1]$.

Proof. See [49, Theorem 2.1]. □

In the special case where η_n has independent entries, Lemma A.2 has the following immediate consequence given in Lemma A.3. This allows to extend Proposition A.1 to certain sequences of dependent random vectors by comparing them with independent ones.

Lemma A.3. *Let η_n, ξ_n be standardized normal random vectors in $\mathbb{R}^{\Omega_n}$. Assume that the entries of η_n are independent, and let κ_n denote the covariance matrix of ξ_n defined by $\kappa_n(\omega, \omega') := \mathbf{Cov}(\xi_n(\omega), \xi_n(\omega'))$. Then, for all $T_n \in \mathbb{R}$,*

$$\bigl|\mathbf{P}\{\max(\eta_n) \leqslant T_n\} - \mathbf{P}\{\max(\xi_n) \leqslant T_n\}\bigr| \leqslant \frac{1}{8}\sum_{\omega\neq\omega'}\bigl|\kappa_n(\omega, \omega')\bigr| \exp\left(-\frac{T_n^2}{1+|\kappa_n(\omega, \omega')|}\right). \tag{A.1}$$

Proof. See [49, Corollary 2.2]. □

Lemmas A.2 and A.3 are significant improvements of the standard versions of the normal comparison lemma [18, Section 4] due to the absence of a singular factor $(1 - |\rho_n(\omega, \omega')|^2)^{-1/2}$ that is contained in earlier versions. It is in fact the absence of this singular term that we require for deriving an inequality similar to the one in Lemma A.3 that compares the distributions of $\|\eta_n\|_\infty$ and $\|\xi_n\|_\infty$ for two normal vectors η_n and ξ_n.

Theorem A.4. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of standardized normal random vectors in $\mathbb{R}^{\Omega_n}$ having covariance matrices $\kappa_n \in \mathbb{R}^{\Omega_n\times\Omega_n}$ satisfying*

$$\lim_{n\to\infty}\sum_{\omega\neq\omega'}\bigl|\kappa_n(\omega, \omega')\bigr|\left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{1/(1+|\kappa_n(\omega,\omega')|)} = 0. \tag{A.2}$$

Then, the maxima $\max(\xi_n)$ are of Gumbel type (see Definition 1.4) with normalization constants $a(N, |\Omega_n|)$, $b(N, |\Omega_n|)$ defined by (3.3), (3.4).

Proof. Fix some $z \in \mathbb{R}$ and define $T_n := a(N, |\Omega_n|)z + b(N, |\Omega_n|)$. Then, the definitions of the normalization sequences $a(N, |\Omega_n|)$ and $b(N, |\Omega_n|)$ imply that $T_n^2 = 2\log|\Omega_n| - \log\log|\Omega_n| + \mathcal{O}(1)$ as $n \to \infty$. Hence there is some constant $C > 0$ and some index $n_0 \in \mathbb{N}$, such that for all $n \geqslant n_0$, we have

$$\exp\left(-\frac{T_n^2}{1+|\kappa_n(\omega,\omega')|}\right) \leqslant C\exp\left(-\frac{\log(|\Omega_n|^2/\log|\Omega_n|)}{1+|\kappa_n(\omega,\omega')|}\right) = C\left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{1/(1+|\kappa_n(\omega,\omega')|)}.$$

Now let Eq. (A.2) be satisfied and let $(\eta_n)_{n\in\mathbb{N}}$ be a sequence of standardized normal vectors with independent entries. Then, the triangle inequality, Lemma A.3, and the estimate just established imply

$$\left|\mathbf{P}\{\max(\xi_n) \leqslant T_n\}\right| \leqslant \left|\mathbf{P}\{\max(\eta_n) \leqslant T_n\}\right| + \frac{C}{8}\sum_{\omega\neq\omega'}\left|\kappa_n(\omega,\omega')\right|\left(\frac{\log|\Omega_n|}{|\Omega_n|^2}\right)^{1/(1+|\kappa_n(\omega,\omega')|)}.$$

Hence the claim follows from Proposition A.1 and assumption (A.2). □

A.2. Maxima of absolute values

In the following we derive results similar to Proposition A.1 and Theorem A.4 for $\|\xi_n\|_\infty$ in place of $\max(\xi_n)$. The first auxiliary result, Proposition A.5, deals with the independent case. It is easy to establish but nevertheless seems to be much less known than the corresponding result in the normal case. We include a short proof based on the known extreme value distribution of independent χ^2-distributed random variables. The second and main result in this section, Theorem A.8, deals with the dependent case. It is a new contribution and based on a novel inequality for comparing the distributions of $\|\eta_n\|_\infty$ and $\|\xi_n\|_\infty$ (given in Lemma A.7).

Proposition A.5. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of standardized normal vectors in $\mathbb{R}^{\Omega_n}$ having independent entries. Then $\|\xi_n\|_\infty$ is of Gumbel type (see Definition 1.4) with normalization sequences $a(\chi, |\Omega_n|)$, $b(\chi, |\Omega_n|)$ defined by (1.9), (1.10).*

Proof. Since $\xi_n(\omega)$ is standard normally distributed for any $\omega \in \Omega_n$, the random variables $|\xi_n(\omega)|^2$ are χ^2-distributed with one degree of freedom. The χ^2-distribution is in turn a member of the family of Gamma distributions $F_{\beta,\gamma}$ corresponding to $\beta = \gamma = 1/2$. The asymptotic extreme value distribution of the Gamma distribution $F_{\beta,\gamma}$ is known (see [51, p. 156]) and implies

$$\lim_{n\to\infty}\mathbf{P}\left\{\|\xi_n\|_\infty^2 \leqslant 2z + 2\log|\Omega_n| - \log\log|\Omega_n| - \log\pi\right\} = \exp\left(-e^{-z}\right). \tag{A.3}$$

Moreover, a Taylor series approximation shows

$$\sqrt{2z + 2\log|\Omega_n| - \log\log|\Omega_n| - \log\pi} = a\left(\chi, |\Omega_n|\right)z + b\left(\chi, |\Omega_n|\right) + o\left(a\left(\chi, |\Omega_n|\right)\right) \quad \text{as } n \to \infty. \tag{A.4}$$

Any $o(a(\chi, |\Omega_n|))$ term can be omitted when computing extreme value distributions (see [18, Theorem 1.2.3]), and hence Eqs. (A.3) and (A.4) imply the desired result. □

Remark A.6. The sequence $b(\chi, |\Omega_n|)$ used for normalizing the maximum $\|\xi_n\|_\infty$ in Proposition A.5 is different from the sequence $b(N, |\Omega_n|)$ used for the normalization of $\max(\xi_n)$ in Proposition A.1. Indeed, as easily verified,

$$b\left(N, 2|\Omega_n|\right) = b\left(\chi, |\Omega_n|\right) + o\left(a\left(N, 2|\Omega_n|\right)\right).$$

Again, the $o(a(N, 2|\Omega_n|))$ term can be omitted in the extreme value distribution and hence $\|\xi_n\|_\infty$ behaves equal to the maximum of $2|\Omega_n|$ (opposed to $|\Omega_n|$) independent standard normally distributed random variables. Using different arguments, this has already been observed in [11, Section 8.3].

If the entries of ξ_n are not independent, then the result of Proposition A.5 does not necessarily hold true. If, however, the correlations of ξ_n are sufficiently small, then, as in the normal case, we will show that the same Gumbel law still holds. This result follows again from a comparison inequality, now between the distributions of $\|\xi_n\|_\infty$ and $\|\eta_n\|_\infty$ with some reference normal vector η_n, to be derived in the following Lemma A.7. For the sake of simplicity we assume that the vector η_n has independent entries; in an analogous manner a similar result could be derived for comparing two dependent random vectors.

Lemma A.7. *Let η_n, ξ_n be standardized normal random vectors in $\mathbb{R}^{\Omega_n}$. Assume that the entries of η_n are independent and denote by $\kappa_n \in \mathbb{R}^{\Omega_n\times\Omega_n}$ the covariance matrix of ξ_n, having entries $\kappa_n(\omega,\omega') := \mathbf{Cov}(\xi_n(\omega), \xi_n(\omega'))$. Then, for all $T_n \in \mathbb{R}$,*

$$\left|\mathbf{P}\{\|\eta_n\|_\infty \leqslant T_n\} - \mathbf{P}\{\|\xi_n\|_\infty \leqslant T_n\}\right| \leqslant \frac{1}{4}\sum_{\omega\neq\omega'}\left|\kappa_n(\omega,\omega')\right|\exp\left(-\frac{T_n^2}{1+|\kappa_n(\omega,\omega')|}\right). \tag{A.5}$$

Proof. The proof uses the normal comparison lemma (Lemma A.2) of Li and Shao applied to the strongly dependent random vectors $Y_n := (\eta_n, -\eta_n)$ and $X_n := (\xi_n, -\xi_n)$ in place of η_n and ξ_n. To that end, we first note that obviously $\{|\xi_n| \leqslant T_n\} = \{X_n < T_n\}$ and $\{|\eta_n| \leqslant T_n\} = \{Y_n < T_n\}$. Moreover, the covariance matrices of Y_n and X_n are block matrices of the form

$$\mathbf{Cov}(Y_n) = \begin{pmatrix} \mathbf{I}_n & -\mathbf{I}_n \\ -\mathbf{I}_n & \mathbf{I}_n \end{pmatrix} \quad \text{and} \quad \mathbf{Cov}(X_n) = \begin{pmatrix} \kappa_n & -\kappa_n \\ -\kappa_n & \kappa_n \end{pmatrix},$$

where $\kappa_n = \mathbf{Cov}(\xi_n)$ denotes the covariance matrix of ξ_n and $\mathbf{I}_n = \mathbf{Cov}(Y_n)$ is the identity matrix in $\mathbb{R}^{\Omega_n \times \Omega_n}$. Now applying Lemma A.2 with Y_n and X_n in place of η_n and ξ_n yields

$$\begin{aligned}
&\mathbf{P}\{\|\eta_n\|_\infty \leqslant T_n\} - \mathbf{P}\{\|\xi_n\|_\infty \leqslant T_n\} \\
&\quad = \mathbf{P}\{\max(\eta_n, -\eta_n) \leqslant T_n\} - \mathbf{P}\{\max(\xi_n, -\xi_n) \leqslant T_n\} \\
&\quad \leqslant \frac{1}{2\pi} \sum_{\omega \neq \omega'} \bigl(\bigl(-\arcsin\bigl(\kappa_n(\omega, \omega')\bigr)\bigr)_+ + \bigl(\arcsin\bigl(\kappa_n(\omega, \omega')\bigr)\bigr)_+\bigr) \exp\biggl(-\frac{T_n^2}{1 + |\kappa_n(\omega, \omega')|}\biggr) \\
&\quad = \frac{1}{2\pi} \sum_{\omega \neq \omega'} \bigl|\arcsin\bigl(-\kappa_n(\omega, \omega')\bigr)\bigr| \exp\biggl(-\frac{T_n^2}{1 + |\kappa_n(\omega, \omega')|}\biggr).
\end{aligned}$$

Here for the first estimate we used that the two sums over the diagonal blocks give the same value, that the same is the case for the two off-diagonal blocks, that all terms having $\omega = \omega'$ cancel and that $\mathbf{Cov}(X_n)(\omega, \omega') = 0$ for $\omega \neq \omega'$. Interchanging the roles of ξ_n and η_n yields the same estimate for $\mathbf{P}\{\|\xi_n\|_\infty \leqslant T_n\} - \mathbf{P}\{\|\eta_n\|_\infty \leqslant T_n\}$ and hence implies

$$\bigl|\mathbf{P}\{\|\eta_n\|_\infty \leqslant T_n\} - \mathbf{P}\{\|\xi_n\|_\infty \leqslant T_n\}\bigr| \leqslant \frac{1}{2\pi} \sum_{\omega \neq \omega'} \bigl|\arcsin\bigl(\kappa_n(\omega, \omega')\bigr)\bigr| \exp\biggl(-\frac{T_n^2}{1 + |\kappa_n(\omega, \omega')|}\biggr). \tag{A.6}$$

Finally, the estimate $|\arcsin y| \leqslant |y| \cdot \pi/2$ for $y \in [-1, 1]$ and inequality (A.6) imply the claimed inequality (A.5). □

The following theorem is the main result of this section and the key for most results established in this paper.

Theorem A.8. *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence of standardized normal vectors in $\mathbb{R}^{\Omega_n}$ having covariance matrices $\kappa_n \in \mathbb{R}^{\Omega_n \times \Omega_n}$ satisfying Eq. (A.2). Then $\|\xi_n\|_\infty$ is of Gumbel type (see Definition 1.4) with normalization constants $a(\chi, |\Omega_n|)$, $b(\chi, |\Omega_n|)$ defined by (1.9), (1.10).*

Proof. This is analogous to the proof of Theorem A.4. Instead of Proposition A.1 and Lemma A.3 one now uses Proposition A.5 and Lemma A.7. □

Eq. (A.2) provides a sufficient condition for the extreme value results of Theorems A.4 and A.8 to hold. However, given a sequence $(\xi_n)_{n\in\mathbb{N}}$ of normal vectors with covariance matrices κ_n, it is not completely obvious whether or not (A.2) is satisfied. In Section 3 we verified that (A.2) indeed holds in the case where $\xi_n = \langle \phi_\omega^n, \epsilon_n \rangle$ are coefficients of standardized normal random vectors ϵ_n having independent entries with respect to an asymptotically stable family of frames $(\phi_\omega^n\colon \omega \in \Omega_n)$.

Occasionally we will make use of the following classical result due to Sidak [52] for bounding the maximum of the magnitudes of dependent random vectors by the maximum of the magnitudes of independent ones.

Lemma A.9 *(Sidak's inequality). Let η_n, ξ_n be standardized normal random vectors in $\mathbb{R}^{\Omega_n}$ and assume that the entries of η_n are independent. Then,*

$$(\forall T \in \mathbb{R}) \quad \mathbf{P}\{\|\xi_n\|_\infty \leqslant T\} \geqslant \mathbf{P}\{\|\eta_n\|_\infty \leqslant T\}. \tag{A.7}$$

Proof. See [52, Corollary 1]. □

Note that a similar result also holds for the maxima without the absolute values, which bounds the probability $\mathbf{P}\{\max(\xi_n) \leqslant T\}$ of dependent standardized normal vectors from below by the probability $\mathbf{P}\{\max(\eta_n) \leqslant T\}$ of independent ones. This one-sided estimate, however, requires the covariances of ξ_n being nonnegative. It is known as Slepian's lemma and has first been derived in [53]. Interestingly, Slepian's lemma immediately follows from the normal comparison lemma (Lemma A.2), whereas this seems not to be the case for Sidak's two sided inequality.

Appendix B. Remaining proofs

B.1. Proof of Proposition 3.10

As already noted in [12, p. 558], for the universal thresholds $\sigma\sqrt{2\log|\Omega_n|}$ this result easily follows by adapting the original proof of [8] (see also [11, Section 8.3] and [12, Theorem 11.7]) from the orthonormal case to the frame case. Indeed, as shown below a similar proof can be made for the extreme value thresholds $T_n = T(\alpha_n, |\Omega_n|)$ defined by Eq. (3.9).

After rescaling we may assume without loss of generality that $\sigma = 1$. Recall that the dual frame $(\tilde{\phi}^n_\omega\colon \omega \in \Omega_n)$ has upper frame bound $1/a_n$, that $\boldsymbol{\Phi}_n^+\boldsymbol{\Phi}_n = \mathrm{Id}$ is the identity on $\mathbb{R}^{I_n}$, and that $\boldsymbol{\Phi}_n\boldsymbol{\Phi}_n^+ = P_{\mathrm{Ran}(\boldsymbol{\Phi}_n)}$ equals the orthogonal projection onto the range $\mathrm{Ran}(\boldsymbol{\Phi}_n) \subset \mathbb{R}^{\Omega_n}$ of the analysis operator $\boldsymbol{\Phi}_n : \mathbb{R}^{I_n} \to \mathbb{R}^{\Omega_n}$. Moreover, we define the parameter $x_n = \boldsymbol{\Phi}_n u_n$ and the data $Y_n = \boldsymbol{\Phi}_n V_n$ as in (2.4). Then we can estimate

$$\begin{aligned}\mathbf{E}\big(\big\|u_n - \boldsymbol{\Phi}_n^+ \circ \mathbf{S}\big(\boldsymbol{\Phi}_n V_n, T\big(\alpha_n, |\Omega_n|\big)\big)\big\|^2\big) &= \mathbf{E}\big(\big\|\boldsymbol{\Phi}_n^+ x_n - \boldsymbol{\Phi}_n^+ \circ \mathbf{S}\big(\boldsymbol{\Phi}_n V_n, T\big(\alpha_n, |\Omega_n|\big)\big)\big\|^2\big)\\ &= \mathbf{E}\big(\big\|\boldsymbol{\Phi}_n^+\boldsymbol{\Phi}_n\boldsymbol{\Phi}_n^+\big(x_n - \mathbf{S}\big(Y_n, T\big(\alpha_n, |\Omega_n|\big)\big)\big)\big\|^2\big)\\ &\leqslant \frac{1}{a_n}\mathbf{E}\big(\big\|P_{\mathrm{Ran}(\boldsymbol{\Phi}_n)}\big(x_n - \mathbf{S}\big(Y_n, T\big(\alpha_n, |\Omega_n|\big)\big)\big)\big\|^2\big)\\ &= \frac{1}{a_n}\sum_{\omega\in\Omega_n}\mathbf{E}\big(\big|x_n(\omega) - S\big(Y_n(\omega), T\big(\alpha_n, |\Omega_n|\big)\big)\big|^2\big).\end{aligned}$$

Now we can proceed similar to [8] (see also [11,12]) to estimate the mean square errors $\mathbf{E}(|x_n(\omega) - S(Y_n(\omega), T(\alpha_n, |\Omega_n|))|^2)$ of one dimensional soft-thresholding.

To that end we use the risk estimate of [11, Section 2.7] for one dimensional soft-thresholding, which states the following: If $y \sim N(\mu, 1)$ is a normal random variable with mean $\mu \in \mathbb{R}$ and unit variance, then

$$(\forall T > 0)\quad \mathbf{E}\big(\big|\mu - S(y, T)\big|^2\big) \leqslant e^{-T^2/2} + \min\big\{1 + T^2, \mu^2\big\}. \tag{B.1}$$

For our purpose we apply the risk estimate (B.1) with threshold $T = T(\alpha_n, |\Omega_n|)$. The definition of the threshold $T(\alpha_n, |\Omega_n|)$ in (3.9) immediately yields the estimate

$$\frac{T(\alpha_n, |\Omega_n|)^2}{2} \geqslant \log|\Omega_n| - \log\log\big(1/(1-\alpha_n)\big) - \frac{\log\log|\Omega_n| + \log\pi}{2}.$$

Inserting these estimates in (B.1) applied with the random variables $y = Y_n(\omega)$ having mean values $\mu = x_n(\omega)$ and using the assumption $T(\alpha_n, |\Omega_n|) \leqslant \sqrt{2\log|\Omega_n|}$ yields

$$\mathbf{E}\big(\big|x_n(\omega) - S\big(Y_n(\omega), T\big(\alpha_n, |\Omega_n|\big)\big)\big|^2\big) \leqslant \frac{\log(1/(1-\alpha_n))\sqrt{\pi\log|\Omega_n|}}{|\Omega_n|} + \big(1 + 2\log|\Omega_n|\big)\min\big\{1, \big|x_n(\omega)\big|^2\big\}.$$

Finally, summing over all $\omega \in \Omega_n$ shows (3.15).

B.2. Proof of Theorem 4.9

Let $\eta = (\eta(t)\colon t \in [0,1])$ denote a white noise process on $[0,1]$ and consider the periodic continuous domain wavelets $\psi_{j,b}(t) = 2^{j/2}\psi(2^j(t-b))$. We then define the random vectors X_n as inner products

$$(\forall j = 0, \ldots, \log n - 1)\big(\forall \ell = 0, \ldots, 2^j n - 1\big)\quad X_n(j, \ell) := \langle\psi_{j,2^j\ell/n}, \eta\rangle.$$

Hence the random variables $X_n(j,\ell)$ are coefficients of the white noise process η with respect to a discrete wavelet transform, that is oversampled by factor n at every scale. Comparing this with the definition of the translation invariant wavelet transform we see that the translation invariant wavelet coefficients $\mathbf{W}_{n,n}\epsilon_n$ are a subset of the elements of X_n. Hence we have

$$(\forall T > 0)\quad \mathbf{P}\big\{\|\mathbf{W}_{n,n}\epsilon_n\|_\infty \leqslant T\big\} \geqslant \mathbf{P}\big\{\|X_n\|_\infty \leqslant T\big\}. \tag{B.2}$$

We proceed by computing the correlations of $X_n(j,\ell)$ for some fixed scale index. Since η is a white noise process, the definition of X_n and some elementary manipulations shows that, for all $j \in \{0, \ldots, \log n - 1\}$ and all indices $\ell, \ell' \in \{0, \ldots, 2^j n - 1\}$, we have

$$\begin{aligned}\mathbf{Cov}\big(X_n(j,\ell), X_n\big(j,\ell'\big)\big) &= \langle\psi_{j,\ell}, \psi_{j,\ell'}\rangle = 2^j\int_{\mathbb{R}}\psi\big(2^j t - \ell/n\big)\psi\big(2^j t - \ell'/n\big)\,dt\\ &= 2^j\int_{\mathbb{R}}\psi\big(2^j t\big)\psi\big(2^j t - \big(\ell' - \ell\big)/n\big)\,dt = \int_{\mathbb{R}}\psi(-t)\psi\big(\big(\ell - \ell'\big)/n - t\big)\,dt = (\bar{\psi} * \psi)\big(\big(\ell - \ell'\big)/n\big).\end{aligned}$$

Next we construct a random vector Y_n with the same index set and pointwise smaller correlations. To that end, for every given j, we group the index set $\{0,\dots,2^j n-1\}$ into 2^j blocks $B_{j,k}=\{kn,\dots,(k+1)n-1\}$ for any $k\in\{0,\dots,2^j-1\}$. We denote $\kappa:=\bar\psi*\psi$ and define the matrix

$$\bar\kappa_n\big((j,\ell),(j',\ell')\big):=\begin{cases}\kappa((\ell-\ell')/n) & \text{if } j=j' \text{ and } (\ell,\ell')\in\bigcup_k B_{j,k}\times B_{j,k}\\ 0 & \text{otherwise.}\end{cases}$$

Hence we have $\bar\kappa_n((j,\ell),(j',\ell'))=\mathbf{Cov}(X_n(j,\ell),X_n(j',\ell'))$ if $j=j'$ and the indices ℓ,ℓ' are in the same block $B_{j,k}$, and the correlations of $\bar\kappa_n$ are zero otherwise. Moreover $\bar\kappa_n$ is obviously symmetric and positive semi-definite and hence there exists a standardized normal random vector Y_n whose covariance matrix is given by $\bar\kappa_n$. By construction of $\bar\kappa_n$, the covariances $|\mathbf{Cov}(X_n(j,\ell),X_n(j,\ell'))|$ pointwise dominate the covariances $|\bar\kappa_n((j,\ell),(j',\ell'))|$. Hence, Lemma A.9 implies

$$(\forall T>0)\quad \mathbf{P}\big\{\|X_n\|_\infty\leqslant T\big\}\geqslant\mathbf{P}\big\{\|Y_n\|_\infty\leqslant T\big\}. \tag{B.3}$$

Inspecting Eqs. (B.2) and (B.3) shows that it remains to compute the asymptotic distribution of $\|Y_n\|_\infty$.

To that end recall that $\mathbf{Cov}(X_n(j,\ell),X_n(j,\ell'))=\kappa((\ell-\ell')/n)$ are densely sampled values of the autocorrelation function of the mother wavelet. This in particular implies that any block in Y_n has the same distribution. Moreover, due to the independence of the blocks this yields

$$\begin{aligned}\mathbf{P}\big\{\|Y_n\|_\infty\leqslant T\big\}&=\mathbf{P}\big\{\max\big|Y_n(0,\ell)\colon \ell=0,\dots,n-1\big|\leqslant T\big\}^n\\&=\big(1-\mathbf{P}\big\{\max\big|Y_n(0,\ell)\colon \ell=0,\dots,n-1\big|>T\big\}\big)^n\\&=\big(1-\mathbf{P}\big\{\max\big|\langle\psi_{0,\ell/n},\eta\rangle\big|\colon \ell=0,\dots,n-1\big\}>T\big)^n\\&=\big(1-\mathbf{P}\big\{\max\big|X(\ell/n)\big|\colon \ell=0,\dots,n-1\big\}>T\big)^n.\end{aligned}$$

Here $X=\{X(t)\colon t\in[0,1]\}$ is defined by $X(t):=\langle\psi_{0,t},\eta\rangle$. One easily verifies that X is a mean square differentiable normal process having covariance function $\kappa(t)$. Moreover the vector $Y_n(0,\ell)=X(\ell/n)$ consist of n equidistant values of that process inside the unit interval. Hence for any sequence of thresholds T_n that tends to infinity as $n\to\infty$ in a sufficiently slowly manner, one has the asymptotic relations (which follow from standard result of continuous extreme value theory [18])

$$\begin{aligned}&\mathbf{P}\big\{\max\big\{\big|X(\ell/n)\big|\colon \ell=0,\dots,n-1\big\}>T_n\big\}\sim\mathbf{P}\big\{\max\big\{\big|X(t)\big|\colon t\in[0,1]\big\}>T_n\big\},\\&\mathbf{P}\big\{\max\big\{\big|X(t)\big|\colon t\in[0,1]\big\}>T_n\big\}\sim 2\mathbf{P}\big\{\max\big\{X(t)\colon t\in[0,1]\big\}>T_n\big\},\\&\mathbf{P}\big\{\max\big\{X(t)\colon t\in[0,1]\big\}>T_n\big\}\sim c/(2\pi)\exp\big(-T_n^2/2\big).\end{aligned}$$

Now fix any $z\in\mathbb{R}$ and define the sequence $T_n:=(2(\log n+z+2\log(c/\pi)))^{1/2}$. Then the definition of T_n immediately yields $\exp(-T_n^2/2)=\pi/(cn)\exp(-z)$. Consequently, by collecting the above estimates, we have

$$\lim_{n\to\infty}\mathbf{P}\big\{\|Y_n\|_\infty\leqslant T_n\big\}=\lim_{n\to\infty}\left(1-\frac{c}{\pi}e^{-T_n^2/2}\right)^n=\lim_{n\to\infty}\left(1-\frac{e^{-z}}{n}\right)^n=\exp\big(-e^{-z}\big).$$

Finally, a simple Taylor series approximation of the square root shows the asymptotic relation

$$T_n=\sqrt{2\log n}+\frac{x+\log(c/\pi)}{\sqrt{2\log n}}+o(1/\sqrt{2\log n}).$$

Recalling, for the last time, that $o(1/a_n)$ terms can be omitted in the rescaling of extreme value distributions finally shows

$$\mathbf{P}\left\{\|Y_n\|_\infty\leqslant\sqrt{2\log n}+\frac{x+\log(c/\pi)}{\sqrt{2\log n}}\right\}\to\exp\big(-e^{-z}\big),$$

and concludes the proof.

References

[1] T. Cai, Adaptive wavelet estimation: A block thresholding and oracle inequality approach, Ann. Statist. 27 (1999) 898–924.
[2] T. Cai, H. Zhou, A data-driven block thresholding approach to wavelet estimation, Ann. Statist. 37 (2009) 569–595.
[3] C. Chesneau, J. Fadili, J.-L. Starck, Stein block thresholding for wavelet-based image deconvolution, Electron. J. Stat. 4 (2010) 415–435.
[4] P. Hall, S. Penev, G. Kerkyacharian, D. Picard, Numerical performance of block thresholded wavelet estimators, Statist. Comput. 7 (1997) 115–124.
[5] A. Antoniadis, J. Bigot, T. Sapatinas, Wavelet estimators in nonparametric regression: A comparative simulation study, J. Stat. Softw. 6 (2001) 1–83.
[6] R.R. Coifman, D.L. Donoho, Translation-invariant de-noising, in: A. Antoniadis, G. Oppenheim (Eds.), Lecture Notes in Statist., vol. 103, Springer, 1995.
[7] D.L. Donoho, De-noising by soft-thresholding, IEEE Trans. Inform. Theory 41 (1995) 613–627.
[8] D.L. Donoho, I.M. Johnstone, Ideal spatial adaptation by wavelet shrinkage, Biometrika 81 (1994) 425–455.
[9] D.L. Donoho, I.M. Johnstone, G.B. Kerkyacharian, D. Picard, Density estimation by wavelet thresholding, Ann. Statist. 24 (1996) 508–539.
[10] W. Härdle, G. Kerkyacharian, D. Picard, A. Tsybakov, Wavelets, Approximation, and Statistical Applications, Lecture Notes in Statist., vol. 129, Springer, New York, 1998.

[11] I.M. Johnstone, Gaussian estimation: Sequence and wavelet models, Draft of a monograph, 2011.
[12] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, third ed., Elsevier/Academic Press, Amsterdam, 2009.
[13] J.-L. Starck, F. Murtagh, J.M. Fadili, Sparse Image and Signal Processing, Cambridge University Press, Cambridge, 2010.
[14] A. Cohen, Numerical Analysis of Wavelet Methods, Stud. Math. Appl., vol. 32, North-Holland, Amsterdam, 2003.
[15] D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, J. Amer. Statist. Assoc. 90 (1995) 1200–1224.
[16] I.M. Johnstone, B.W. Silverman, Wavelet threshold estimators for data with correlated noise, J. R. Stat. Soc. Ser. B Stat. Methodol. 59 (1997) 319–351.
[17] L. de Haan, A. Ferreira, Extreme Value Theory, Springer Ser. Oper. Res. Financ. Eng., Springer, New York, 2006.
[18] M.R. Leadbetter, G. Lindgren, H. Rootzén, Extremes and Related Properties of Random Sequences and Processes, Springer Ser. Statist., Springer, New York, 1983.
[19] A. Antoniadis, J. Fan, Regularization of wavelet approximations, J. Amer. Statist. Assoc. 96 (2001) 939–967, with discussion and a rejoinder by the authors.
[20] L.D. Brown, L. Brown, Variance estimation in nonparametric regression via the difference sequence method, Ann. Statist. 35 (2007) 2219–2232.
[21] H. Dette, A. Munk, T. Wagner, Estimating the variance in nonparametric regression—what is a reasonable choice?, J. R. Stat. Soc. Ser. B Stat. Methodol. 60 (1998) 751–764.
[22] P. Hall, J.W. Kay, D.M. Titterington, Asymptotically optimal difference-based estimation of variance in nonparametric regression, Biometrika 77 (1990) 521–528.
[23] P. Hall, J.S. Marron, On variance estimation in nonparametric regression, Biometrika 77 (1990) 415–419.
[24] A. Munk, N. Bissantz, T. Wagner, G. Freitag, On difference-based variance estimation in nonparametric regression when the covariate is high dimensional, J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2005) 19–41.
[25] O. Christensen, Frames and Bases, Appl. Numer. Harmon. Anal., Birkhäuser Boston Inc., Boston, MA, 2008.
[26] L. Breiman, Better subset regression using the nonnegative garrote, Technometrics 37 (1995) 373–384.
[27] H. Gao, Wavelet shrinkage denoising using the non-negative garrote, J. Comput. Graph. Statist. 7 (1998) 469–488.
[28] C. Chesneau, J. Fadili, J.-L. Starck, Stein block thresholding for image denoising, Appl. Comput. Harmon. Anal. 28 (2010) 67–88.
[29] P. Hall, G. Kerkyacharian, D. Picard, On the minimax optimality of block thresholded wavelet estimators, Statist. Sinica 9 (1999) 33–49.
[30] G. Kerkyacharian, R. Nickl, D. Picard, Concentration inequalities and confidence bands for needlet density estimators on compact homogeneous manifolds, Probab. Theory Related Fields 153 (1–2) (June 2012) 363–404.
[31] K. Berkner, R.O. Wells Jr., Smoothness estimates for soft-threshold denoising via translation-invariant wavelet transforms, Appl. Comput. Harmon. Anal. 12 (2002) 1–24.
[32] A. Cohen, I. Daubechies, J.-C. Feauveau, Biorthogonal bases of compactly supported wavelets, Comm. Pure Appl. Math. 45 (1992) 485–560.
[33] I. Daubechies, Orthonormal bases of compactly supported wavelets, Comm. Pure Appl. Math. 41 (1988) 909–996.
[34] I. Daubechies, Ten Lectures on Wavelets, SIAM, Philadelphia, PA, 1992.
[35] M. Lang, H. Guo, J.E. Odegard, C.S. Burrus, R.O. Wells Jr., Noise reduction using an undecimated discrete wavelet transform, IEEE Signal Process. Lett. 3 (1996) 10–12.
[36] G.P. Nason, B.W. Silverman, The stationary wavelet transform and some statistical applications, in: A. Antoniadis, G. Oppenheim (Eds.), Lecture Notes in Statist., vol. 103, Springer, 1995.
[37] J.-C. Pesquet, H. Krim, H. Carfantan, Time-invariant orthonormal wavelet representations, IEEE Trans. Signal Process. 44 (1996) 1964–1970.
[38] E.J. Candès, D.L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, Comm. Pure Appl. Math. 57 (2004) 219–266.
[39] E.J. Candès, D.L. Donoho, Continuous curvelet transform. I. Resolution of the wavefront set, Appl. Comput. Harmon. Anal. 19 (2005) 162–197.
[40] E.J. Candès, D.L. Donoho, Continuous curvelet transform. II. Discretization and frames, Appl. Comput. Harmon. Anal. 19 (2005) 198–222.
[41] K. Guo, G. Kutyniok, D. Labate, Sparse multidimensional representations using anisotropic dilation and shear operators, in: Wavelets and Splines, Athens 2005, in: Mod. Methods Math., Nashboro Press, Brentwood, TN, 2006, pp. 189–201.
[42] K. Guo, D. Labate, W.-Q. Lim, G. Weiss, E. Wilson, Wavelets with composite dilations and their MRA properties, Appl. Comput. Harmon. Anal. 20 (2006) 202–236.
[43] D. Labate, W.-Q. Lim, G. Kutyniok, G. Weiss, Sparse multidimensional representation using shearlets, Proc. SPIE 5914 (2005) 59140U.
[44] K. Guo, D. Labate, Optimally sparse multidimensional representation using shearlets, SIAM J. Math. Anal. 39 (2007) 298–318.
[45] M.N. Do, M. Vetterli, Contourlets, in: P.M.C.K. Chui, L. Wuytack (Eds.), Beyond Wavelets, in: Stud. Comput. Math., vol. 10, Elsevier, 2003, pp. 83–105.
[46] M.N. Do, M. Vetterli, The contourlet transform: an efficient directional multiresolution image representation, IEEE Trans. Image Process. 14 (2005) 2091–2106.
[47] E. Candès, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, Multiscale Model. Simul. 5 (2006) 861–899.
[48] L. Borup, M. Nielsen, Frame decomposition of decomposition spaces, J. Fourier Anal. Appl. 13 (2007) 39–70.
[49] W.V. Li, Q.-M. Shao, A normal comparison inequality and its applications, Probab. Theory Related Fields 122 (2002) 494–508.
[50] A.K. Ng, T.S. Koh, K. Puvanendran, U.R. Abeyratne, Snore signal enhancement and activity detection via translation-invariant wavelet transform, IEEE Trans. Biomed. Eng. 55 (2008) 2332–2342.
[51] P. Embrechts, C. Kluppelberg, T. Mikosch, Modelling Extremal Events for Insurance and Finance, Springer, Berlin, 1997.
[52] Z. Šidák, Rectangular confidence regions for the means of multivariate normal distributions, J. Amer. Statist. Assoc. 62 (1967) 626–633.
[53] D. Slepian, The one-sided barrier problem for Gaussian noise, Bell System Tech. J. 41 (1962) 463–501.