Haar Wavelets, Gradients and Approximate Total Variation Regularization

Tomas Sauer¹,²,³† and Andreas Michael Stock³‡

¹Chair of Mathematical Image Processing, University of Passau, Innstr. 43, Passau, 94032, Germany.
²Development Center X-ray Technology, Fraunhofer Institute of Integrated Circuits, Innstr. 43, Passau, 94032, Germany.
³FORWISS, University of Passau, Innstr. 43, Passau, 94032, Germany.

*Corresponding author(s). E-mail(s): stock@forwiss.uni-passau.de; Contributing authors: tomas.sauer@uni-passau.de; †All authors have contributed equally.

Abstract

Image denoising by means of total variation (TV) regularization is still a standard procedure. For very large images, especially three-dimensional voxel datasets, however, this can be computationally infeasible. We show how this TV regularization can be approximately performed even in arbitrary dimensions by applying appropriate shrinkage to selected and properly weighted Haar wavelet coefficients, all of which depends even on the dimensionality of the data. Our approach acts entirely on the wavelet coefficients which represent the compressed image, and is therefore suited for the application on large three-dimensional images represented in the Haar wavelet basis, e.g., volumes from computed tomography.

Keywords: Haar wavelets, shrinkage, total variation, denoising

1 Introduction

Modern imaging techniques produce larger and larger images that need to be processed and analyzed. This is particularly evident in industrial computed tomography (CT) where huge datasets of one terabyte and more are generated on an almost routine basis by either scanning very large objects [1] or working with a very high resolution. Even on modern consumer hardware, such big data cannot be handled without substantial compression. A natural approach for that purpose is to use well-established tensor product wavelet methods [2]. For two-dimensional images, wavelets have become part of the JPEG2000 standard [3], and they have also been applied successfully for the compression of tomographic volume data recently [4, 5].

To illustrate the type of objects that we are concerned with, we refer to the CT scan of a Peruvian mummy, located at the Linden-Museum in Stuttgart, that was performed at the Fraunhofer IIS Development Center X-ray Technology (EZRT), see Fig. 1 and 2. The size of the dataset is roughly 970 GB and in order to handle and visualize it on consumer hardware, it has been reduced to about 30 GB by the wavelet compression method from [5]. While
high-resolution regions of interest (ROIs) of the scan can be obtained on-the-fly, the full resolution image is not tractable with reasonable effort and this necessitates algorithms that work exclusively on the wavelet coefficients.

The simplest choice for selecting a wavelet is the Haar wavelet [6] and its associated multiresolution analysis generated by characteristic functions of the unit interval and their dyadic refinement properties. Haar wavelets have the advantages of being compactly supported, orthonormal and symmetric, and they are the only wavelet system with these properties [7]. Moreover, Haar wavelets have minimal support among all discrete multiresolution systems and thus provide optimal localization. Fortunately, they can also be implemented in a very efficient way, allowing for fast decomposition and on-the-fly reconstruction of images. This by itself makes them interesting and useful despite their well-known drawbacks like a lack of smoothness and vanishing moments.

In this paper we show that and how Haar wavelets can be used for gradient estimation and an approximate total variation (TV) denoising directly on the wavelet coefficients without any need to reconstruct the full image. This is a fundamental requirement for handling very large images, especially in 3D. By this direct and computationally cheap manipulation of wavelet coefficients, TV denoising can even be integrated into an almost real-time visualization pipeline.

Moreover, the proofs will show that the approximation of the TV norm is a particular property of tensor product Haar wavelets: the fact that they have only one vanishing moment is responsible for their ability to approximate the gradient on a grid. Their symmetry and analytic expression allow us to determine the explicit level-dependent renormalization coefficients that guarantee that direction and length of the gradients are met accurately. Altogether, this leads to a novel shrinkage scheme, where vectors consisting of a subset of the wavelet coefficients are thresholded with respect to their length after a proper dimension-dependent and level-dependent renormalization.

The paper is organized as follows: after briefly discussing some related work in Section 2, we recall the definition of Haar wavelets in Section 3, set up the notation for arbitrary
dimensions and derive a saturation result for the wavelet coefficients depending on their type, more precisely, on the distribution of scaling and wavelet components in the tensor product function. Section 4 shows how the TV norm of a function can be approximated by a properly renormalized subset of wavelet coefficients and gives explicit error estimates for this approximation. How TV denoising based on the ROF functional can be done directly on the wavelet coefficients is demonstrated in Section 5. Finally, in Section 6 we give some example applications of the method on CT datasets.

2 Related work

It has long been observed that soft thresholding of (especially Haar) wavelet coefficients is a useful tool for data denoising. A careful mathematical analysis and discussion of the relationship between Haar wavelet shrinkage and TV regularization has been given in [9] and [10], starting from the study of diffusion processes. More precisely, the authors showed that their diffusion-based process of denoising has strong connections to Haar wavelet shrinkage. Though this approach is likely superior to ours for the goal of denoising of two-dimensional images, it is not applicable to our needs as it relies on an iterative diffusion process of the complete image that is not practically feasible for gigavoxel datasets.

The geometric meaning of Haar wavelet coefficients has also been used implicitly in [11] to determine similarity indices which are essentially based on the role of a subset of the Haar wavelet coefficients as approximations of gradient vectors. This indeed might be carried over to three dimensions and could be directly applied for quality measurements between voxel datasets.

Bounded variation in the context of piecewise constant functions was studied in depth in [12] where it was also shown that approximate solutions of the TV denoising problem can be obtained by suitable thresholding of a Haar wavelet decomposition. While their results rely on significantly weaker assumptions, they are developed only for the bivariate case where the dimension-dependent renormalization of the approximative gradients with respect to the level does not appear, see Remark 2.

3 Haar wavelets

To set up notation, we begin by recalling the well-known concept of tensor product Haar wavelets on \( \mathbb{R}^s \), where \( s \in \mathbb{N} \) stands for the number of variables. We denote the characteristic function on an interval \( I \subset \mathbb{R} \) by \( \chi_I \). In the one-dimensional case, the Haar wavelets are based on the scaling function \( \psi_0 \) and the wavelet \( \psi_1 \), defined as

\[
\psi_0 := \chi_{[0,1]}, \quad \psi_1 := \chi_{[0,\frac{1}{2}]} - \chi_{[\frac{1}{2},1]},
\]

cf. [2]. Since \( \psi_{0,k} := \psi_0(\cdot - k) \) and the normalized wavelets \( \psi_{1,k}^n := 2^{n/2} \psi_1(2^n \cdot - k) \), \( k \in \mathbb{Z} \), \( n \in \mathbb{N}_0 \), form an orthonormal basis of \( L_2(\mathbb{R}) \), any function \( f \in L_2(\mathbb{R}) \) can be expressed, for \( k \in \mathbb{Z} \) and \( n \in \mathbb{N}_0 \), in terms of its scaling coefficients

\[
c_k(f) := \int_\mathbb{R} f(t)\psi_0(t - k) \, dt,
\]

and wavelet coefficients

\[
d_k^n(f) := 2^{n/2} \int_\mathbb{R} f(t)\psi_1(2^n t - k) \, dt,
\]

via

\[
f = \sum_{k \in \mathbb{Z}} c_k(f) \psi_{0,k} + \sum_{n=0}^{\infty} \sum_{k \in \mathbb{Z}} d_k^n(f) \psi_{1,k}^n,
\]

and the wavelet coefficients of different levels give rise to a multiresolution analysis.

The extension to \( s \) variables by means of tensorization is straightforward. Let \( \theta \in \{0,1\}^s \) be the index for a basis function and define, for \( x = (x_1, \ldots, x_s) \in \mathbb{R}^s \),

\[
\psi_\theta(x) := \prod_{j=1}^s \psi_\theta_j(x_j), \quad \theta \in \{0,1\}^s \setminus \{0\}.
\]

We call \( \psi_0 = \chi_{[0,1]} \) the scaling function while all other functions \( \psi_\theta, \theta \in \{0,1\}^s \setminus \{0\} \), are wavelets, yielding \( 2^s - 1 \) wavelet functions at each level. With the scaling coefficients

\[
c_\alpha(f) = \int_\mathbb{R} f(t)\psi_0(t - \alpha) \, dt, \quad \alpha \in \mathbb{Z}^s,
\]
and the wavelet coefficients

\[ d^n_{\theta,\alpha}(f) = 2^{ns/2} \int_{\mathbb{R}} f(t) \psi_n(2^n t - \alpha) \, dt \]  

(7)

with indices \( \alpha \in \mathbb{Z}^s \) and \( n \in \mathbb{N}_0 \), we obtain the orthogonal representation

\[ f = \sum_{\alpha \in \mathbb{Z}^s} c_\alpha(f) \psi_{0,\alpha} + \sum_{n=0}^{\infty} \sum_{\theta,\alpha \in \mathbb{Z}^s} d^n_{\theta,\alpha}(f) \psi^n_{\theta,\alpha} \]  

(8)

with the normalized wavelet functions

\[ \psi^n_{0,\alpha} := 2^{ns/2} \psi_n(2^n \cdot -\alpha). \]  

(9)

Of course, in applications one usually does not work with the infinite series, but only with a finite sum of wavelet levels up to some maximal level \( n_1 \).

### 3.1 Wavelet moments

As a first auxiliary result, we compute moments of the wavelets \( \psi_n(\cdot - \alpha) \) with respect to the polynomials \( (1 + \frac{1}{2} \epsilon)^\gamma \), \( |\gamma| \leq |\theta| \), where \( \epsilon := (1, \ldots, 1) \in \mathbb{Z}^s \) and where \( |\theta| \) denotes the length of the multiindex \( \theta \). The scaled and shifted wavelets have the center of their support at the points

\[ x^n_\alpha := 2^{-n} \left( \alpha + \frac{1}{2} \theta \right). \]  

(10)

Note that these points form a set of non-interlacing grids in \( \mathbb{R}^s \). Indeed, if there were \( n < n' \) and \( \alpha, \alpha' \in \mathbb{Z}^s \) such that \( x^n_\alpha = x^{n'}_{\alpha'} \), then this would mean that \( 2^{n'-n} (\alpha + \frac{1}{2} \theta) = \alpha' + \frac{1}{2} \epsilon \), which is impossible since the left hand side is a point in \( \mathbb{Z}^s \) while the one on the right hand side lies in the shifted grid \( \mathbb{Z}^s + \frac{1}{2} \epsilon \). The moments of monomials centered at such midpoints with respect to the Haar wavelets have a simple explicit form which we record first.

**Lemma 1.** For \( \theta \in \{0,1\}^s \) and \( |\gamma| \leq |\theta| \) one has

\[ \int_{\mathbb{R}^s} (x - x^n_\alpha)^\gamma \psi^n_{\theta,\alpha}(x) \, dx = (-1)^{\theta|\theta|} 2^{-(n+2)|\theta|} \cdot 2^{ns/2} \delta_{\gamma,\theta}. \]  

(11)

**Proof.** The integral in (11) can be written as

\[ \int_{\mathbb{R}^s} (x - x^n_\alpha)^\gamma \psi^n_{\theta,\alpha}(x) \, dx \]

\[ = \prod_{j=1}^{n} \int_{\mathbb{R}} \left( x_j - 2^{-n} \left( \alpha_j + \frac{1}{2} \right) \right)^\gamma \times \psi^n_{\theta,\alpha,j}(x_j) \, dx_j. \]  

(12)

We first note that if \( |\gamma| \leq |\theta| \) and \( \gamma \neq \theta \) then there exists \( k \in \{1, \ldots, s\} \) such that \( \gamma_k = 0 \) and \( \theta_k = 1 \) and therefore

\[ \int_{\mathbb{R}} (x_j - x^n_{\alpha,j})^0 \psi^n_{\theta,\alpha,j}(x_j) \, dx_j = 0. \]  

(13)

Hence, one factor in (12) vanishes and therefore the whole product. In consequence, the integral in (11) is nonzero only for \( \gamma = \theta \). To compute the value of the integral, we note that by shift invariance we can restrict ourselves to the case \( \alpha = 0 \). In the univariate case we note that

\[ \int_{\mathbb{R}} (x - 2^{-n-1})^0 \psi^n_{0,0}(x) \, dx = 2^{-n/2}; \]  

(14)

as well as

\[ \int_{\mathbb{R}} (x - 2^{-n-1}) \psi^n_{1,0}(x) \, dx = -2^{-\frac{3}{2}n-2}; \]  

(15)

so that the integral in (11) takes on the value

\[ \left( \prod_{j \in \{1, \ldots, s\}, \theta_j = 0} 2^{-n/2} \right) \left( \prod_{j \in \{1, \ldots, s\}, \theta_j = 1} (-2^{-\frac{3}{2}n-2}) \right) \]

\[ = (-1)^{\theta|\theta|} 2^{-2(n+1)|\theta|} \cdot 2^{-(\frac{3}{2}n-2)|\theta|} \]

\[ = (-1)^{\theta|\theta|} 2^{-(n+2)|\theta|-ns/2}, \]  

(16)

whenever \( \gamma = \theta \). \( \square \)

### 3.2 Limits of coefficients

The simple observation of Lemma 1 allows us to determine the limits of Haar wavelet coefficients for sufficiently smooth functions and their natural rate of decay.

**Theorem 2.** For \( \theta \in \{0,1\}^s \) suppose that \( f \in C^{[\theta]+1} (\mathbb{R}^s) \). Then, for any \( \alpha \in \mathbb{Z}^s \),

\[ \lim_{n \to \infty} \left| \kappa_{n,|\theta|} d^n_{\theta,\alpha}(f) - \frac{D^{|\theta|} f(x^n_\alpha)}{|\theta|!} \right| = 0, \]  

(17)

where \( \kappa_{n,|\theta|} := (-1)^{|\theta|} 2^{(n+2)|\theta|+ns/2} \).
Proof. For $n \in \mathbb{N}_0$, $\alpha \in \mathbb{Z}^s$, we consider the $(|\theta| + 1)$-st Taylor expansion of $f$ at $x_\alpha$, to find that for any $x \in x_\alpha + \left[-\frac{1}{2}, \frac{1}{2}\right]^s$,

$$f(x) = \sum_{|\gamma| \leq |\theta|} \frac{D^\gamma f(x_\alpha)}{\gamma!} (x - x_\alpha)^\gamma + \sum_{|\gamma| = |\theta| + 1} \frac{D^\gamma f(\xi)}{\gamma!} (x - x_\alpha)^\gamma,$$

where $\xi = \xi(x) \in x_\alpha + 2^{-n} \left[-\frac{1}{2}, \frac{1}{2}\right]^s$. Now, by Lemma 1,

$$d_{\theta,\alpha}^n(f) = \int_{\mathbb{R}^s} f(x) \psi_{\theta,\alpha}(x) \, dx = (-1)^{|\theta|} 2^{-(n+2)|\theta| - ns/2} D^\theta f(x_\alpha) \theta! + \sum_{|\gamma| = |\theta| + 1} \int_{\mathbb{R}^s} \frac{D^\gamma f(\xi(x))}{\gamma!} (x - x_\alpha)^\gamma \psi_{\theta,\alpha}(x) \, dx.$$  

The estimate

$$\left| \int_{\mathbb{R}^s} \frac{D^\gamma f(\xi(x))}{\gamma!} (x - x_\alpha)^\gamma \psi_{\theta,\alpha}(x) \, dx \right| \leq \max_{x \in x_\alpha + 2^{-n} \left[-\frac{1}{2}, \frac{1}{2}\right]^s} \left| \frac{D^\gamma f(x)}{\gamma!} \right| \times 2^{n/2} \int_{x_\alpha + 2^{-n} \left[-\frac{1}{2}, \frac{1}{2}\right]^s} |x - x_\alpha|^\gamma \, dx = 2^{-(n+1)|\theta| + ns/2} \max_{x \in x_\alpha + 2^{-n} \left[-\frac{1}{2}, \frac{1}{2}\right]^s} \left| \frac{D^\gamma f(x)}{\gamma!} \right|$$

holds for any $\gamma$ with $|\gamma| = |\theta| + 1$, and immediately yields that

$$\left| \kappa_{n,|\theta|} d_{\theta,\alpha}^n(f) - \frac{D^\theta f(x_\alpha)}{\theta!} \right| \leq \left( |\theta| + s \right) 2^{n+|\theta|-n-1} \max_{|\gamma| = |\theta| + 1} \max_{x \in x_\alpha + 2^{-n} \left[-\frac{1}{2}, \frac{1}{2}\right]^s} \left| \frac{D^\gamma f(x)}{\gamma!} \right|, \quad (21)$$

whose right hand side indeed tends to zero as $O(2^{-n})$ for $n \to \infty$. \qed

If the $(|\theta| + 1)$-st derivative of $f$ is globally bounded in the sense that

$$\left\| D^{(|\theta|+1)} f \right\|_\infty := \sup_{x \in \mathbb{R}^s} \max_{|\gamma| = |\theta| + 1} \left| D^\gamma f(x) \right| < \infty,$$

then (21) holds independently of $\alpha$, and we get the following improvement of Theorem 2.

Corollary 3. If $f \in C^{n+1}(\mathbb{R}^s)$ and $\| D^k f \|_\infty < \infty$, $k = 1, \ldots, s + 1$, then, for any $\theta \in \{0, 1\}^s$,

$$\lim_{n \to \infty} \max_{\alpha \in \mathbb{Z}^s} \left| \kappa_{n,|\theta|} d_{\theta,\alpha}^n(f) - \frac{D^\theta f(x_\alpha)}{\theta!} \right| = 0. \quad (23)$$

Theorem 2 and its uniform version, Corollary 3, have some consequences. The first one is that Haar wavelet coefficients show what is known as a saturation behavior in Approximation Theory [13], i.e., they cannot decay faster than a given rate, namely $|\kappa_{n,|\theta|}|^{-1}$, no matter how smooth the underlying (nonpolynomial) function $f$ is. Moreover, we record the following for later reference.

Remark 1 (Wavelet coefficient decay). The decay rate of the coefficients depends on $|\theta|$, which means that for smooth functions the coefficients decay faster whenever more “wavelet contribution” is contained in the respective wavelet function $\psi_\theta$. In particular, in wavelet compression as in JPEG2000 where a hard thresholding is applied to the normalized wavelet coefficients, the coefficients $d_{\theta,\alpha}^n(f)$ with a large $|\theta|$ are more likely to be eliminated by the thresholding process. This phenomenon is frequently observed in wavelet compression for images.

3.3 Gradients

From the wavelet coefficients we can form, for $\alpha \in \mathbb{Z}^s$ and $n \in \mathbb{N}_0$, the renormalized vectors

$$d_{\alpha}^n(f) := -2^{n(1-s/2)+2} d_{\alpha}^n(f) = 2^{-ns} \kappa_{n,1} d_{\alpha}^n(f) \quad (24)$$

with

$$d_{\alpha}^n(f) := \left( d_{\alpha_{j,\alpha}}^n(f) : j = 1, \ldots, s \right) \in \mathbb{R}^s. \quad (25)$$

By Theorem 2, the vector $2^{ns} d_{\alpha}^n(f)$ is an approximation for the gradient $\nabla f(x_\alpha)$, $\alpha \in \mathbb{Z}^s$. 

\[ \text{Page 5} \]
Fig. 3 The level image (top) and their vector fields of gradients for multiple levels $n$ (bottom) for $f(x, y) = |x| + |y|$ where $x, y > 0$.

Fig. 4 The level image (top) and their vector fields of gradients for multiple levels $n$ (bottom) for $f(x, y) = x^2 + y^2$ where $x, y > 0$.

$n \in \mathbb{N}_0$. This can be used to extract gradient information directly from the wavelet coefficient vectors $\kappa_{n,1} \left( d_{c,j}^n(f) : j = 1, \ldots, s \right)$. As an example, we consider the subsampled gradient fields for piecewise smooth functions, computed directly from the wavelet coefficients. For the 1-norm and the squared 2-norm, these are shown in Fig. 3 and 4, respectively. The situation changes in the case of non-smooth functions with sharp contours. While in the binary image in Fig. 5 the directions of the gradients are recognized correctly along the straight lines, their lengths are upscaled by a factor of $2^n$ which is due to the different resolution levels as the difference between two neighboring pixels is always either zero or one, but one divides by the resolution-dependent distance between the pixels at different levels to obtain the gradient. This can be compensated by rescaling the gradients by a factor of $2^n$ as shown on Fig. 5.

The estimate (23) is still a pointwise result, even if the error is bounded uniformly with respect
Fig. 5 Simple synthetic image (top) and the resulting gradient field (based on a low-resolution version for increased visual clarity), with the normalization factor $2^n (1 + s/2) + 2 = 2^n + 2$ (middle) and with a normalization $2^n + 2$ that compensates the resolution effects (bottom). In the images, the three colors stand for the three highest resolutions to $\alpha$. We will explore this further in the next section to give an estimate of the TV norm of $f$ by means of wavelet coefficients, which also explains the normalization chosen in (24). So far, the estimates in (17) and (23) only work in the supremum norm.

3.4 Why Haar wavelets?

The above derivation has been restricted to Haar wavelets for good reasons. Besides the fact that their small support and simple structure allows for fast computational methods like octrees, they are the ones that can approximate gradients. In fact, Lemma 1 can be easily extended to arbitrary tensor product wavelets $\psi(x) = \psi_1(x_1) \otimes \cdots \otimes \psi_s(x_s)$ where $\psi_j$ has precisely $\nu_j$ vanishing moments. Then, by the same arguments as in Lemma 1,

$$\int (x - x_{\alpha}^n) \psi(x) \, dx = c_\gamma \delta_{\gamma, \nu}$$

whenever $|\gamma| \leq |\nu|$ and the properly normalized wavelet coefficients $d_{\theta, \alpha}^n(f)$ converge to $D^\theta \nu f(x_\alpha)$, where $\theta \cdot \nu = (\theta_1 \nu_1, \ldots, \theta_s \nu_s)$ for sufficiently smooth $f$. Hence, an approximation of the gradient can be achieved if and only if one uses tensor products of wavelets with only one vanishing moments, i.e., $\nu = (1, \ldots, 1)$. Among these, Haar wavelets are the ones of minimal support and we can compute the normalization constants explicitly in an elementary way.

4 Approximation of the TV norm

We will now derive estimates to show that the vectors $\widehat{d}_{\theta, \alpha}^n(f)$ formed from the wavelet coefficients with only one wavelet contribution, i.e., $d_{\theta_j, \alpha}^n(f)$ for $j = 1, \ldots, s$ in Eq. (25), give a good approximation for the TV norm of $f$ again directly from the wavelet coefficients.

4.1 General remarks

In the rest of this section, we make the assumption that the Haar wavelet coefficients $d_{\theta, \alpha}^n(f)$ stem from a function $f \in C^2(\mathbb{R}^s)$ and the estimate will be given in terms of the magnitude of its second derivatives. This is mainly for technical
reasons and will enable us to give quantitative estimates for the quality of the approximate TV norm. This assumption includes, for example, the case where \( f \) is a bandlimited function, i.e., a function whose Fourier transform \( \hat{f} \) is supported on some compact set \( \Omega \), a frequent assumption in computed tomography, cf. [14]. Moreover, it could be extended to piecewise \( C^2 \) functions, another common assumption in industrial CT, by locality of the wavelet coefficients.

Despite of all that, we are well aware that this is a very restrictive class of functions and that this assumption cannot be made in general. However, the main point of the paper is to derive the proper normalization of the wavelet coefficients, which will depend on the dimensionality of the data, and the proper thresholding process induced by the approximate problem, which will be a thresholding with respect to the length of certain vectors formed by the wavelet coefficients. These two aspects are determined on the dense subset of \( C^2 \) functions and if at all a wavelet thresholding is going to work for approximate TV, this way is straightforward.

4.2 The approximate TV norm

First, let us recall that the TV norm of \( f \in C^2 (\mathbb{R}^s) \), defined as

\[
\|f\|_{TV} := \|\nabla f\|_2 = \int_{\mathbb{R}^s} |\nabla f(x)|_2 \, dx, \tag{27}
\]

where \( | \cdot |_2 \) denotes the Euclidean norm in \( \mathbb{R}^s \). This functional plays a fundamental role in many imaging applications, at least since the work of Rudin, Osher and Fatemi [8], which made TV regularization a standard method in image processing, in particular for image denoising. For a survey on TV based applications, we refer the reader to [15].

In the following we will need the Frobenius norm of the second derivative of \( f \in C^2 (\mathbb{R}^s) \), i.e., for \( x \in \mathbb{R}^s \) we set

\[
|D^2 f(x)|_F := \left( \sum_{j,k=1}^s \left( \frac{\partial^2 f(x)}{\partial x_j \partial x_k} \right)^2 \right)^{1/2}. \tag{28}
\]

We also define for \( n \in \mathbb{N}_0 \), the functional

\[
H_n(f)
\]

\[
:= \sum_{\alpha \in \mathbb{Z}^s} 2^{-n\alpha} \max_{x \in \mathbb{R}^s, 2^{-n/2} \leq |x| \leq 2^{-n/2}} |D^2 f(x)|_F, \tag{29}
\]

which may in general take on the value \( \infty \), but has the property that \( H_{n+1}(f) \leq H_n(f) \), \( n \in \mathbb{N}_0 \), and is clearly bounded from below. Functions for which \( H_n(f) < \infty, n \in \mathbb{N}_0 \), are for example compactly supported functions in \( C^2 (\mathbb{R}^s) \) as then the Frobenius norm is globally bounded and the sum in (29) becomes a finite one. Since bandlimited functions decrease exponentially and have a bandlimited second derivative, they also satisfy \( H_n(f) < \infty, n \in \mathbb{N}_0 \).

We will prove in this section that the \( \ell_1 \) norm of the sequence \( |d^n_\alpha(f)|_0, \alpha \in \mathbb{Z}^s \), given as

\[
\left\| |d^n_\alpha(f)|_0 \right\|_1 := \sum_{\alpha \in \mathbb{Z}^s} |d^n_\alpha(f)|_0
\]

\[
= 2^{n(1-s/2)+2} \sum_{\alpha \in \mathbb{Z}^s} \left( \sum_{j=1}^s (d^n_{j,\alpha}(f))^2 \right)^{1/2} \tag{30}
\]

with \( \hat{d}^n(f) := (\hat{d}^n_\alpha(f): \alpha \in \mathbb{Z}^s) \), is an approximation to \( \|f\|_{TV} \), and we provide an explicit estimate for the error of this approximation. The result is as follows.

**Theorem 4.** If \( f \in C^2 (\mathbb{R}^s) \) and \( H_n(f) < \infty \) for some \( n \in \mathbb{N}_0 \), then there exists, for any \( n_0 \in \mathbb{N}_0 \), a constant \( C > 0 \) that depends only on \( n_0 \) and \( f \), such that

\[
\left\| \left|d^n(f)\right|_0 \right\|_1 - \|f\|_{TV} \leq C 2^{-n} \left\| |D^2 f|_F \right\|_1
\]

holds for all \( n \geq n_0 \).

The slightly strange formulation of Theorem 4 will become clear from the proof. Indeed, the number \( n_0 \) can be used to control and reduce the constant \( C \) by relying on high-pass information only. More details on that in Remark 3 later. Our main point in Theorem 4 is that considering some part of the wavelet coefficients with a proper dimension-dependent normalization gives a discrete approximation of the TV norm of the underlying function with a quantitative error.

**Remark 2** (Normalization of coefficients).

1. The normalization of the coefficients \( \hat{d}^n_\alpha(f) \) in (24) may be somewhat surprising at first view because of its dependency on the
dimensionality $s$. For $s = 1$, coefficients at higher level are reweighted with an increasing weight $2^{n/2+2}$, for the image case $s = 2$ the orthonormal normalization is just perfect, while for $s > 2$ the weights decrease and penalize higher levels, which is particularly relevant for our main application, namely the volume case $s = 3$.

2. Nevertheless, there is an explanation for this behavior. The TV norm considers the Euclidean length of the gradient, hence a one-dimensional feature that scales like $2^{-n}$, while all other normalizations within the integrals are based on volumes that scale like $2^{-ns}$. This feature also appears when considering coefficients with respect to unnormalized wavelets.

3. It is important to keep in mind that it is the proper renormalization of the relevant coefficients and their treatment as vectors that allows us to approximate the TV norm by means of summing the coefficients.

4.3 Proof of Theorem 4

We split the proof of Theorem 4 into two parts, first showing that $\|d^n(f)\|_2$ is close to a cubature formula and then estimating the quality of the cubature formula. To that end, we define the sequence

$$f^n_\alpha := (\nabla f(x^n_\alpha) : \alpha \in \mathbb{Z}^s)$$

of sampled gradients of $f$.

**Lemma 5.** If $f \in C^2(\mathbb{R}^s)$ and $H_n(f) < \infty$ for some $n \in \mathbb{N}_0$, then there exists, for any $n_0 \in \mathbb{N}_0$, a constant $C > 0$ that depends only on $n_0$ and $f$ such that

$$\|d^n(f)\|_2 - 2^{-ns} \|\nabla f^n_\alpha\|_2 \leq 2^{-n}C \|D^2f\|_F \leq 1$$

(33)

holds for any $n \geq n_0$.

**Proof.** We again use a Taylor expansion at $x^n_\alpha$, this time with an integral remainder. It follows directly by applying the univariate formula to $t \mapsto f(x_\alpha + t(x - x^n_\alpha))$, $t \in [0, 1]$, and takes the form

$$f(x) = f(x^n_\alpha) + \nabla f(x^n_\alpha)^T (x - x^n_\alpha) + \int_0^1 (1 - t)D^2f(x^n_\alpha + tz_\alpha) \, dz_\alpha \, dt$$

+ $\int_0^1 (1 - \xi)(x - x^n_\alpha)^T D^2f(x^n_\alpha + \xi(x - x^n_\alpha)) \times (x - x^n_\alpha) \, d\xi.$

(34)

Using Lemma 1, it follows for $j = 1, \ldots, s$ that

$$d^n_{j,\alpha}(f) = \int_{\mathbb{R}^s} f(x)\psi^n_{j,\alpha}(x) \, dx$$

$$= -2^{-n(1+s/2)} - \partial f(x^n_\alpha + h^n_{j,\alpha}(f)) \times (x - x^n_\alpha) \, d\xi.$$ 

(35)

where

$$h^n_{j,\alpha}(f) := \int_{\mathbb{R}^s} \int_0^1 (1 - \xi) \, d\xi.$$ 

(36)

The magnitude of the latter can be estimated as

$$|h^n_{j,\alpha}(f)| \leq 2^{ns/2} \int_0^1 (1 - \xi) \, d\xi.$$ 

(37)

This estimate is independent of $j$. Hence, if we define

$$h_\alpha^n(f) = (h^n_{j,\alpha}(f) : j = 1, \ldots, s),$$ 

(38)

we get that

$$|h^n_{j,\alpha}(f)| \leq 2^{ns/2} \int_0^1 (1 - \xi) \, d\xi.$$ 

(39)
The step functions
\[
g_n^\alpha(f) := \sum_{\alpha \in \mathbb{Z}^s} \left( \max_{x \in x_n^\alpha + 2^{-n}[-\frac{1}{2}, \frac{1}{2})} |D^2 f(x)|_F \right) \times \chi_{[0,1]}(2^n \cdot -\alpha),
\]
(40)

converge monotonically decreasing and pointwise to \( |D^2 f|_F \) as \( n \to \infty \) and satisfy \( 0 \leq g_n^\alpha(f) \leq H_n(f) \) which is finite for sufficiently large \( n \). Hence, using that \( 0 \leq |D^2 f|_F \leq g_n^\alpha(f), |D^2 f| \in L_1(\mathbb{R}^s) \). In addition,
\[
g_n^\alpha(f) := \sum_{\alpha \in \mathbb{Z}^s} \left( \min_{x \in x_n^\alpha + 2^{-n}[-\frac{1}{2}, \frac{1}{2})} |D^2 f(x)|_F \right) \times \chi_{[0,1]}(2^n \cdot -\alpha),
\]
(41)
is trivially bounded from below by 0, from above by \( |D^2 f|_F \) and converges monotonically increasing to \( |D^2 f|_F \). Since for any \( \xi \in [0,1] \) we have that
\[
\min_{x \in x_n^\alpha + 2^{-n}[-\frac{1}{2}, \frac{1}{2})} |D^2 f(x)|_F \\
\leq 2^{-ns} \int_{2^{-n}[-\frac{1}{2}, \frac{1}{2})} |D^2 f(x_n^\alpha + \xi) d\xi|_F dx \\
\leq \max_{x \in x_n^\alpha + 2^{-n}[-\frac{1}{2}, \frac{1}{2})} |D^2 f(x)|_F,
\]
(42)
and since \( \|g_n^\alpha(f)\|_1 \to \|D^2 f\|_F \) as \( n \to \infty \), it follows that
\[
\lim_{n \to \infty} \sum_{\alpha \in \mathbb{Z}^s} \int_{2^{-n}[-\frac{1}{2}, \frac{1}{2})} |D^2 f(x_n^\alpha + \xi)|_F d\xi = \int_{\mathbb{R}^s} |D^2 f(x)|_F dx
\]
(43)
uniformly in \( \xi \). In particular, there exists, for any \( n_0 \in \mathbb{N}_0 \), a constant \( C = C(n_0, f) \) such that
\[
\sum_{\alpha \in \mathbb{Z}^s} \int_{2^{-n}[-\frac{1}{2}, \frac{1}{2})} |D^2 f(x_n^\alpha + \xi)|_F d\xi \\
\leq C \int_{\mathbb{R}^s} |D^2 f(x)|_F dx
\]
(44)
holds for \( n \geq n_0 \). Moreover,
\[
\lim_{n_0 \to \infty} C(n_0, f) = 1.
\]
(45)
Under the assumption that \( n \geq n_0 \) this yields that
\[
\sum_{\alpha \in \mathbb{Z}^s} |h_n^\alpha(f)|_2 \leq C \frac{s^{3/2}2^{-2n-2+ns/2}}{2} \|D^2 f\|_F.
\]
(46)
Multiplying (35) by \( \kappa_{n,1} = -2^{n(1-s/2)+2} \) we get that
\[
\left| \kappa_{n,1} d_{1,\alpha}^n(f) - 2^{-ns} \frac{\partial f(x_n^\alpha)}{\partial x_j} \right| \\
\leq |\kappa_{n,1}| \|h_{j,\alpha}(f)\|,
\]
(47)
and substituting (46) we then obtain
\[
\sum_{\alpha \in \mathbb{Z}^s} \left| \left| \frac{\partial}{\partial x_j} d_n^\alpha(f) \right|_2 - 2^{-ns} |\nabla f(x_n^\alpha)|_2 \right| \\
\leq |\kappa_{n,1}| \|h_n^\alpha(f)\|_2.
\]
(48)
Summing this over \( \alpha \in \mathbb{Z}^s \) gives
\[
\sum_{\alpha \in \mathbb{Z}^s} \left| \left| \frac{\partial}{\partial x_j} d_n^\alpha(f) \right|_2 - 2^{-ns} |\nabla f(x_n^\alpha)|_2 \right| \\
\leq C \kappa_{n,1} \frac{s^{3/2}}{2} \|D^2 f\|_F,
\]
(49)
which is (33).
\( \square \)

**Remark 3** (Choice of \( C \) and \( n_0 \)). The constant \( C \) in (33) can be chosen as \( \frac{s^{3/2}}{2} + \varepsilon \) for any \( \varepsilon > 0 \) by making \( n_0 \) sufficiently large. In practice this means to avoid the low-pass content of the wavelet transformation which approximates the gradient only in a rather poor way. Of course, the dependency of the constant \( C \) on \( f \) would also have to be taken into account.

Since (33) immediately implies for \( n \geq n_0 \) that
\[
\left| \left| \frac{\partial}{\partial x_j} d_n^\alpha(f) \right|_2 \right| - \sum_{\alpha \in \mathbb{Z}^s} 2^{-ns} |\nabla f(x_n^\alpha)|_2 \\
\leq C 2^{-n} \|D^2 f\|_F,
\]
(50)
the discrete sum \( \left| \left| \frac{\partial}{\partial x_j} d_n^\alpha(f) \right|_2 \right| \) approximates the uniform cubature formula for the TV norm. To complete the proof, we only have to recall the approximation quality of the cubature formula. This follows by standard arguments which we include for completeness.
Lemma 6. If $f \in C^2(\mathbb{R}^s)$ and $H_n(f) < \infty$ for some $n \in \mathbb{N}_0$, then there exists, for any $n_0 \in \mathbb{N}_0$, a constant $C > 0$ that depends only on $n_0$ and $f$ such that

$$\|2^{-ns}|f^n_\varepsilon|_2 - \|f\|_{TV}\| \leq C 2^{-n} \|D^2 f\|_{F,1},$$

(51)

for $n \geq n_0$.

Proof. Writing

$$\nabla f(x) = \nabla f(x^n_\alpha) + \int_{0}^{1} D^2 f(x^n_\alpha + \xi(x-x^n_\alpha)) \cdot (x-x^n_\alpha) d\xi,$$

(52)

we get by similar transformations as in the proof of Lemma 5 that

$$\int_{x_n + 2^{-n}[-\frac{1}{2}, \frac{1}{2}]^s} |\nabla f(x) - \nabla f(x^n_\alpha)|_2 \ dx$$

$$\leq \int_{2^{-n}[-\frac{1}{2}, \frac{1}{2}]^s} \int_{0}^{1} |D^2 f(x^n_\alpha + \xi x)|_F |x|_2 \ dx \ d\xi,$$

(53)

hence

$$2^{-ns} |\nabla f(x^n_\alpha)|_2 - \int_{x_n + 2^{-n}[-\frac{1}{2}, \frac{1}{2}]^s} |\nabla f(x)|_2 \ dx$$

$$\leq \sqrt{2} 2^{-n-1} \int_{2^{-n}[-\frac{1}{2}, \frac{1}{2}]^s} \int_{0}^{1} |D^2 f(x^n_\alpha + \xi x)|_F \ d\xi.$$  

(54)

By the same arguments as in Lemma 5, there exist $n_0 \in \mathbb{N}_0$ and $C > 0$ such that

$$\left| \sum_{\alpha \in \mathbb{Z}^s} 2^{-ns} |\nabla f(x^n_\alpha)|_2 - \int_{\mathbb{R}^s} |\nabla f(x)|_2 \ dx \right|$$

$$\leq \frac{\sqrt{2}}{2} 2^{-n} C \|D^2 f\|_{F,1},$$

(55)

giving (51).

Now it is easy to complete the proof of the main theorem.

Proof of Theorem 4. Combining (50) and (51), the triangle inequality gives, for $n \geq n_0$,

$$\left\|\left|d^n(f)\right|_F - \|f\|_{TV}\right\| \leq C 2^{-n} \|D^2 f\|_{F,1},$$

(56)

where the constant can be anything of the form

$$\frac{s^{3/2}}{2} + \frac{s^{1/2}}{2} + \varepsilon = \sqrt{s} \frac{s+1}{2} + \varepsilon, \quad \varepsilon > 0,$$

(57)

by selecting a sufficiently high value for $n_0$. \qed

### 4.4 Estimation over several levels

The estimate in (31) holds for any sufficiently high wavelet level separately and the error of the estimate decreases with $n$, so it might appear reasonable to approximate the TV norm just by the maximal level. Unfortunately, wavelet coefficients of high level are most affected by high frequency noise which would lead to an overdetection of gradients. Therefore, it makes sense to incorporate also wavelet coefficients of lower resolution.

Indeed, in practical applications one starts with finite data on the finest resolution $n_1 + 1$, i.e.,

$$c^{n_1+1}_\alpha(f) := \int_{\mathbb{R}^s} f(x) \psi^{n_1+1}_{0,\alpha}(x) \ dx, \quad \alpha \in \mathbb{Z}^s,$$

(58)

and then computes the coefficients in the wavelet decomposition

$$\sum_{\alpha \in \mathbb{Z}^s} c^{n_1+1}_\alpha(f) \psi^{n_1+1}_{0,\alpha} = \sum_{k=0}^{n_1} \sum_{\theta \in (0,1) \setminus \{0\}} \sum_{\alpha \in \mathbb{Z}^s} d^{n_1}_{\theta,\alpha}(f) \psi^{k}_{\theta,\alpha}.$$  

(59)

To get an approximation of the TV norm that uses as many levels as possible at the same time, namely $d^n(f)$, $n = n_0, \ldots, n_1$, we make use of suitable averaging to obtain almost the same rate of accuracy as by the highest level alone.

**Proposition 7.** If, under the assumptions of Theorem 4, we define

$$\mu_n := \frac{2^{n-n_1}}{2^{n_0-n_1}}, \quad n = n_0, \ldots, n_1,$$

(60)
then
\[
\left| \sum_{n=n_0}^{n_1} \mu_n \left\| \hat{d}^n(f) \right\|_2 - \|f\|_{TV} \right| \\
\leq C(n_1 + 1 - n_0)2^{-n_1} \left\| D^2 f \right\|_F.
\] (61)

Proof. By construction,
\[
\sum_{n=n_0}^{n_1} \mu_n = 1,
\] (62)
and we thus have that
\[
\left| \sum_{n=n_0}^{n_1} \mu_n \left\| \hat{d}^n(f) \right\|_2 - \|f\|_{TV} \right| \\
= \left| \sum_{\alpha \in \mathbb{Z}^s} \sum_{n=n_0}^{n_1} \mu_n \left\| \hat{d}^n_\alpha(f) \right\|_2 - \sum_{n=n_0}^{n_1} \mu_n \|f\|_{TV} \right| \\
\leq \sum_{n=n_0}^{n_1} \mu_n \left| \left\| \hat{d}^n(f) \right\|_2 - \|f\|_{TV} \right| \\
\leq \frac{C}{2 - 2{n_0 - n_1}} \left\| D^2 f \right\|_F \sum_{n=n_0}^{n_1} 2^{n-n_1} 2^{-n} \\
\leq (n_1 + 1 - n_0)C 2^{-n_1} \left\| D^2 f \right\|_F,
\] (63)
which is (61) with C being the constant in Theorem 4 for initial level \(n_0\).

In fact, any averaging of \(\left\| \hat{d}^n(f) \right\|_2\), \(n = n_0, \ldots, n_1\), would yield an approximation for the TV norm, but the particular choice of the weights in (60) ensures that the rate of convergence obtained by this averaging process is the same as that on the highest level \(n_1\), only affected by the “logarithmic” number \(n_1 - n_0\) of the levels incorporated in the approximation process.

5 Approximate TV regularization

TV regularization is a standard procedure for many imaging applications nowadays, especially for denoising. It consists of solving, for a given image \(f\), an optimization problem of the basic form
\[
\min_{u} \frac{1}{2} \|f - u\|_2^2 + \lambda \|u\|_{TV}, \quad \lambda > 0,
\] (64)

where the regularization term \(\lambda \|u\|_{TV}\) encourages a smooth or less noisy behavior of \(u\) whose influence is controlled by the parameter \(\lambda\). In most applications, \(\|u\|_{TV}\) is computed for discrete data \(u\) by numerical differentiation, usually by means of differences. In particular, this not only requires access to the full image, but the discrete gradient needs an additional amount of \(s\) times the memory consumption of the original image. Since this is unacceptable in realistic applications, where the image is larger than the available memory, the straightforward approach is to use the wavelet coefficients as a computationally efficient approximation for \(\|u\|_{TV}\).

A relaxation of the optimization problem (64) can be solved explicitly by standard methods that we are going to explain now. To that end, we assume that \(f\) and \(u\) are given as finite orthonormal wavelet expansions
\[
f = \sum_{\alpha \in \mathbb{Z}^s} c_\alpha(f) \psi_0, \alpha
\]
\[
u = \sum_{\alpha \in \mathbb{Z}^s} c_\alpha(u) \psi_0, \alpha
\]

The first term in (64) can be differentiated with respect to the wavelet coefficients yielding
\[
\frac{\partial}{\partial c_\alpha(u)} \frac{1}{2} \|f - u\|_2^2 = c_\alpha(u) - c_\alpha(f),
\]
\[
\frac{\partial}{\partial d^n_{\alpha, \theta}(u)} \frac{1}{2} \|f - u\|_2^2 = d^n_{\alpha, \theta}(u) - d^n_{\alpha, \theta}(f).
\] (66)

For the regularization term we now use the approximation from Proposition 7 for \(\|u\|_{TV}\), i.e., we solve the approximate problem
\[
\min_{u} \frac{1}{2} \|f - u\|_2^2 + \lambda \sum_{n=n_0}^{n_1} \mu_n \left\| \hat{d}^n(u) \right\|_2
\] (67)
with the regularization term
\[
F(u) = \sum_{\alpha \in \mathbb{Z}^s} \sum_{n=n_0}^{n_1} \mu_n \left\| \hat{d}^n_\alpha(u) \right\|_2.
\] (68)
whose (generally set-valued) subgradient is composed, for \( \alpha \in \mathbb{Z}^* \), and \( n = n_0, \ldots, n_1 \), of
\[
\partial \tilde{d}_n^\alpha (u) F(u) = \mu_n \left\{ \begin{array}{ll}
B_1(0), & \tilde{d}_n^\alpha (u) = 0, \\
\tilde{d}_n^\alpha (u) - \alpha \tilde{d}_n^\alpha (u) & \tilde{d}_n^\alpha (u) \neq 0,
\end{array} \right.
\]
where \( B_1(0) := \{ x \in \mathbb{R}^n : |x|_2 \leq 1 \} \) denotes the unit ball. A necessary and sufficient condition for \( u \) to be a solution of the convex optimization problem (67) is that
\[
0 \in \partial \left( \frac{1}{2} \| f - \cdot \|_2^2 + \lambda \| d(\cdot) \|_1 \right)(u),
\]
cf. [16], which is equivalent to setting all coefficients in the wavelet expansion of \( u \) equal to those of \( f \), except the nontrivial conditions
\[
0 \in \left( \tilde{d}_n^\alpha (f) - \tilde{d}_n^\alpha (u) \right) + \lambda \partial \tilde{d}_n^\alpha (u) F(u)
\]
for \( \alpha \in \mathbb{Z}^* \), \( n = n_0, \ldots, n_1 \). (71) means that
\[
\tilde{d}_n^\alpha (f) \in \tilde{d}_n^\alpha (u) + \mu_n \lambda \partial \tilde{d}_n^\alpha (u) F(u),
\]
i.e., either
\[
\tilde{d}_n^\alpha (f) \in \mu_n \lambda B_1(0),
\]
if \( \tilde{d}_n^\alpha (u) = 0 \), or
\[
\tilde{d}_n^\alpha (f) = \left( 1 + \frac{\mu_n \lambda}{\| \tilde{d}_n^\alpha (u) \|_2} \right) \tilde{d}_n^\alpha (u)
\]
if \( \tilde{d}_n^\alpha (u) \neq 0 \). Solving (73) and (74) for the coefficients of \( u \), we get the explicit representation of the solution of (67). Note that this will be a block thresholding operation and that the thresholding is in terms of the length of a whole vector of wavelet coefficients.

**Proposition 8.** The solution of (71) can be computed by soft thresholding of the length of the normalized coefficient vectors \( \tilde{d}_n^\alpha (f) \), i.e., as
\[
\tilde{d}_n^\alpha (u) = \left( 1 - \frac{\mu_n \lambda}{\| \tilde{d}_n^\alpha (f) \|_2} \right) \tilde{d}_n^\alpha (f),
\]
for \( \alpha \in \mathbb{Z}^* \), \( n = n_0, \ldots, n_1 \).

**Proof.** The case (73) describes \( \tilde{d}_n^\alpha (u) = 0 \) and is equivalent to \( \| \tilde{d}_n^\alpha (f) \|_2 \leq \mu_n \lambda \). Otherwise, (74) yields
\[
\tilde{d}_n^\alpha (f) \leq \left( 1 + \frac{\mu_n \lambda}{\| \tilde{d}_n^\alpha (f) \|_2} \right) \tilde{d}_n^\alpha (u)
\]
can be easily checked to have the solution
\[
\tilde{d}_n^\alpha (u) = \left( 1 - \frac{\mu_n \lambda}{\| \tilde{d}_n^\alpha (f) \|_2} \right) \tilde{d}_n^\alpha (f),
\]
which is (75). \( \square \)

One could also approximate the TV norm in (64) by a single set of wavelet coefficients on some level \( n \) and solve
\[
\min_u \left\{ \frac{1}{2} \| f - u \|_2^2 + \lambda \| \tilde{d}_n^\alpha (u) \|_2 \right\}_1
\]
by
\[
\tilde{d}_n^\alpha (u) = \left( 1 - \frac{\lambda}{\| \tilde{d}_n^\alpha (f) \|_2} \right) \tilde{d}_n^\alpha (f), \quad \alpha \in \mathbb{Z}^*,
\]
only applied to coefficients of level \( n \). Also, the weights \( \mu_n \) can be chosen arbitrarily.

To conclude, we again make a short comparison with [9, 10] where wavelet shrinkage applied to each wavelet coefficient separately is related to diffusion filtering processes. Nevertheless, our findings show that for an approximation of the TV functional the shrinkage process has to be adapted:

1. The shrinkage has to be applied to the length of the vectors \( \tilde{d}_n^\alpha (f) \) and not to its components separately. Also only those coefficients have to be taken into account that contain a single wavelet component, as already mentioned at the beginning of Section 4.
2. The coefficients have to be properly renormalized and this renormalization depends on the level of the wavelet coefficients and the dimensionality of the problem.
3. It matters that we use Haar wavelets here. The reconstruction of first derivatives requires, at least in our approach, univariate wavelets with only a single vanishing moment, and the simple explicit expressions for proper renormalization are even due to the explicit nature and support size of Haar wavelets.

Under these conditions, we can give precise error estimates for the procedure that works entirely on the wavelet coefficients.

6 Applications to volume data

We finally give some numerical results obtained from applying the method to a dataset from industrial computed tomography.

6.1 Example dataset

To illustrate the results, we will use a typical industrial dataset, a Mahle motor piston, as a running example. In Fig. 6, we see a single 2D slice view of that data, where the different materials are visible: surrounding (noisy) air, styrofoam (piston fixation), aluminum (piston body) and iron (ring). The dataset itself is of size $464 \times 464 \times 414$ voxels and contains nonnegative 16-bit integer values. Due to its well-distinguishable materials and the fact that, by means of discrete differences, we can also compute its TV norm explicitly for comparison, it is nevertheless useful for illustration purposes despite its small size. As described in Section 3.3, the single-wavelet-component coefficients can be used for approximating the gradients, see Fig. 7.

6.2 Numerical examples and heuristics

Thresholding on the wavelet coefficients with $\theta = \epsilon_j$ clearly has its limitations that will become visible when applying very strong regularization to a dataset as the approach thresholds only $s$ blocks of coefficients, while the other $2^s - s - 1$ ones with $|\theta| > 1$ remain unaffected. This phenomenon becomes more and more prominent in higher dimensions and is very well observable for $s = 3$ already.

To compensate this behavior, we propose a heuristic sparsification in the following way: whenever the vector $d_n^\alpha(f)$ is thresholded to zero for some index $\alpha \in \mathbb{Z}^s$ and level $n \in \mathbb{N}_0$, we set all wavelet coefficients $d_{n,\theta}^\alpha(f)$, $\theta \in \{0,1\} \setminus \{0\}$, to zero and not only $d_n^\alpha(f)$, $j = 1, \ldots, s$, which make up the vector $d_n^\alpha(f)$. The rationale behind this heuristic is, on the one hand, the assumption of locally homogeneous data combined with the observation stated in Remark 1, that in such a
| \( \lambda \) | \( 10^2 \) | \( 10^3 \) | \( 10^4 \) | \( 10^5 \) |
|----------------|-------------|-------------|-------------|-------------|
| \( \|u\|_{TV}/\|f\|_{TV} \) | 97% | 81% | 72% | 76% |
| \( \sum \mu_n \|d^n(u)\|_1 \) | 93% | 49% | 20% | 8.5% |
| \( \sum \mu_n \|d^n(f)\|_1 \) | 0.38% | 2.9% | 6.5% | 13% |
| PSNR | 71 | 53 | 46 | 41 |
| Zeros | 19% | 40% | 52% | 53% |

Table 1: Results (up to two digits of accuracy) of approximate TV for multiple regularization parameters, relative to the values for the original data, e.g., to \( \|f\|_{TV} \). “Zeros” stands for wavelet coefficients that have been thresholded to zero.

| \( \lambda \) | \( 10^2 \) | \( 10^3 \) | \( 10^4 \) | \( 10^5 \) |
|----------------|-------------|-------------|-------------|-------------|
| \( \|u\|_{TV}/\|f\|_{TV} \) | 97% | 79% | 55% | 53% |
| \( \sum \mu_n \|d^n(u)\|_1 \) | 93% | 49% | 20% | 8.5% |
| \( \sum \mu_n \|d^n(f)\|_1 \) | 0.41% | 3.4% | 7.3% | 13% |
| PSNR | 70 | 52 | 45 | 40 |
| Zeros | 20% | 60% | 98% | 99.8% |

Table 2: Results (up to two digits of accuracy) obtained by also zeroing coefficients with \( |\theta| > 1 \) if all coefficients with \( |\theta| = 1 \) are thresholded to zero, affecting all metrics except the Haar-wavelet-based TV approximation. Note that the behavior of the TV norm is now more similar to the behavior of the TV approximation.

situation, the wavelet coefficients with \( |\theta| > 1 \) should decay faster than those with \( |\theta| = 1 \). On the other hand, higher order wavelet coefficients are usually more sensitive to noise and therefore they may be nonzero just because of higher order reactions to noise.

6.3 Results

Given the motor piston dataset \( f \), we compare the solutions to the approximate TV minimization problem \( u \) for different parameters \( \lambda \) satisfying (75) in Tables 1 and 2. The used metrics consider data fidelity, regularization performance, and memory efficiency, also normalized with respect to the corresponding values of the original data \( f \). The proposed method does indeed lower the approximate wavelet TV norm. However, it seems to not influence all the coefficients necessary to reduce the TV norm in the standard basis. The heuristic sparsification, on the other hand, shows the same reduction of the approximate wavelet TV norm but also consistently leads to a smaller TV norm in the standard basis, and a higher overall percentage of coefficients of value zero as well. Fig. 8 and 9 indicate that, for larger thresholds, the reduction of noise texture is stronger when using the heuristic sparsification while the image quality at material transitions is comparable. The three-dimensional views show similar noise texture properties.

In this practical example, the heuristic sparsification also leads to less speckle-like phenomena regarding the surrounding noise.
compared to soft thresholding, where the coefficients are manipulated independently, which can be seen in Fig. 10 and 11. The additional heuristic sparsification addresses this kind of discontinuous behavior, which is typical for componentwise thresholding of tensorized Haar wavelet coefficients, by considering all wavelet coefficients $d_{\theta,\alpha}^n(f), \theta \in \{0, 1\}^s \setminus \{0\}$, at once for some index $\alpha \in \mathbb{Z}^s$ and level $n \in \mathbb{N}_0$ instead.

6.4 Timing

Given Haar-wavelet-transformed data, we now compare the runtimes of a standard wavelet-based regularization technique, soft thresholding, with the proposed method and the heuristic sparsification. Componentwise soft thresholding with parameter $\lambda$ modifies a wavelet
Reconstructed motor piston slice region after soft thresholding (top), and after heuristic sparsification (bottom), exhibiting less speckle-like noise in different resolutions, and similar sharpness at the important edges at the borders between air and aluminum. Both reconstructions are based on the wavelet data that is visualized in Fig. 10. Gamma correction was applied to both images to highlight the noise details.

\[ x \mapsto \text{sgn}(x)(|x| - \lambda)_+ . \]  

(80)

The CPU-only implementation was run on an AMD Ryzen 7 4800H system. The two datasets to test are the motor piston wavelet coefficients without any manipulation, and the 30 GB mummy wavelet coefficients that remain after zeroing small coefficients (via hard thresholding) that are mentioned in the introduction. For the piston, the complete dataset of size 464 \times 464 \times 414 voxels is compared. For the mummy, not the whole volume of size 7584 \times 7584 \times 9216 voxels, but

| \( \lambda \) | \( 10^2 \) | \( 10^3 \) | \( 10^4 \) | \( 10^5 \) |
|---|---|---|---|---|
| min. time (s) | 0.086 | 0.036 | 0.025 | 0.034 |
| avg. time (s) | 0.093 | 0.039 | 0.037 | 0.037 |
| max. time (s) | 0.108 | 0.044 | 0.041 | 0.040 |

| \( \lambda \) | \( 10^2 \) | \( 10^3 \) | \( 10^4 \) | \( 10^5 \) |
|---|---|---|---|---|
| min. time (s) | 0.034 | 0.036 | 0.031 | 0.031 |
| avg. time (s) | 0.047 | 0.048 | 0.045 | 0.043 |
| max. time (s) | 0.053 | 0.052 | 0.048 | 0.050 |

| \( \lambda \) | \( 10^2 \) | \( 10^3 \) | \( 10^4 \) | \( 10^5 \) |
|---|---|---|---|---|
| min. time (s) | 0.342 | 0.293 | 0.297 | 0.297 |
| avg. time (s) | 0.360 | 0.306 | 0.306 | 0.308 |
| max. time (s) | 0.382 | 0.329 | 0.318 | 0.321 |

| \( \lambda \) | \( 10^2 \) | \( 10^3 \) | \( 10^4 \) | \( 10^5 \) |
|---|---|---|---|---|
| min. time (s) | 0.464 | 0.460 | 0.457 | 0.445 |
| avg. time (s) | 0.486 | 0.480 | 0.472 | 0.475 |
| max. time (s) | 0.535 | 0.506 | 0.494 | 0.498 |

| \( \lambda \) | \( 10^2 \) | \( 10^3 \) | \( 10^4 \) | \( 10^5 \) |
|---|---|---|---|---|
| min. time (s) | 0.574 | 0.596 | 0.580 | 0.551 |
| avg. time (s) | 0.624 | 0.615 | 0.592 | 0.584 |
| max. time (s) | 0.709 | 0.665 | 0.644 | 0.635 |

Table 3: Regularization times in seconds for the piston dataset. Soft thresholding (top), the proposed method (middle), and the heuristic sparsification (bottom) are applied to the wavelet coefficients with ten repetitions.

Table 4: Regularization times in seconds for the mummy teeth ROI. Soft thresholding (top), the proposed method (middle), and the heuristic sparsification (bottom) are applied to the wavelet coefficients with ten repetitions.

a region of interest of size $1024 \times 1024 \times 1024$ containing the teeth similar to Fig. 2 is considered. In both cases, regularization time was measured for ten repetitions. Tables 3 and 4 show that, compared to pointwise soft thresholding, the proposed technique needs up to 50% more computation time, whereas the additional heuristic sparsification may take 100% more time on average. However, note that in relation to the calculation of the inverse wavelet transform (around 24 seconds for the mummy teeth ROI), this increase in runtime is negligible and demonstrates that the proposed Haar-wavelet-based approximate TV minimization is a computationally cheap regularization strategy.

7 Conclusion

In contrast to computationally expensive TV regularization methods, our approach only considers an approximation of the TV norm, and does it in an efficient way by directly shrinking the Haar wavelet coefficients. This renders TV-like regularization feasible for very large datasets. In two examples, the performance of our method was discussed and compared to a standard wavelet-based regularization approach. The results can even be improved by a computationally cheap and straightforward heuristic modification of the thresholding process.

Acknowledgements. This work was supported by the German Federal Ministry of Education and Research (BMBF) within the project KI4D4E: Ein KI-basiertes Framework für die Visualisierung und Auswertung der massiven Datenmengen der 4D Tomographie für Endanwender von Beamlines under the title 05D2022 in collaboration between the Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V., the University of Stuttgart, the University of Passau, the Friedrich-Alexander University of Erlangen, the Karlsruhe Institute of Technology, the Helmholtz center Berlin, the Helmholtz center Hereon, and the Forschungszentrum Jülich.

We thank Christoph Heinzl for providing us with Figures 1 and 2.

Conflict of interest. The authors declare that they have no conflict of interest.

ORCID iDs.

- Tomas Sauer https://orcid.org/0000-0002-3182-2141
- Andreas Michael Stock https://orcid.org/0000-0003-2969-3990

References

[1] M. Salamon, N. Reims, M. Böhnel, K. Zarbe, M. Schmitt, N. Uhlmann, and R. Hanke, XXL-CT capabilities for the inspection of modern electric vehicles, International Symposium on Digital Radiology and Computed Tomography – DIR 2019, (2019).
[2] S. Mallat, A Wavelet Tour of Signal Processing: The Sparse Way, Academic Press, 3rd ed., 2009.
[3] C. Christopoulos, A. Skodras, and T. Ebrahimi, The JPEG2000 still image coding system: an overview, IEEE transactions on consumer electronics, 46.4 (2000), pp. 1103–1127.
[4] B. Diederichs, T. Sauer, A. M. Stock, Mathematical aspects of computerized tomography: compression and compressed computing, Fifteenth International Conference Zaragoza-Pau on Mathematics and its Applications, 42 (2019), pp. 79–93.
[5] A. M. Stock, G. Herl, T. Sauer, and J. Hiller, Edge preserving compression of CT scans using wavelets, Insight, 62 (2020), pp. 345–351, https://doi.org/10.1784/insi.2020.62.6.345.
[6] A. Haar, Zur Theorie der orthogonalen Funktionensysteme, Math. Ann., 69 (1910), pp. 331–371.
[7] Y. Meyer, Wavelets – Algorithms and Applications, SIAM, 1993.
[8] L. I. Rudin, S. Osher, and S. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), pp. 259–268.
[9] G. Steidl and J. Weickert, Relations between soft wavelet shrinkage and total variation denoising, in Pattern Recognition. DAGM 2002, L. V. Gool, ed., vol. 2449 of Lecture Notes in Computer Science, Springer, 2002.
[10] M. Welk, G. Steidl, and J. Weickert,
Locally analytic schemes: A link between diffusion filtering and wavelet shrinkage, Appl. Comput. Harmon. Anal., 24 (2008), pp. 195–224, https://doi.org/10.1016/j.acha.2007.05.004.

[11] R. Reisenhofer, S. Bosse, G. Kutyniok, and T. Wiegand, A Haar wavelet-based perceptual similarity index for image quality assessment, Signal Proc. Image Comm., 61 (2018), pp. 33–43.

[12] A. Cohen, R. DeVore, P. Petrushev and H. Xu, Nonlinear Approximation and the Space BV(R²), Amer. J. Math, Ann., 121 (1999), pp. 587–628.

[13] G. G. Lorentz, Approximation of Functions, Chelsea Publishing Company, 1966.

[14] F. Natterer, The Mathematics of Computerized Tomography, John Wiley & Sons, 1986.

[15] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, An introduction to Total Variation for Image Analysis, in Theoretical Foundations and Numerical Methods for Sparse Recovery, M. Fournasier, vol. 9 of Radon Series Comp. Appl. Math., De Gruyter, 2010, pp. 263–340, https://doi.org/10.1515/9783110226157.263.

[16] R. T. Rockafellar, Convex Analysis, Princeton University Press, 1970.