J. Math. Imaging and Vision manuscript No.
(will be inserted by the editor)

Pointwise Besov Space Smoothing of Images

Gregery T. Buzzard · Antonin Chambolle · Jonathan D. Cohen · Stacey E. Levine · Bradley J. Lucier

the date of receipt and acceptance should be inserted later

Abstract We formulate various variational problems in which the smoothness of functions is measured using Besov space semi-norms. Equivalent Besov space semi-norms can be defined in terms of moduli of smoothness or sequence norms of coefficients in appropriate wavelet expansions. Wavelet-based semi-norms have been used before in variational problems, but existing algorithms do not preserve edges, and many result in blocky artifacts. Here we devise algorithms using moduli of smoothness for the $B^1_\infty(L_1(I))$ Besov space semi-norm. We choose that particular space because it is closely related both to the space of functions of bounded variation, $\mathrm{BV}(I)$, that is used in Rudin–Osher–Fatemi (ROF) image smoothing, and to the $B^1_1(L_1(I))$ Besov space, which is associated with wavelet shrinkage algorithms. It contains all functions in $\mathrm{BV}(I)$, which include functions with discontinuities along smooth curves, as well as "fractal-like" rough regions; examples are given in an appendix. Furthermore, it prefers affine regions to staircases, potentially making it a desirable regularizer for recovering piecewise affine data. While our motivations and computational examples come from image processing, we make no claim that our methods "beat" the best current algorithms. The novelty in this work is a new algorithm that incorporates a translation-invariant Besov regularizer that does not depend on wavelets, thus improving on earlier results. Furthermore, the algorithm naturally exposes a range of scales that depends on the image data, noise level, and the smoothing parameter. We also analyze the norms of smooth, textured, and random Gaussian noise data in $B^1_\infty(L_1(I))$, $B^1_1(L_1(I))$, $\mathrm{BV}(I)$ and $L^2(I)$ and their dual spaces. Numerical results demonstrate properties of solutions obtained from this moduli-of-smoothness–based regularizer.

BJL was supported in part by the Office of Naval Research, Contract N00014-91-J-1076, the Institute for Mathematics and its Applications, Minneapolis, USA, and the Simons Foundation (Awards #209418 and #229816). SEL and JDC were supported in part by NSF-DMS 1320829. AC acknowledges the support of the Isaac Newton Institute, Cambridge, and of a grant of the Simons Foundation.

G. T. Buzzard
Department of Mathematics, Purdue University, 150 N. University St., West Lafayette, IN 47907, USA
E-mail: buzzard@purdue.edu

A. Chambolle
CMAP, École Polytechnique, CNRS, 91128 Palaiseau Cedex, France
E-mail: antonin.chambolle@polytechnique.fr

J. D. Cohen
Department of Mathematics and Computer Science, Duquesne University, 440 College Hall, Pittsburgh, PA 15282, USA
Currently at Google Inc.
E-mail: jonathandcohen222@gmail.com

S. E. Levine
Department of Mathematics and Computer Science, Duquesne University, 440 College Hall, Pittsburgh, PA 15282, USA
E-mail: sel@mathcs.duq.edu

B. J. Lucier
Department of Mathematics, Purdue University, 150 N. University St., West Lafayette, IN 47907, USA
E-mail: lucier@purdue.edu

1 Introduction

This paper has several goals.

We formulate various variational problems in which the smoothness of functions is measured using Besov space semi-norms. Equivalent Besov space semi-norms can be defined in terms of moduli of smoothness or sequence norms of coefficients in appropriate wavelet ex-

pansions. Wavelet-based semi-norms have been used before in variational problems; here we devise algorithms using moduli of smoothness.

In particular, we build algorithms around the $B^1_\infty(L_1(I))$ Besov space semi-norm. We chose that particular space because it is closely related both to the space of functions of bounded variation (incorporated in Rudin–Osher–Fatemi (ROF) image smoothing) and to the $B^1_1(L_1(I))$ Besov space, which is associated with wavelet shrinkage algorithms.

While our motivations and computational examples come from image processing, we make no claim that our methods "beat" the best current algorithms in these areas. Perhaps others can improve on our results.

We give further background and motivation in Section 2. Section 3 introduces the modulus-of-smoothness definition of the $B^1_\infty(L_1(I))$ semi-norm and discusses its properties. Section 4 introduces discrete versions of Besov space and BV semi-norms based on moduli of smoothness. The next Section 5 discusses previously-introduced algorithms for minimizing various variational problems and how they can be adapted to our circumstances. Section 6 gives an algorithm for a projection needed in our algorithms. The very brief Section 7 summarizes our main algorithm. We then examine the norms of smooth features, Gaussian noise, and periodic textures in various function spaces in Section 8. We provide computational examples in Section 9. Finally, an Appendix contains a one-parameter family of bounded, self-similar functions whose variation tends to infinity while the $B^1_\infty(L(I))$ norms are bounded.

2 Background and motivation

In this section we discuss some past image smoothing methods and motivations for the approach taken here.

Rudin, Osher, and Fatemi [20] introduced an image smoothing method that is equivalent to the following: given a noisy image f, represented as a function on the unit square $I = [0,1]^2$, and a positive *smoothing parameter* λ, find the function $\tilde{f}$ that achieves the minimum of the functional

$$K(f,\lambda) = \inf_g \Big(\frac{1}{2}\|f-g\|^2_{L_2(I)} + \lambda \|g\|_{\mathrm{BV}(I)} \Big), \tag{1}$$

where $\|g\|_{\mathrm{BV}(I)}$ is the bounded-variation semi-norm of g. (We will not give precise definitions here, see the references for details.) Around the same time, Sauer and Bouman [21] used a discrete version of the $\mathrm{BV}(I)$ semi-norm to regularize a tomography inversion problem.

Not much later, Donoho and Johnstone [15] introduced the notion of *wavelet shrinkage* for image smoothing. Here one calculates an orthogonal wavelet decomposition of the function f,

$$f = \sum_{j\geq 0,\ k\in\mathbb{Z}^2,\ \psi\in\Psi} c_{j,k,\psi}\, \psi_{j,k},$$

where $\psi_{j,k}(x) = 2^k\psi(2^k x - j) = 2^k\psi(2^k(x - j/2^k))$ with 2^k the dyadic scale, $j/2^k$ a dyadic translation, and Ψ a basic set of wavelet functions (often three functions in two dimensions). Given a positive parameter λ, one then smooths f by *shrinking* the wavelet coefficients to calculate

$$\bar{f} = \sum_{j\in\mathbb{Z}^2,\ k\geq 0,\ \psi\in\Psi} S_\lambda(c_{j,k,\psi})\, \psi_{j,k}, \tag{2}$$

where

$$S_\lambda(t) = \begin{cases} t-\lambda, & \lambda \leq t, \\ 0, & -\lambda \leq t \leq \lambda, \\ t+\lambda, & t \leq -\lambda. \end{cases} \tag{3}$$

They derived similar results when using *hard thresholding*

$$T_\lambda(t) = \begin{cases} t, & \lambda \leq t, \\ 0, & -\lambda \leq t \leq \lambda, \\ t, & t \leq -\lambda, \end{cases}$$

instead of $S_\lambda(t)$ to calculate the wavelet coefficients of $\tilde{f}$.

In 1992, DeVore and Lucier [11] related both variational image smoothing as in (1) and wavelet shrinkage as in (2) to *interpolation of function spaces* using so-called (power) K*-functionals*: given two (quasi-)normed function spaces X and Y and any $p, q > 0$, one can consider

$$K(f,\lambda,X,Y) = \inf_g (\|f-g\|^p_X + \lambda\|g\|^q_Y) \tag{4}$$

for any function f and smoothing parameter $\lambda > 0$. One sees that (up to a constant) (1) is of the form (4) with $X = L_2(I)$, $p = 2$, and $Y = \mathrm{BV}(I)$, $q = 1$. It is noted in [11], however, that because Sobolev spaces and Besov spaces have topologies that can be defined by norms of the wavelet coefficients of functions in those spaces, we have that for appropriately chosen λ_k, $k \geq 0$,

$$\bar{f} = \sum_{j\in\mathbb{Z}^2,\ k\geq 0,\ \psi\in\Psi} T_{\lambda_k}(c_{j,k,\psi})\, \psi_{j,k} \tag{5}$$

is, to within a constant, a minimizer of $K(f,\lambda,L_2(I),Y)$ whenever Y is one of a family of Sobolev spaces $W^\alpha(L_2(I))$, $\alpha > 0$, or Besov spaces $B^\alpha_q(L_q(I))$, with $\alpha > 0$ and $1/q = \alpha/2 + 1/2$.

There was speculation in [11] that because $\mathrm{BV}(I)$ satisfies the strict inclusion

$$B^1_1(L_1(I)) \subset \mathrm{BV}(I) \subset B^1_\infty(L_1(I))$$

between the Besov spaces $B^1_1(L_1(I))$ and $B^1_\infty(L_1(I))$, wavelet shrinkage or wavelet thresholding will likely give a solution "similar" to the minimizer of (1). However, because the Besov space $B^1_1(L_1(I))$ does not include functions that are discontinuous across one-dimensional curves, what we shall call *images with edges*, the minimizing image $\tilde{f}$ of

$$\frac{1}{2}\|f-g\|^2_{L_2(I)} + \lambda\|g\|_{B^1_1(L_1(I))},$$

cannot have such discontinuities, i.e., *it cannot have edges*.

Later, Chambolle, DeVore, Lee, and Lucier [7] noted that because

$$\|f\|^2_{L_2(I)} = \sum_{j,k,\psi} |c_{j,k,\psi}|^2$$

for an orthonormal wavelet transform, and the topology of $B^1_q(L_1(I))$ is generated by the wavelet sequence norm

$$\|f\|^q_{B^1_q(L_1(I))} := \sum_{k\geq 0}\left(\sum_{j,\psi}|c_{j,k,\psi}|\right)^q, \quad 0<q<\infty,$$

and

$$\|f\|_{B^1_\infty(L_1(I))} := \sup_{k\geq 0}\sum_{j,\psi}|c_{j,k,\psi}|,$$

when Ψ contains smooth enough functions (smoother than Haar wavelets), one can derive algorithms for *exact* minimizers of

$$\frac{1}{2}\|f-g\|^2_{L_2(I)} + \lambda\|g\|^q_{B^1_q(L_1(I))} \tag{6}$$

for $q=1$ and 2 and

$$\frac{1}{2}\|f-g\|^2_{L_2(I)} + \lambda\|g\|_{B^1_\infty(L_1(I))}. \tag{7}$$

We'll call the latter problem $B^1_\infty(L_1(I))$ smoothing.

It was noted in [7] that when $q=1$, wavelet shrinkage by λ gives the *exact* minimizer of (6). Further algorithms were suggested for $q=2$ and $q=\infty$; basically, the minimizer of (6) and (7) is given by

$$\tilde{f} = \sum_k \sum_{j,k} S_{\lambda_k}(c_{j,k,\psi})\,\psi_{j,k},$$

where now the shrinkage parameter λ_k is not constant, but depends on the *scale* k of the wavelet term $c_{j,k,\psi}\,\psi_{j,k}$. The case $q=\infty$ was also discussed in [16].

Because $B^1_\infty(L_1(I))$ contains $\mathrm{BV}(I)$, $B^1_\infty(L_1(I))$ certainly contains images with edges, and it must contain functions that are *more singular* than functions in $\mathrm{BV}(I)$, i.e., it contains functions with certain fractal-like structures that are not of bounded variation. Indeed, in the appendix we give an example of a one-parameter family of bounded, self-similar, "fractal", functions whose norms in $B^1_\infty(L_1(I))$ are bounded but whose norms in $\mathrm{BV}(I)$ tend to ∞.

$B^1_\infty(L_1(I))$ smoothing should have another nice property. Minimizers of (1) applied to noisy images generally exhibit stair-casing or terrace-like effects because the $\mathrm{BV}(I)$ semi-norm does not distinguish between a staircase or a plane that rises the same distance, while the $B^1_\infty(L_1(I))$ semi-norm of a plane is zero, and the $B^1_\infty(L_1(I))$ semi-norm of a staircase is nonzero. Thus, the terracing effects should not appear in otherwise smooth regions of minimizers of (7).

In practice, however, if one applies the wavelet shrinkage algorithm for (7) to images, one finds that the minimizers do not generally look much different, qualitatively, than minimizers of (6) with $q=1$.

This leads to the natural question about whether other, *equivalent*, norms for $B^1_\infty(L_1(I))$ would give different results. Perhaps such $B^1_\infty(L_1(I))$ minimizers will have not only edges but also fractal-like structures that are not seen in $\mathrm{BV}(I)$ smoothing.

As it happens, Besov spaces have norms defined via moduli of smoothness, what we'll call "pointwise" norms for Besov spaces. In this paper we develop image smoothing algorithms based on the pointwise Besov space norm for the space $B^1_\infty(L_1(I))$. As we'll see, the resulting images have properties that are qualitatively different from the properties of minimizers of both $\mathrm{BV}(I)$ smoothing and wavelet $B^1_\infty(L_1(I))$ smoothing.

Previous work by Bredies, Kunish, and Pock [4] has been motivated by similar considerations. In their variational problem they use what they call the Total Generalized Variation (TGV), which includes, roughly speaking, the variation of higher order derivatives, not just of the function itself. We include comparisons of their techniques with ours in Section 9.

3 Pointwise norms for Besov spaces

We begin by defining the Besov space $B^\alpha_q(L_p(I))$, where $I=[0,1]^2$ is the domain of our images. Roughly speaking, functions in this space have α "derivatives" in $L_p(I)$; the parameter q is useful in measuring finer gradations of smoothness.

The rth-order forward difference at the point $x\in I$ in the direction $h\in\mathbb{R}^2$ of a function $f\colon I\to\mathbb{R}$ is defined for nonnegative integers r by

$$\begin{aligned}\Delta^0 f(x,h) &= f(x),\\ \Delta^{r+1} f(x,h) &= \Delta^r f(x+h,h) - \Delta^r f(x,h).\end{aligned} \tag{8}$$

The difference $\Delta^r f(x,h)$ is defined for those $x\in I$ for which all of $f(x), f(x+h),\ldots,f(x+rh)$ are defined; we'll denote this set by $I_{rh} = \{x\in I \mid x+rh\in I\}$.

For $0 < p \leq \infty$, the rth order modulus of smoothness in $L_p(I)$ is defined by

$$\omega_r(f,t)_p = \sup_{|h|\leq t} \|\Delta^r f(\cdot,h)\|_{L_p(I_{rh})}. \tag{9}$$

Given parameters $0 < \alpha < \infty$ and $0 < p, q \leq \infty$ we choose the integer r such that $r - 1 \leq \alpha < r$ and define the $B^\alpha_q(L_p(I))$ semi-norm as

$$|f|_{B^\alpha_q(L_p(I))} = \left(\int_0^\infty \left[t^{-\alpha} \omega_r(f,t)_p \right]^q \frac{dt}{t} \right)^{1/q}$$

for $0 < q < \infty$ and

$$|f|_{B^\alpha_\infty(L_p(I))} = \sup_{t>0} t^{-\alpha} \omega_r(f,t)_p.$$

The $B^\alpha_q(L_p(I))$ norm we use is then

$$\|f\|_{B^\alpha_q(L_p(I))} = |f|_{B^\alpha_q(L_p(I))} + \|f\|_{L_p(I)}.$$

In particular, the semi-norm for $B^1_\infty(L_1(I))$ is

$$|f|_{B^1_\infty(L_1(I))} = \sup_{t>0} \frac{1}{t} \omega_2(f,t)_1. \tag{10}$$

Combining (10) and (9), we have

$$|f|_{B^1_\infty(L_1(I))} = \sup_{|h|>0} \frac{1}{|h|} \|\Delta^2 f(\cdot,h)\|_{L_1(I_{2h})} \tag{11}$$

and

$$\|f\|_{B^1_\infty(L_1(I))} = \|f\|_{L_1(I)} + |f|_{B^1_\infty(L_1(I))}.$$

Note that we are not dividing by $|h|^2$; if we were, the functions in the space would consist of functions whose second derivatives are measures. More to the point, because

$$\omega_2(f,nt)_1 \leq n^2 \omega_2(f,t)_1,$$

(see Chapter 2 of [10]) we have

$$\frac{1}{(nt)^2} \omega_2(f,nt)_1 \leq \frac{1}{t^2} \omega_2(f,t)_1,$$

so

$$\sup_{t>0} \frac{1}{t^2} \omega_2(f,t)_1 = \lim_{t\to 0} \frac{1}{t^2} \omega_2(f,t)_1.$$

The supremum in (10) is generally not taken in the limit as $t \to 0$, however. For example, if f has two continuous derivatives in $L_1(I)$, then as $\delta \to 0$, $|\delta h|^{-2} \Delta^2 f(x, \delta h) \to D^2_h f(x)$, the second directional derivative of f in the direction h at x, and

$$\frac{1}{t} \omega_2(f,t)_1 = t\Big(\frac{1}{t^2} \omega_2(f,t)_1 \Big) \to 0$$

as $t \to 0$, as the quantity in parentheses is finite. So computations with the $B^1_\infty(L_1(I))$ semi-norm must take into account all the values of $|h| > 0$, not just the limit as $|h| \to 0$. On the other hand, we have that $\omega_r(f,t)_p \leq 2^r \|f\|_{L_p(I)}$ for $1 \leq p \leq \infty$, so $t^{-1}\omega_2(f,t)_1 \to 0$ as $t \to \infty$. So the supremum in (11) is found on some finite interval $[0, t^*]$.

Note also that $\mathrm{BV}(I)$, the space of functions of bounded variation on I is contained in $B^1_\infty(L_1(I))$, because an equivalent semi-norm for $\mathrm{BV}(I)$ is

$$|f|_{\mathrm{BV}(I)} = \sup_{t>0} \frac{1}{t} \omega_1(f,t)_1 = \lim_{t\to 0} \frac{1}{t} \omega_1(f,t)_1,$$

and because $\omega_2(f,t)_1 \leq 2\omega_1(f,t)_1$ by (8), functions in $B^1_\infty(L_1(I))$ can be "more singular" than functions in $\mathrm{BV}(I)$.

The supremum in (11) is taken over all directions h, and we cannot compute over *all* directions. For effective computations, we note as a consequence of a result of [14] that an equivalent semi-norm $|f|_{B^1_\infty(L_1(I))}$ arises if we replace the set of all directions h in (11) by multiples of any three of the directions $(1,0)$, $(0,1)$, $(1,1)$, and $(1,-1)$. For the sake of symmetry, we use all four directions.

4 Discrete norms, spaces, and variational problems

We consider a number of variational approaches to noise removal and image smoothing in general.

The first model uses the $L_2(I)$ norm as a measure of the difference between the original and smooth function: Given a function f defined on I and a parameter $\lambda > 0$, find the minimizer over all g with $f - g \in L_2(I)$ and $g \in B^1_\infty(L_1(I))$ of

$$E(g) = \frac{1}{2\lambda} \|f - g\|^2_{L_2(I)} + |g|_{B^1_\infty(L_1(I))}. \tag{12}$$

One could replace the $L_2(I)$ norm in the previous expression with an $L_1(I)$ norm, and so come up with the second model: Given a function f defined on I and a parameter $\lambda > 0$, find the minimizer over all g with $f - g \in L_1(I)$ and $g \in B^1_\infty(L_1(I))$ of

$$E(g) = \frac{1}{\lambda} \|f - g\|_{L_1(I)} + |g|_{B^1_\infty(L_1(I))}. \tag{13}$$

We discuss algorithmic issues related to this model, but leave discussion of computational results to others.

The existence of minimizers of energies (12) and (13) is not a difficult issue. For (12), one just needs to observe that a minimizing sequence has weakly converging subsequences in $L_2(I)$ and that both terms of the energy are lower semicontinuous with respect to this convergence. For (13), minimizing sequences are bounded in $L_1(I)$ and in theory their limits could be measures, however the Besov semi-norm term prevents this from

happening. In fact, sublevel sets of energy (13) are compact subsets of $L_p(I)$ for any $p < 2$; this is easily seen by adapting the arguments of the proof of Theorem 7.1 in [12].

One could wonder whether this extends to linear inverse problems (recover g from $f = Kg + \text{noise}$, where K is a linear operator), for which one would for instance minimize energies such as

$$E(g) = \frac{1}{2\lambda}\|f - Kg\|^2_{L_2(I)} + |g|_{B^1_\infty(L_1(I))}. \tag{14}$$

Again, if one assumes that K is continuous from $L_p(I)$ to $L_2(I)$ for some $1 \le p < 2$ (in which case it is also weakly continuous), existence can be deduced by the classical direct method.

4.1 Discrete L_p and Besov norms

For computations, we want to extend the idea of spaces, semi-norms, etc., from the continuous domain to the discrete domain. So for discrete functions f, g defined for $\mathcal{I} = \{i = (i_1, i_2) \mid 0 \le i_1, i_2 < N\}$, $h = 1/N$, we define a discrete inner product,

$$\langle f, g\rangle = \sum_{i\in\mathcal{I}} f_i g_i \, h^2,$$

and the discrete $L_p(\mathcal{I})$ norm for $1 \le p < \infty$

$$\|g\|^p_{L_p(\mathcal{I})} = \sum_{i\in\mathcal{I}} |g_i|^p \, h^2.$$

We need a discrete $B^1_\infty(L_1(I))$ semi-norm. For notational convenience we introduce the translation operator $\mathcal{T}_j$ defined for any $j = (j_1, j_2)$ by $(\mathcal{T}_j g)_i = g_{i+j}$.

Then for any pixel offset j we define discrete divided second differences

$$\nabla^2_j g = \frac{1}{|jh|}(\mathcal{T}_{2j} g - 2\mathcal{T}_j g + g),$$

which is defined only for $i \in \mathcal{I}^{2j} = \{i \in \mathcal{I} \mid i + 2j \in \mathcal{I}\}$ (which will be empty if $|j|$ is large enough). Pointwise, we have

$$\nabla^2_j g_i = \frac{(\mathcal{T}_{2j} g)_i - 2(\mathcal{T}_j g)_i + g_i}{|jh|} = \frac{g_{i+2j} - 2g_{i+j} + g_i}{|jh|}.$$

Following (11), our discrete $B^1_\infty(L_1(\mathcal{I}))$ semi-norm is then

$$|g|_{B^1_\infty(L_1(\mathcal{I}))} = \max_{k>0} \max_{\substack{j=(k,0),(0,k),\\(k,k),(k,-k)}} \sum_{i\in\mathcal{I}^{2j}} |\nabla^2_j g_i| \, h^2.$$

The sum is zero if $\mathcal{I}^{2j}$ is empty. As shorthand we write

$$\begin{aligned}\nabla^2 g &= \{\nabla^2_j g \mid j = (k,0),\ (0,k),\ (k,k),\ (k,-k);\ k > 0\}\\ &= \{\nabla^2_j g_i \in \mathbb{R} \mid j = (k,0),\ (0,k),\ (k,k),\ (k,-k);\\ &\qquad k > 0;\ i \in \mathcal{I}^{2j}\}.\end{aligned}$$

If we introduce a sequence of scalar fields

$$\begin{aligned}p &= \{p^j \mid k > 0;\ j = (k,0),\ (0,k),\ (k,k),\ (k,-k)\}\\ &= \{p^j_i \in \mathbb{R} \mid j = (k,0),\ (0,k),\ (k,k),\ (k,-k);\\ &\qquad k > 0;\ i \in \mathcal{I}^{2j}\},\end{aligned}$$

with the inner product

$$\langle p, q\rangle = \sum_{k\ge1} \sum_{\substack{j=(k,0),(0,k),\\(k,k),(k,-k)}} \sum_{i\in\mathcal{I}^{2j}} p^j_i \cdot q^j_i \, h^2,$$

we can consider the operator adjoint in this inner product to $\nabla^2 g$, which we denote by $\nabla^2 \cdot p$. One sees immediately that

$$\nabla^2 \cdot p = \sum_{k>0} \sum_{\substack{j=(k,0),(0,k),\\(k,k),(k,-k)}} \frac{\mathcal{T}_{-2j} p^j - 2\mathcal{T}_{-j} p^j + p^j}{|jh|},$$

which is defined now wherever *any* of $\mathcal{T}_j p^j_i$, etc., are defined. (Equivalently, assume that each scalar field p^j is extended as zero outside its domain.) Thus

$$\begin{aligned}\nabla^2 \cdot p_i &= \sum_{k>0} \sum_{\substack{j=(k,0),(0,k),\\(k,k),(k,-k)}} \frac{(\mathcal{T}_{-2j} p^j)_i - 2(\mathcal{T}_{-j} p^j)_i + p^j_i}{|jh|}\\ &= \sum_{k>0} \sum_{\substack{j=(k,0),(0,k),\\(k,k),(k,-k)}} \frac{p^j_{i-2j} - 2p^j_{i-j} + p^j_i}{|jh|}.\end{aligned}$$

Then we can write

$$\begin{aligned}|g|_{B^1_\infty(L_1(\mathcal{I}))} &= \max_{k>0} \max_{\substack{j=(k,0),(0,k),\\(k,k),(k,-k)}} \sum_{i\in\mathcal{I}^{2j}} |\nabla^2_j g_i| \, h^2\\ &= \sup_{\|\{\|p^j\|_{L_\infty(\mathcal{I}^{2j})}\}\|_{\ell_1(P)}\le 1} \big\langle \nabla^2 g, p\big\rangle\\ &= \sup_{\|\{\|p^j\|_{L_\infty(\mathcal{I}^{2j})}\}\|_{\ell_1(P)}\le 1} \big\langle g, \nabla^2 \cdot p\big\rangle,\end{aligned}$$

where the ℓ_1 norm is taken over the (finite) set

$$\begin{aligned}P = \{j \mid j = (k,0),\ (0,k),\ (k,k),\ (k,-k);\\ k > 0;\ \mathcal{I}^{2j} \ne \emptyset\}.\end{aligned}$$

We note that our formulation doesn't incorporate any notion of "boundary conditions." Each scalar field p^j is defined on its own distinct subdomain $\mathcal{I}^{2j}$ of $\mathcal{I}$. Even for a fixed k, the four scalar fields for $j = (k,0)$, $(0,k)$, (k,k), and $(k,-k)$ are defined on different domains, so there is no "vector field" on a common domain as is usually defined for discrete $\mathrm{BV}(I)$ (sometimes two-dimensional, sometimes four or more—see, e.g., [8]).

Our discrete version of (12) is then: Given a discrete function f and a positive parameter λ, find the minimizer over all g of the functional

$$E^h(g) = \frac{1}{2\lambda}\|f-g\|^2_{L_2(\mathcal{I})} + |g|_{B^1_\infty(L_1(\mathcal{I}))}. \tag{15}$$

Similarly, our discrete version of (13) is: Given a discrete function f and a positive parameter λ, find the minimizer over all g of the functional

$$E^h(g) = \frac{1}{\lambda}\|f-g\|_{L_1(\mathcal{I})} + |g|_{B^1_\infty(L_1(\mathcal{I}))}. \tag{16}$$

5 Discrete algorithms

We can formulate our discrete variational problems (15) and (16) as saddle-point problems that can readily be solved by algorithms from [6]. In this setting, we consider two Hilbert spaces X and Y and the bounded linear operator $K\colon X \to Y$ with the usual norm $\|K\| = \sup_{\|g\|\le 1}\|Kg\|$. We consider the general saddle-point problem

$$\min_{g\in X}\max_{p\in Y}(\langle Kg,p\rangle + G(g) - F^*(p)), \tag{17}$$

where G and F^* are lower semicontinuous (l.s.c.), proper, convex functions from X and Y (respectively) to $[0,+\infty]$; F^* is itself the convex conjugate of a convex l.s.c. function F. This saddle-point problem is a primal-dual formulation of the nonlinear primal problem

$$\min_{g\in X}(F(Kg)+G(g))$$

and of the corresponding dual problem

$$\max_{p\in Y} -(G^*(-K^*p)+F^*(p)),$$

where $K^*\colon Y \to X$ is the adjoint of K. An overview of these concepts from convex analysis can be found in [19]. We use the notation ∂F for the (multi-valued) subgradient of a l.s.c. function F

A number of algorithms are given in [6] to compute approximate solutions to (17); whether a particular algorithm is appropriate depends on properties of F^* and G. That paper also includes comparisons between the algorithms introduced there and competing algorithms. Our focus here is not on that comparison, but just to indicate that relatively efficient algorithms exist to solve the discrete minimization problems in the following sections.

Algorithm 1 in [6] is appropriate for completely general F and G; a special case can be stated as follows.

Algorithm 1 General F and G

1. Let $L = \|K\|$. Choose positive τ and σ with $\tau\sigma L^2 < 1$, $(g^0,p^0)\in X\times Y$ and set $\bar g^0 = g^0$.
2. For $n\ge 0$, update g^n, p^n, and $\bar g^n$ as follows:

$$\begin{cases} p^{n+1} = (I+\sigma\partial F^*)^{-1}(p^n+\sigma K\bar g^n),\\ g^{n+1} = (I+\tau\partial G)^{-1}(g^n-\tau K^*p^{n+1}),\\ \bar g^{n+1} = 2g^{n+1}-g^n. \end{cases}$$

Then there exists a saddle point (g^*,p^*) such that $g^n\to g^*$ and $p^n\to p^*$. This algorithm exhibits $O(1/N)$ convergence after N steps.

Algorithm 2 in [6] is appropriate when G (or F^*) is uniformly convex, which means that there exists $\gamma>0$ such that for any g in the domain of ∂G, $w\in\partial G(g)$, and $g'\in X$,

$$G(g') \ge G(g) + \langle w, g'-g\rangle + \frac{\gamma}{2}\|g-g'\|^2.$$

Algorithm 2 can then be written as follows.

Algorithm 2 G (or F^*) is uniformly convex

1. Let $L = \|K\|$. Choose positive τ_0 and σ_0 with $\tau_0\sigma_0L^2 \le 1$, $(g^0,p^0)\in X\times Y$ and set $\bar g^0 = g^0$.
2. For $n\ge 0$, update g^n, p^n, $\bar g^n$ and σ_n, τ_n, θ_n as follows:

$$\begin{cases} p^{n+1} = (I+\sigma_n\partial F^*)^{-1}(p^n+\sigma_nK\bar g^n),\\ g^{n+1} = (I+\tau_n\partial G)^{-1}(g^n-\tau_nK^*p^{n+1}),\\ \theta_n = 1/\sqrt{1+2\gamma\tau_n},\ \tau_{n+1}=\theta_n\tau_n,\ \sigma_{n+1}=\sigma_n/\theta_n,\\ \bar g^{n+1} = g^{n+1}+\theta_n(g^{n+1}-g^n). \end{cases}$$

Then there exists a unique saddle point (g^*,p^*) such that $g^n\to g^*$ and $p^n\to p^*$.

These algorithms can directly be applied to solve problems (15) and (16). For example, we can rewrite (15) as

$$E^h(g) = \sup_p\Big(\langle\nabla^2 g,p\rangle + \frac{1}{2\lambda}\|f-g\|^2_{L_2(\mathcal{I})} - \delta(p)\Big),$$

where

$$\delta(p) = \begin{cases} 0, & \text{if } \|\{\|p^j\|_{L_\infty(\mathcal{I}^{2j})}\}\|_{\ell_1(P)}\le 1,\\ \infty, & \text{otherwise.}\end{cases}$$

This matches formula (17) with

$$Kg = \nabla^2 g,\ K^*p = \nabla^2\cdot p,$$
$$G(g) = \frac{1}{2\lambda}\|f-g\|^2_{L_2(\mathcal{I})},\ \text{and } F^*(p) = \delta(p).$$

We know that G is uniformly convex with $\gamma=\lambda^{-1}$, so Algorithm 2 applies. In fact, we have that the iterates

g^n and K^*p^n in Algorithm 2 converge to a pair u and $\nabla^2 \cdot p$ that satisfy

$$u = f - \lambda \nabla^2 \cdot p.$$

For (16), we can rewrite

$$E^h(g) = \sup_p \Big(\langle \nabla^2 g, p \rangle + \frac{1}{\lambda} \|f - g\|_{L_1(\mathcal{I})} - \delta(p) \Big).$$

This matches formula (17) with

$$Kg = \nabla^2 g, \;\; K^* p = \nabla^2 \cdot p,$$
$$G(g) = \frac{1}{\lambda} \|f - g\|_{L_1(\mathcal{I})}, \text{ and } F^*(p) = \delta(p).$$

Neither F^* nor G are uniformly convex, so in contrast to problem (15), where we can use Algorithm 2, we are limited to using Algorithm 1 for (16).

The function F^* is the same for both problems, and for any $\sigma > 0$ we have $(I + \sigma \partial F^*)^{-1}(p) = P_{\mathcal{K}}(p)$, the projection of p onto the set

$$\mathcal{K} = \{ p \mid \| \{ \|p^j\|_{L_\infty(\mathcal{I}^{2j})} \} \|_{\ell_1(P)} \le 1 \}. \tag{18}$$

Although this projection is nontrivial, it is reasonable to compute; we give an algorithm for this operation in the next section.

For both (15) and (16), computing $(I + \tau \partial G)^{-1}$ involves simple, pointwise operations. For the first case we have

$$(I + \tau \partial G)^{-1}(g)_i = \frac{g_i + (\tau/\lambda) f_i}{1 + \tau/\lambda},$$

while for (16) we have

$$(I + \tau \partial G)^{-1}(g)_i = S_{\tau/\lambda}(g_i - f_i) + f_i,$$

where $S_{\tau/\lambda}$ is given by (3).

To apply Algorithms 1 and 2 we need a bound on $L = \|K\|$. We have (using $(a+2b+c)^2 \le 6(a^2+b^2+c^2)$)

$$\sum_{i \in \mathcal{I}^{2j}} |\nabla_j^2 g_i|^2 \, h^2 \le \frac{6}{|jh|^2} \sum_{i \in \mathcal{I}^{2j}} \Big(g_{i+j}^2 + g_i^2 + g_{i-j}^2 \Big) h^2$$
$$\le \frac{18}{|jh|^2} \|g\|_{L_2(\mathcal{I})}^2.$$

Because, for each $k > 0$, $j = (k,0)$, $(0,k)$, (k,k), and $(k,-k)$, for which $|j| = k$, k, $\sqrt{2}k$, and $\sqrt{2}k$, respectively, we have

$$\|\nabla^2 g\|^2 := \langle \nabla^2 g, \nabla^2 g \rangle \le \sum_{k>0} \frac{54}{k^2 h^2} \|g\|_{L_2(\mathcal{I})}^2$$
$$= \frac{9\pi^2}{h^2} \|g\|_{L_2(\mathcal{I})}^2.$$

So $L = \|K\| \le 3\pi/h \approx 9.4248/h$.

Theorem 2 of [6] gives an a priori error bound for the iterates of Algorithm 2 of the following form: Choose $\tau_0 > 0$, $\sigma_0 = 1/(\tau_0 L^2)$. Then for any $\epsilon > 0$ there is a N_0 (depending on ϵ and $\gamma\tau_0$) such that for any $N > N_0$

$$\|\hat{g} - g^N\|^2 \le \frac{1+\epsilon}{N^2} \Big(\frac{\|\hat{g} - g^0\|^2}{\gamma^2 \tau_0^2} + \frac{L^2}{\gamma^2} \|\hat{p} - p^0\|^2 \Big).$$

Computational results are given in [6] that indicate that N_0 is quite moderate (≈ 100) in many situations.

In our case, we just want $\|\hat{g} - g^N\| \lessapprox \varepsilon$ greyscales. So we take

$$\frac{\|\hat{g} - g^0\|^2}{\gamma^2 \tau_0^2} \le \varepsilon^2$$

or (because $\|\hat{g} - g^0\| \le 256$ greyscales, as pixel values lie between 0 and 256)

$$\frac{256}{\gamma \varepsilon} \le \tau_0.$$

Then we take

$$\frac{1}{N^2} \frac{L^2}{\gamma^2} \|\hat{p} - p^0\|^2 \le \varepsilon^2,$$

but $p^0 = 0$ and $\|\hat{p}\| \le 1$, so

$$L/(\gamma\varepsilon) \le N$$

We choose the least τ_0 and N that satisfy these two inequalities.

6 Computing $P_{\mathcal{K}}$

It remains to give the algorithm for $P_{\mathcal{K}}$, where $\mathcal{K}$ is given by (18). The problem can be formulated as: Given the sequence of scalar fields $p = \{p_i^j \mid i \in \mathcal{I}^{2j},\ j \in P = \{(k,0), (0,k), (k,k), (k,-k) \mid k > 0\}\}$ find $q = \{q_i^j\}$ that minimizes

$$\frac{1}{2} \sum_{i \in \mathcal{I}^{2j}, j \in P} |p_i^j - q_i^j|^2 \text{ with } \sum_{j \in P} \max_{i \in \mathcal{I}^{2j}} |q_i^j| \le 1. \tag{19}$$

The set P has at most $2N$ elements j ($0 < k < N/2$, four js for each k), but in practice we compute p_i^j for only a finite number of scales $k \le k_{\max}$, where $k_{\max}$ is defined in Section 7.

Instead of directly solving (19), we introduce variables μ^j, $j \in P$, and search for $q = \{q_i^j\}$ that minimizes

$$\frac{1}{2} \sum_{i,j} |p_i^j - q_i^j|^2 \text{ with } |q_i^j| \le \mu^j,$$
$$\sum_j \mu^j \le 1, \text{ and } \mu^j \ge 0. \tag{20}$$

If $\mu = \{\mu^j\}$ is known, we know that at a minimizer, q satisfies

$$q_i^j = \begin{cases} \mu^j \dfrac{p_i^j}{|p_i^j|}, & \mu^j < |p_i^j|, \\ p_i^j, & \mu^j \geq |p_i^j|. \end{cases} \tag{21}$$

Therefore, we aim to find appropriate μ, a problem that can be simplified even further. To this end, denote the characteristic function on the set $\{i \in \mathcal{I}^{2j} \mid \mu^j < |p_i^j|\}$ as

$$[\mu^j < |p_i^j|] := \begin{cases} 1, & \text{if } \mu^j < |p_i^j|, \\ 0, & \text{otherwise,} \end{cases}$$

so at a minimum we have

$$\begin{aligned} \frac{1}{2}\sum_{i,j} |p_i^j - q_i^j|^2 &= \frac{1}{2}\sum_{i,j} \left| p_i^j - p_i^j \frac{\mu^j}{|p_i^j|} \right|^2 [\mu^j < |p_i^j|] \\ &= \frac{1}{2}\sum_{i,j} |p_i^j|^2 \left(1 - \frac{\mu^j}{|p_i^j|}\right)^2 [\mu^j < |p_i^j|]. \end{aligned}$$

The Karush–Kuhn–Tucker (KKT) theorem states that there are nonnegative Lagrange multipliers η^j and ν such that the projector $P_{\mathcal{K}}p$ can be computed from the μ that minimizes

$$\frac{1}{2}\sum_{i,j} |p_i^j|^2 \left(1 - \frac{\mu^j}{|p_i^j|}\right)^2 [\mu^j < |p_i^j|] - \sum_j \mu^j \eta^j - \left(1 - \sum_j \mu^j\right)\nu.$$

Note that for each i, j the expression

$$(|p_i^j| - \mu^j)^2 \, [\mu^j < |p_i^j|]$$

is continuously differentiable with respect to μ^j, so the previous displayed expression is continuously differentiable with respect to μ.

Differentiating with respect to μ^j gives

$$\sum_i |p_i^j|^2 \left(1 - \frac{\mu^j}{|p_i^j|}\right) \frac{-1}{|p_i^j|} [\mu^j < |p_i^j|] - \eta^j + \nu = 0,$$

or

$$\sum_i (|p_i^j| - \mu^j)[\mu^j < |p_i^j|] + \eta^j - \nu = 0. \tag{22}$$

The sum is nonnegative and the KKT theorem says that $\nu > 0 \Rightarrow \sum_j \mu^j = 1$ and $\eta^j > 0 \Rightarrow \mu^j = 0$.

Thus, if $\nu = 0$, then (22) implies that $\eta^j = 0$ for all j and $\mu^j \geq |p_i^j|$ for all i, j, so (21) implies $q = p$.

If $\nu > 0$ then $\sum_j \mu^j = 1$. If $\eta^j > 0$ for some j, then $\mu^j = 0$ and so $q_i^j = 0$ for all i by (21). If $\eta^j = 0$ then (22) implies

$$\sum_i (|p_i^j| - \mu^j)[\mu^j < |p_i^j|] = \nu. \tag{23}$$

Note that as ν increases, μ^j decreases, so each of the μ^j can be considered a function of ν: $\mu^j = \mu^j(\nu)$.

If we set $A(\mu) = \{i \mid |p_i^j| > \mu\}$, the left-hand side of (23) can be written for general μ as

$$G(\mu) = \sum_{A(\mu)} (|p_i^j| - \mu) = \sum_{A(\mu)} |p_i^j| - |A(\mu)| \times \mu,$$

where $|A(\mu)|$ is the number of items in $A(\mu)$.

G is a continuous, decreasing, piecewise linear function of μ. In Algorithm 4 we describe how to find the solution μ^j of $G(\mu) = \nu$. We do this by precomputing a table of values $G(|p_i^j|)$ for $i \in \mathcal{I}^{2j}$. For each ν, we use binary search to find adjacent values in this table for which the linear function $G(\mu)$, restricted to lie between these two adjacent values, takes on the value ν precisely once. Linear interpolation gives the proper value of μ^j. Note that μ^j is a continuous, piecewise linear function of ν, being the inverse of G.

Once Algorithm 4 is established, the projection $(P_{\mathcal{K}}p)^j$ can be computed using Algorithm 3. The function $F(\nu)$ in Algorithm 3 is, yet again, a continuous, piecewise linear function of ν, being a sum of functions of the same type. Except in the trivial case where F is a first-degree polynomial, it does not have any higher-order smoothness. To find the unique zero of F, we employ an algorithm from [18] that requires the evaluation of no derivatives, yet which has second-order convergence properties for smooth F.

Algorithm 3 Computing the projection $(P_{\mathcal{K}}p)^j$

1. If $\|\{\|p^j\|_{L_\infty(\mathcal{I}^{2j})}\}\|_{\ell_1(P)} \leq 1$, then $P_{\mathcal{K}}p = p$.
2. Otherwise, do the following:
   (a) Find a root $\bar{\nu}$ of
   $$F(\nu) = 1 - \sum_{j \in P} \mu^j(\nu),$$
   which guarantees the constraint $\sum_j \mu^j = 1$. This can be done in a number of ways. We chose to use the algorithm from [18], and this part of the algorithm takes an immeasurably small amount of time.
   (b) With this value of $\bar{\nu}$ we can calculate $\mu^j = \mu^j(\bar{\nu})$ for $j \in P$ and subsequently calculate $q^j = (P_{\mathcal{K}}p)^j$ using (21).

7 The main iteration

The iterations of Algorithm 1 or 2 must estimate $k_{\max}$ as n increases. Here we have no theoretical justification for our argument, and possibly in pathological cases we could get it wrong.

Algorithm 4 Computing $\mu^j(\nu)$ for any $\nu > 0$

1. Sort the numbers $\{|p_i^j| \mid i \in \mathcal{I}^{2j}\}$ into nonincreasing order and add a single element 0 at the end; call this sequence $\{a_\ell^j\}$. There are at most N^2 elements of $\{a_\ell^j\}$, so this takes at most $O(N^2 \log N)$ time.
2. Calculate the sequence $\{b_\ell^j\}$ by

$$b_\ell^j = \sum_{k<\ell} a_k^j - \ell \times a_\ell^j.$$

   For each j, this takes at most $O(N^2)$ time. Note that b_ℓ^j increases as ℓ increases, and if $\nu = b_\ell^j$, then $\mu^j(\nu) = a_\ell^j$, i.e., the pairs (b_ℓ^j, a_ℓ^j) lie on the graph of the one-to-one relation (23) between μ^j and ν.
3. Given any $\nu \geq 0$, we have the following mechanism for calculating $\mu^j(\nu)$:
   (a) If $\nu \geq b_{N^2}^j$, set $\mu^j(\nu) = 0$.
   (b) Otherwise, use binary search to calculate an index ℓ such that $b_\ell^j \leq \nu < b_{\ell+1}^j$. Since the relationship between μ^j and ν in (23) is linear between any two consecutive pairs (b_ℓ^j, a_ℓ^j) and $(b_{\ell+1}^j, a_{\ell+1}^j)$, we can interpolate linearly between (b_ℓ^j, a_ℓ^j) and $(b_{\ell+1}^j, a_{\ell+1}^j)$ to find the point $(\nu, \mu^j(\nu))$:

$$\mu^j(\nu) = a_\ell^j + \frac{\nu - b_\ell^j}{b_{\ell+1}^j - b_\ell^j}(a_{\ell+1}^j - a_\ell^j).$$

   This takes $O(\log N)$ time.

1. We pick an initial value of $k_{\max}$ (typically 4) and set p^j to be zero scalar fields for $j \in P$.
2. We iterate Algorithm 1 or 2 a fixed number of times.
3. We have a parameter $z_{\min}$ (we've used 4), and we want the scalar fields p^j, $j = (k, 0)$, $(0, k)$, (k, k), $(k, -k)$, to be zero for the $z_{\min}$ largest values of k. Heuristically, if we have zero scalar fields p^j for the $z_{\min}$ largest values of k, then we think it unlikely that at a minimizer of (15) or (16) any scalar fields p^j with $k > k_{\max}$ will be nonzero. So we adjust $k_{\max}$ so that the all scalar fields p^j for the $z_{\min}$ largest values of k are zero.
4. We then decide whether to stop the iteration or go back to step 2.

8 Norms of image features

Generally speaking, image features are classified as: smooth regions; edges; textures; and noise. We follow the framework of [3] to compute some sample solutions to the following problem.

One goal of image processing is to separate an image into its various features or components. In this section we show that one can separate smooth regions, texture, and noise by looking at the asymptotic behavior of various norms applied to these features. Because edge features have components at all wavelengths, parts of edges may appear as texture, noise, or smooth regions.

In this section we operate on the (two-dimensional) unit interval $[0, 1]^2$ with $N \times N$ images, with $N = 2^K$ for some K reasonably large. We set $h = 1/N$ and deal with the weighted $L_2(I)$ norm $\|f\|^2 = \sum_i f_i^2 h^2$, where f_i is the value of the pixel at $i = (i_1, i_2)$.

8.1 Smooth features

Basically, if a function is smooth (a broad Gaussian, say), then all its norms, in $L_2(I)$, $B_\infty^1(L_1(I))$, $B_\infty^1(L_1(I))^*$, or other spaces, will be of moderate size. (If X is a Banach space, we denote its dual by X^*.) If we're to distinguish among smooth features, noise, and textures, then we'll need to find spaces for which the norms of noise and textures either tend to infinity or tend to zero as a parameter (the frequency of a texture, the size of pixels, etc.) changes.

8.2 Gaussian noise

We consider here issues that are similar to those considered in Section 3.6 of [2].

We begin by studying the behavior of Gaussian pixel noise. Then the pixel values $f_i = \epsilon_i$ are i.i.d. Gaussian random variables with mean zero and variance σ^2, and

$$E(\|\epsilon\|_{L_2(I)}^2) = \sum_i E(\epsilon_i^2)\, h^2 = \sigma^2,$$

since there are N^2 pixels and $N^2 h^2 = 1$.

We consider the orthonormal wavelet transform of ϵ with wavelets $\psi_{j,k}(x) = 2^k \psi(2^k x - j)$, $k \geq 0$, $j = (j_1, j_2)$, $\psi \in \Psi$, that are smoother than Haar (so that the following norm equivalences work). Then we can write

$$\epsilon = \sum_k \sum_{j,\psi} \eta_{j,k,\psi} \psi_{j,k}.$$

Because the transform is orthonormal (so $\|\psi_{j,k}\|_{L_2(I)} = 1$), the $\eta_{j,k,\psi}$ are i.i.d. Gaussian with mean zero and

$$\sigma^2 = E(\|\epsilon\|^2) = \sum_k \sum_{j,\psi} E(\|\eta_{j,k,\psi}\psi_{j,k}\|^2) = N^2 E(\eta_{j,k,\psi}^2)$$

for any $\eta_{j,k,\psi}$, so $E(\eta_{j,k,\psi}^2) = \sigma^2/N^2$. For similar reasons,

$$E(|\eta_{j,k,\psi}|) = C_1 \frac{\sigma}{N} \text{ and } \operatorname{Var}(|\eta_{j,k,\psi}|) = C_2 \frac{\sigma^2}{N^2} \qquad (24)$$

for some known C_1 and C_2.

We also note that if $X_1, X_2, \ldots, X_n$ are i.i.d. $N(0,1)$ random variables, then for large n [9, p. 374],

$$\begin{aligned} E(\max_i X_i) &\approx \sqrt{\log\left(\frac{n^2}{2\pi\log\left(\frac{n^2}{2\pi}\right)}\right)} \\ &= \sqrt{2\log n - \log\left(2\pi\log\left(\frac{n^2}{2\pi}\right)\right)} \\ &\approx \sqrt{2\log n}. \end{aligned} \tag{25}$$

Wavelet-based Besov semi-norms have the following characterizations:

$$\begin{aligned} |\epsilon|_{B^1_1(L_1)} &= \sum_{j,k,\psi} |\eta_{j,k,\psi}| \text{ and} \\ |\epsilon|_{B^1_\infty(L_1)} &= \sup_k \sum_{j,\psi} |\eta_{j,k,\psi}|, \end{aligned}$$

and the norms in the dual spaces (for finite images) are

$$\begin{aligned} |\epsilon|_{B^1_1(L_1)^*} &= \sup_{j,k,\psi} |\eta_{j,k,\psi}| \text{ and} \\ |\epsilon|_{B^1_\infty(L_1)^*} &= \sum_k \sup_{j,\psi} |\eta_{j,k,\psi}|. \end{aligned}$$

Now we take expectations. Because $\sup_{j,\psi} |\eta_{j,k,\psi}| = \max(\sup_{j,\psi} \eta_{j,k,\psi}, -\inf_{j,\psi} \eta_{j,k,\psi})$, we use formulas (24) and (25) to see

$$\begin{aligned} E(|\epsilon|_{B^1_1(L_1)}) &= N^2 \times C_1 \frac{\sigma}{N} = C_1 N \sigma, \\ E(|\epsilon|_{B^1_1(L_1)^*}) &= E(\sup_{j,k,\psi} |\eta_{j,k,\psi}|) \approx \sqrt{4\log N}\frac{\sigma}{N}. \end{aligned}$$

We note that there are 3×2^{2k} coefficients $\eta_{j,k\psi}$ at level k, and because the $\eta_{j,k,\psi}$ are independent, $\sum_{j,\psi} |\eta_{j,k,\psi}|$ is asymptotically a Gaussian random variable with

$$\begin{aligned} E\left(\sum_{j,\psi} |\eta_{j,k,\psi}|\right) &= 3 \times 2^{2k} C_1 \frac{\sigma}{N} \text{ and} \\ \mathrm{Var}\left(\sum_{j,\psi} |\eta_{j,k,\psi}|\right) &= 3 \times 2^{2k} C_2 \frac{\sigma^2}{N^2}. \end{aligned}$$

When $k = K-1$, a maximum, we have $2^{2k} = N^2/4$ and

$$\begin{aligned} E\left(\sum_{j,\psi} |\eta_{j,k,\psi}|\right) &= \frac{3}{4} C_1 \sigma N \text{ and} \\ \mathrm{Var}\left(\sum_{j,\psi} |\eta_{j,k,\psi}|\right) &= \frac{3}{4} C_2 \sigma^2. \end{aligned}$$

When $k = K-2$, we get the same expressions divided by 4. So the mean when $k = K-1$ is four times the mean when $k = K-2$, and that difference is roughly N times the standard deviation of the two random variables; a similar argument holds for smaller k.

Thus

$$E(|\epsilon|_{B^1_\infty(L_1)}) = E\left(\sup_k \sum_{j,\psi} |\eta_{j,k,\psi}|\right)$$

will most likely occur for $k = K-1$, so

$$E(|\epsilon|_{B^1_\infty(L_1)}) \approx \frac{3}{4} C_1 \sigma N.$$

Finally, we want to estimate

$$|\epsilon|_{B^1_\infty(L_1)^*} = \sum_k \sup_{j,\psi} |\eta_{j,k,\psi}|.$$

Again we have

$$E(\sup_{j,\psi} |\eta_{j,k,\psi}|) \approx C\frac{\sigma}{N}\sqrt{2\log 2^{2k}} = C\sqrt{k}\frac{\sigma}{N}.$$

So

$$\begin{aligned} E(|\epsilon|_{B^1_\infty(L_1)^*}) &\approx \sum_{k=0}^{K-1} C\sqrt{k}\frac{\sigma}{N} \\ &\approx CK^{3/2}\frac{\sigma}{N} = C(\log_2 N)^{3/2}\frac{\sigma}{N}. \end{aligned}$$

Thus, because of the continuous inclusions

$$B^1_1(L_1) \subset \mathrm{BV} \subset B^1_\infty(L_1)$$

and

$$B^1_\infty(L_1)^* \subset \mathrm{BV}^* \subset B^1_1(L_1)^*,$$

we find that

$$CN\sigma \le E\big(|\epsilon|_{\mathrm{BV}}\big) \le CN\sigma,$$

and

$$C\sqrt{\log N}\frac{\sigma}{N} \le E\big(|\epsilon|_{\mathrm{BV}^*}\big) \le C(\log N)^{3/2}\frac{\sigma}{N}.$$

8.3 Periodic textures

We compute some norms of the function $f(x_1, x_2) = \sin(2\pi n x_1)$, $n = 1, 2, \ldots$, which we use as a model for periodic textures.

Note that because f is constant in x_2, a second difference of f at x in any direction $h = (h_1, h_2)$ takes the same value as the second difference of f at x in direction $(h_1, 0)$ and $|(h_1, h_2)| \ge |(h_1, 0)|$, so the supremum in (11) occurs for some $h = (h_1, 0)$. We take a centered second difference in the horizontal direction, and abuse notation a bit with $x_1 = x$ and $h_1 = h$:

$$\begin{aligned} &\sin(2n\pi(x+h)) - 2\sin(2n\pi x) + \sin(2n\pi(x-h)) \\ &\quad = \big(\sin(2n\pi x)\cos(2n\pi h) + \cos(2n\pi x)\sin(2n\pi h)\big) \\ &\quad\quad - 2\sin(2n\pi x) \\ &\quad\quad + \big(\sin(2n\pi x)\cos(2n\pi h) - \cos(2n\pi x)\sin(2n\pi h)\big) \\ &\quad = -4\sin(2n\pi x)\frac{1-\cos(2n\pi h)}{2} \\ &\quad = -4\sin(2n\pi x)\sin^2(2n\pi h). \end{aligned}$$

So the $L_1(I)$ norm of the second difference is

$$4\sin^2(2n\pi h)\int_I |\sin(2n\pi x_1)|\,dx_1\,dx_2.$$

The value of the integral, which we'll call I_0, doesn't depend on n.

We see that $\sin^2(2n\pi h)/h$ takes a maximum value when

$$\begin{aligned}0 &= h\times 2\sin(2n\pi h)\cos(2n\pi h)2n\pi - \sin^2(2n\pi h)\\ &= \sin(2n\pi h)\big(4n\pi h\cos(2n\pi h) - \sin(2n\pi h)\big).\end{aligned}$$

So this is a maximum when h satisfies $\tan(2n\pi h) = 4n\pi h$, i.e, if we take y_0 to be the smallest positive solution of $\tan y = 2y$, then $y_0 = 2n\pi h$, or $h = y_0/(2n\pi)$ and

$$\begin{aligned}|f|_{B^1_\infty(L_1(I))} &= \sup_{t>0} t^{-1}\omega_2(\sin(2\pi n x_1),t)_1\\ &= \frac{2n\pi}{y_0}\times 4\sin^2(y_0)I_0 = Cn.\end{aligned}$$

A trivial calculation shows that $L_2(I)$ norm of f is $\sqrt{2}$.

We know of no rigorous argument to estimate the size of f in the dual space of $B^1_\infty(L_1(I))$. We will note, however, that if we write g in a Fourier basis of a product of sines and cosines in x_1 and x_2, then

$$|f|_{B^1_\infty(L_1(I))^*} = \sup\frac{\int_I fg}{|g|_{B^1_\infty(L_1(I))}},$$

and for $f(x_1,x_2) = \sin(2\pi n x_1)$, the only nonzero term in g that contributes to the integral in the numerator is the $\sin(2\pi n x_1)$ term. Trying only $g = \sin(2\pi n x_1)$ gives a numerator of $1/2$ and a denominator that is $\asymp n$.

So unless adding extra Fourier terms to $g = \sin(2\pi n x_1)$ (which will not affect the numerator) decreases the norm in the denominator (which we doubt), we expect that $|f|_{(B^1_\infty(L_1(I)))^*} \asymp (1/n)$.

These calculations support what one might suspect heuristically: If the frequency n of a texture is small, a smoothness norm cannot distinguish between that texture and a smooth function of the same amplitude. And if the frequency of a texture is large (so the wavelength is on the order of a small power of $\log N$ in an $N\times N$ image), then the dual of a smoothness norm cannot distinguish that texture from Gaussian noise whose standard deviation is the amplitude of the texture. Similarly, the dual of a smoothness norm will not distinguish low-contrast textures of moderate frequency from Gaussian noise with standard deviation much higher than the contrast of the texture.

9 Computational examples

9.1 Noise removal, smoothing

We compute the minimizers of

$$\frac{1}{2\lambda}\|f-u\|^2_{L_2(\mathcal{I})} + |u|_{B^1_\infty(L_1(\mathcal{I}))}. \tag{26}$$

In Figure 1, we begin with a 481×321 natural image from [4] together with their noisy input image "containing zero mean Gaussian noise with standard deviation $= 0.05$." We measured the RMS difference between the clean image and the noisy as 12.34 greyscales, less than the $256\times 0.05 = 12.8$ greyscales one might expect because of limiting pixel values to between 0 and 255.

We computed our $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ smoothing with parameter values $\lambda = 1/64$, $3/128$, $1/32$, $3/64$, and $1/16$. Because this image is not square, we had to decide whether to take h to be the inverse of the number of pixels in the smaller or larger side; we chose the smaller. Algorithm 2 applies. We applied Formula (12) of [5] to ensure that the bound on the final error was less than $1/4$ greyscale. (The final error could be much less than that, the error *bound* was $< 1/4$ greyscale.)

Table 1 Nonzero limiters μ^j of the vector fields p^j for various values of λ for the image in Figure 1.

λ	k	$\mu^{(k,0)}$	$\mu^{(0,k)}$	$\mu^{(k,k)}$	$\mu^{(k,-k)}$
1/64	1	.3944	.3381	.1358	.1316
3/128	1	.2567	.2425	.2059	.2049
	2	.0901	.0000	.0000	.0000
1/32	1	.1773	.1725	.1597	.1561
	2	.1559	.0745	.0433	.0606
3/64	1	.1078	.1138	.0977	.0972
	2	.0930	.0688	.1357	.1204
	3	.0078	.0000	.0000	.0561
	4	.0000	.0000	.0000	.0565
	5	.0000	.0000	.0000	.0453
1/16	1	.0782	.0854	.0737	.0732
	2	.0622	.0579	.1054	.0935
	3	.0000	.0000	.0353	.0721
	4	.0000	.0000	.0000	.0000
	5	.0000	.0000	.0000	.0807
	6	.0000	.0000	.0000	.1823

In Table 1 we show how more scales are involved in the computation as λ increases. We include all the nonzero limiters μ^j of Section 6 when the computation ended (with our error bound $< 1/4$). Our computations always involve second differences over a distance of $k = 1$ pixels, and as λ increases to $1/16$, computations incorporate second differences over $k = 2,\dots,6$ pixels.

We show the results of [4] using ROF and L_2–TGV^2_α smoothings, where the authors report choosing α experimentally to achieve the best RMSE results. We also

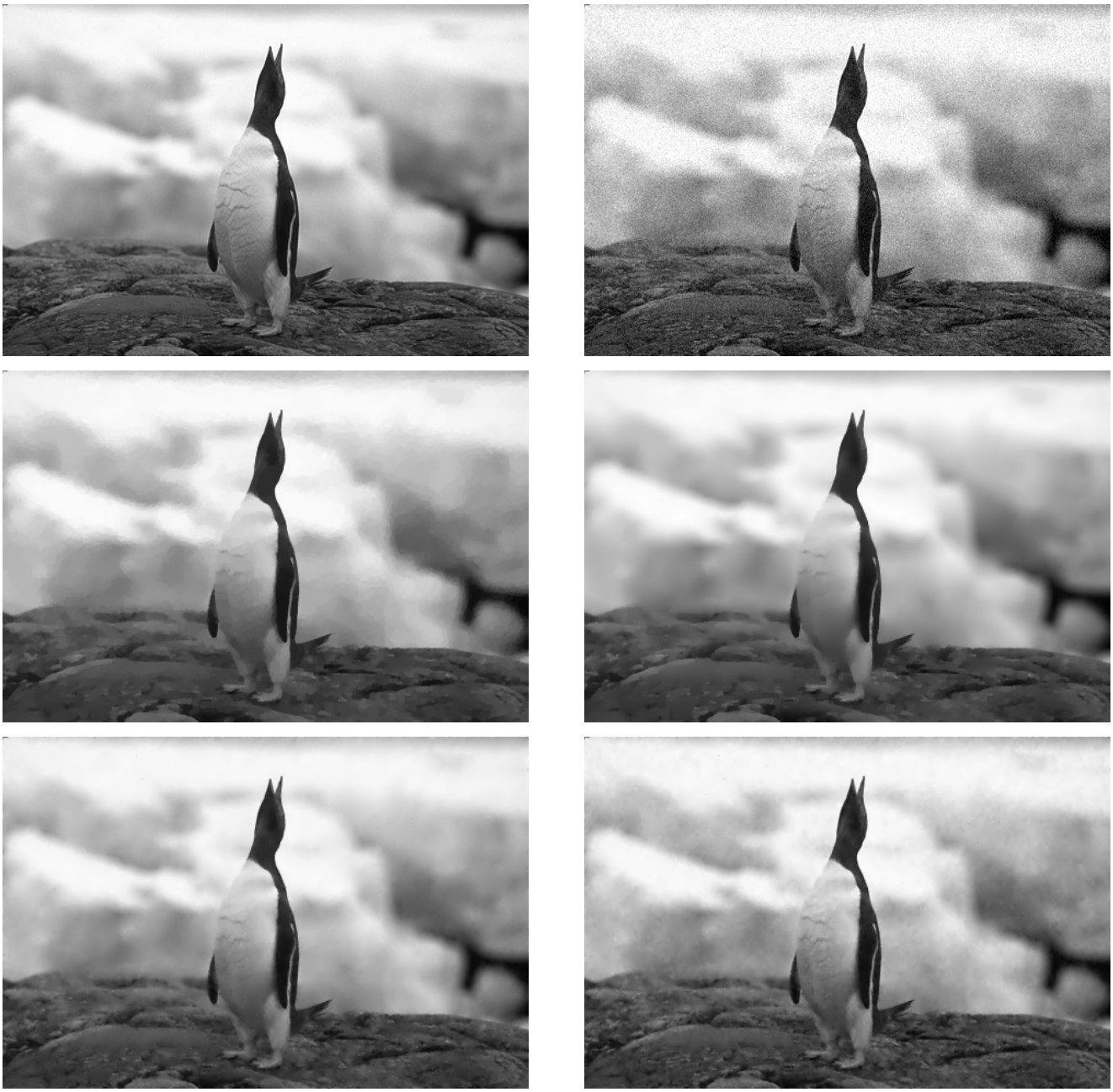

Fig. 1 Top left: original 480 × 321 image; top right: noisy image, RMSE = 12.34, SSIM = .4979. (All RMSE in greyscales.) Middle left: ROF, RMSE = 6.29, SSIM = .8710; middle right: L_2–TGV^2_α, RMSE = 6.17, SSIM = .8844. Bottom left: $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ with $\lambda = 3/64$, RMSE = 5.66, SSIM = .9030; bottom right: $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ with $\lambda = 1/32$, RMSE = 5.43, SSIM = .9022.

show our $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ smoothing with parameter values $\lambda = 3/64$ and $\lambda = 1/32$; these parameter values produced the best RMS error ($\lambda = 1/32$) and SSIM index [24, 23] ($\lambda = 3/64$) from among our choices of λ.

In Table 2, we report for various smoothing methods measures of image differences in a smooth subregion, a rough subregion, and a subregion with a relatively sharp edge, of the image in Figure 1; these subregions are shown in Figure 2. As might be expected, methods that smooth more ($L_2(I)$–TGV^2_α and $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ with $\lambda = 3/64$) do better in the smooth region and worse in the rough region than methods that smooth less (ROF and $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ with $\lambda = 1/32$, respectively). What may be of more interest is that smoothing with the largest space $B^1_\infty(L_1(I)) \supset \mathrm{BV}(I) = \mathrm{TGV}^2_\alpha(I)$ leads to the best results in the fractal-like rough regions.

Table 2 Image difference measures in subregions shown in Figure 2 for Rudin–Osher–Fatemi, L_2–TGV^2_α, and $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ smoothing for the image in Figure 1.

		ROF	TGV^2_α	$B^1_\infty(L_1(\mathcal{I}))$, $\lambda = 3/64$	$B^1_\infty(L_1(\mathcal{I}))$, $\lambda = 1/32$
Smooth	RMSE	2.9824	2.2243	2.6488	2.9861
	SSIM	0.9521	0.9735	0.9694	0.9521
Rough	RMSE	6.5958	6.6423	5.8526	5.3570
	SSIM	0.6727	0.6617	0.7426	0.7895
Near beak	RMSE	5.9354	5.8434	5.8347	6.4221
	SSIM	0.9596	0.9725	0.9654	0.9682

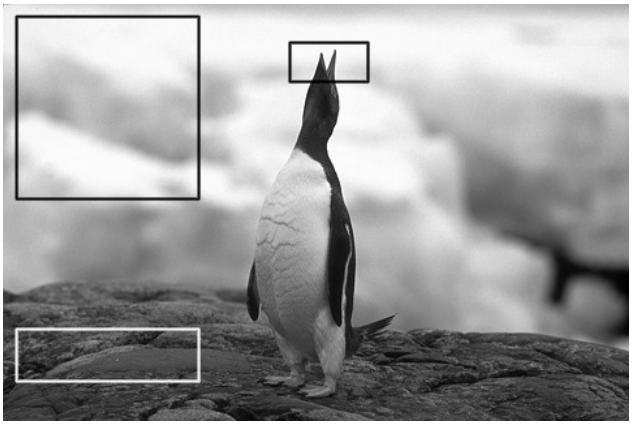

Fig. 2 Smooth subregion of image in Figure 1, beginning at row 10, column 10, of size 140×140; rough subregion beginning at row 250, column 10, of size 140×50; region near beak beginning at row 30, column 220, of size 60×30.

We next present some results that indicate the qualitative nature of the smoothed images produced by our method, including the style of artifacts when noise levels and smoothing parameters are large.

In Figure 3, we started with a 512×512 image of a girl; we then added i.i.d. Gaussian noise with mean zero and standard deviation 16 greyscales, so 99.7% of the noise added to pixels takes values within three standard deviations, or between -48 and 48 greyscales. For our image the added noise caused many of the noisy pixels to be "clipped" to either 0 or 255. We see that even when removing high noise levels there are relatively sharp edges, and in areas where noise was added to smooth regions, smooth reconstructions appear. The image with the smaller λ shows an intermediate state where noise has been smoothed but not entirely removed. Higher-contrast textures in the sweater survive smoothing that removes the noise, while lower-contrast textures are removed.

We now apply pointwise Besov smoothing to a synthetic image found in [17] with different smoothing parameters to indicate the scale space generated by the method. We take (26) with $\lambda = 3/32, 1/8, 1/4, 1$. Results are in Figure 4.

While $B^1_\infty(L_1(\mathcal{I}))$ minimizers do not suffer from staircasing, the example in Figure 4 also demonstrates their potential lack of smoothness. This is not so surprising, since $B^1_\infty(L_1(\mathcal{I}))$ is a "weak" smoother and in particular, weaker than both BV and TGV. While $B^1_\infty(L_1(\mathcal{I}))$ minimizers do not suffer from staircasing, they are also not required to have even the gradient a bounded measure, like BV and TGV. The image quality metrics reported in [17] for the BV $-L^2$ and TGV $- L^2$ minimizers of the noisy image in Figure 4 were SSIM=0.8979 and SSIM=0.9429, as opposed to the peak we experimentally obtained with $B^1_\infty(L_1(\mathcal{I})) - L^2$ of SSIM=0.8801 for $\lambda = 1/8$, also indicative of this lack of smoothness.

9.2 Separating smooth features, periodic textures, and noise

We follow the framework of Section 3.2 of [2] to compute some sample solutions to the following problem.

Informally speaking, given an image f and positive parameters λ and μ, we wish to find functions $\hat{u}$ and $\hat{v}$ that minimize over all u and v the functional

$$\frac{1}{2\lambda}\|f - u - v\|^2_{L_2(\mathcal{I})} + \|u\|_X + \delta_{\mu K}(v),$$

where X is a "smoothness space" and $K = \{v \mid \|v\|_Y \leq 1\}$ for some "dual smoothness space" Y, and $\delta_S(v) = 0$ if $v \in S$ and $+\infty$ otherwise.

We take X to be the (discrete) Besov space $B^1_\infty(L_1(\mathcal{I}))$ and Y to be the dual of the same space. We use the 512×512 synthetic image whose pixels are given by

$$f_i = \mathrm{round}(64 + 128e^{-16|x-(\frac{1}{2},\frac{1}{2})|^2} + 16\sin(2\pi(x_2 - 1/2) \times 16 \times 4^{x_2 - 1/2}) + 16\epsilon_i),$$

where $i = (i_1, i_2)$, $x = i/512$, and ϵ_i are i.i.d. $N(0,1)$ random variables. As usual, we "clip" the values of f_i so they cannot go above 255 or below 0, but the choice of parameters means that ϵ_i would need to be at least 3 standard deviations away from the mean (depending on the phase of the sinusoid and the value of the Gaussian at that point) for a pixel to be clipped, so clipping happens rarely.

Fig. 3 Top left: original 512 × 512 image; top right: noisy 512 × 512 image. Bottom left: smoothed image with $\lambda = 1/32$; bottom right: smoothed image with $\lambda = 3/64$.

The sinusoidal part of this image has wavelength 1/16 near the center, 1/8 near the top, and 1/32 near the bottom of the image.

In [2], the authors set λ to be "very small", to achieve "a maximum norm of $f - u - v$ of about 0.5 (for values ranging from 0 to 255)." Their goal was to have very little in the $f - u - v$ component.

Our goal here is different—to have a nontrivial component of the image in $f - u - v$. We use $\lambda = 1/4$, $\mu = 1/32$ (the standard deviation of the noise divided by N), and apply Algorithm 5.1.1 of [2] (equivalently, Algorithm 2.5 of [3]) with $\epsilon = 1/4$ (so we expect results to be accurate to within about 1/4 RMS greyscales). We also use Formula (12) from [5] to ensure that each iterated solution u^n, v^n has an error of $< 1/4$ RMS greyscales. (Again we note that this just bounds the error from above, the true error could be much smaller.)

We make several comments about the results. First, we find the noise contained in v (the "dual smoothness space" term) and the texture in $f - u - v$ (the $L_2(\mathcal{I})$ term). This allocation of noise and texture corresponds to the calculations in Section 8. Second, the calculations

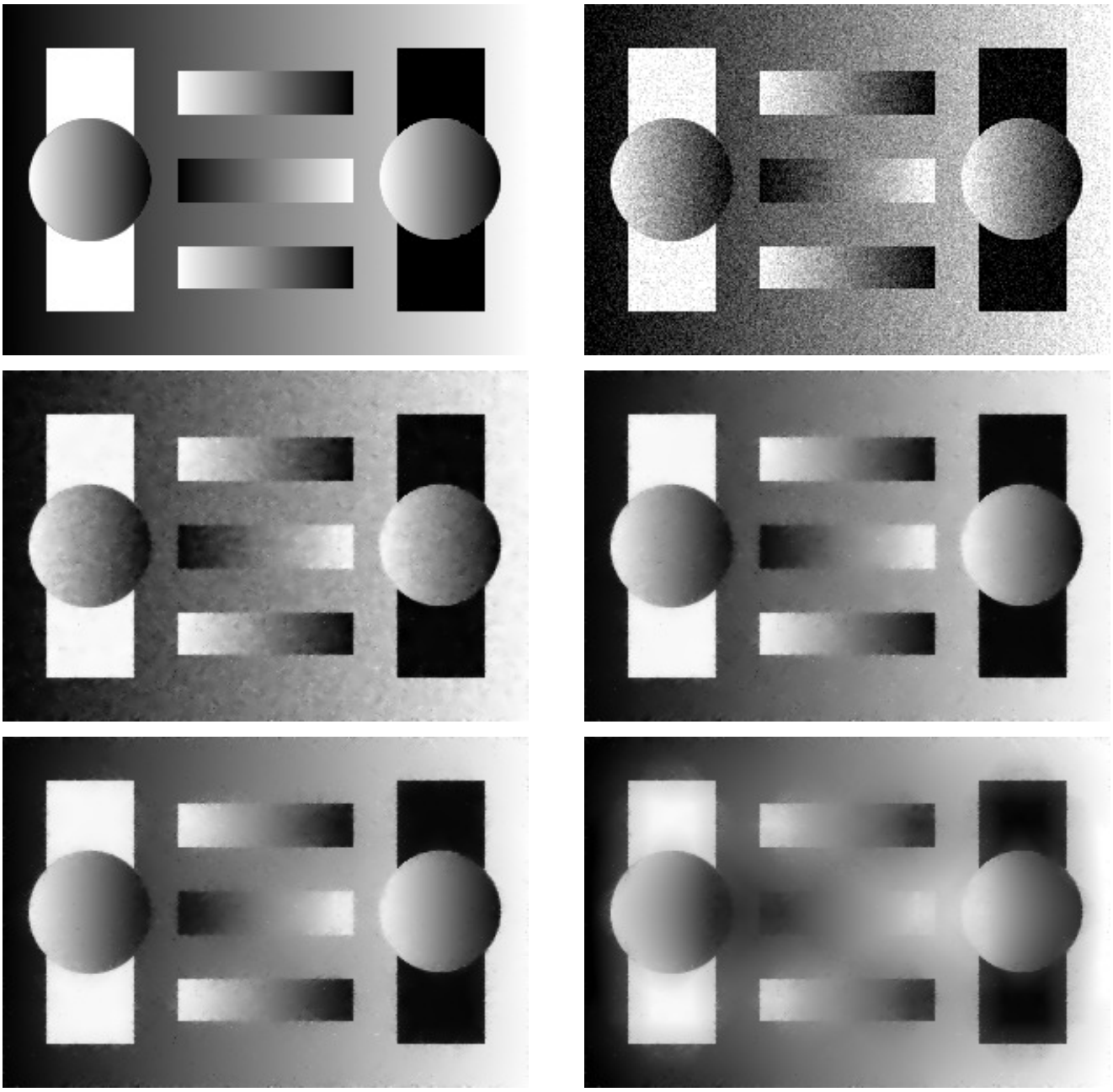

Fig. 4 Top left: original 300 × 200 image; top right: noisy 300 × 200 image. Middle left: $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ smoothed image with $\lambda = 3/32$: SSIM=0.8640 and RMSE=7.751; middle right: $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ smoothed image with $\lambda = 1/8$: SSIM=0.8801 and RMSE=8.0484. Bottom left: $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ smoothed image with $\lambda = 1/4$: SSIM=0.8784 and RMSE=9.6573; bottom right: $L_2(\mathcal{I})$–$B^1_\infty(L_1(\mathcal{I}))$ smoothed image with $\lambda = 1$: SSIM=0.7936 and RMSE=20.97744.

in Section 8 suggest that when the frequency of a texture is small, one cannot distinguish it from a smooth feature, and when the frequency of a texture is large, one cannot separate it from the noise. This computation supports the theory—in the images of Figure 5, low-frequency texture can be found in the smooth part and is absent from the noise image, while high-frequency texture is (mostly) absent from the smooth image while "ghosting" into the noise image. The texture image has small contributions from both the smooth Gaussian and the noise.

The relative amount of this contribution can be observed in Figure 6. This is the vertical slice of all four images with $i_1 = 250$.

We hesitated to include slice images for a number of reasons. Functions in $\mathrm{BV}(I)$ can be discontinuous across a vertical line in I, functions in $B^1_\infty(L_1(I))$ are less smooth, and "natural" images have significantly

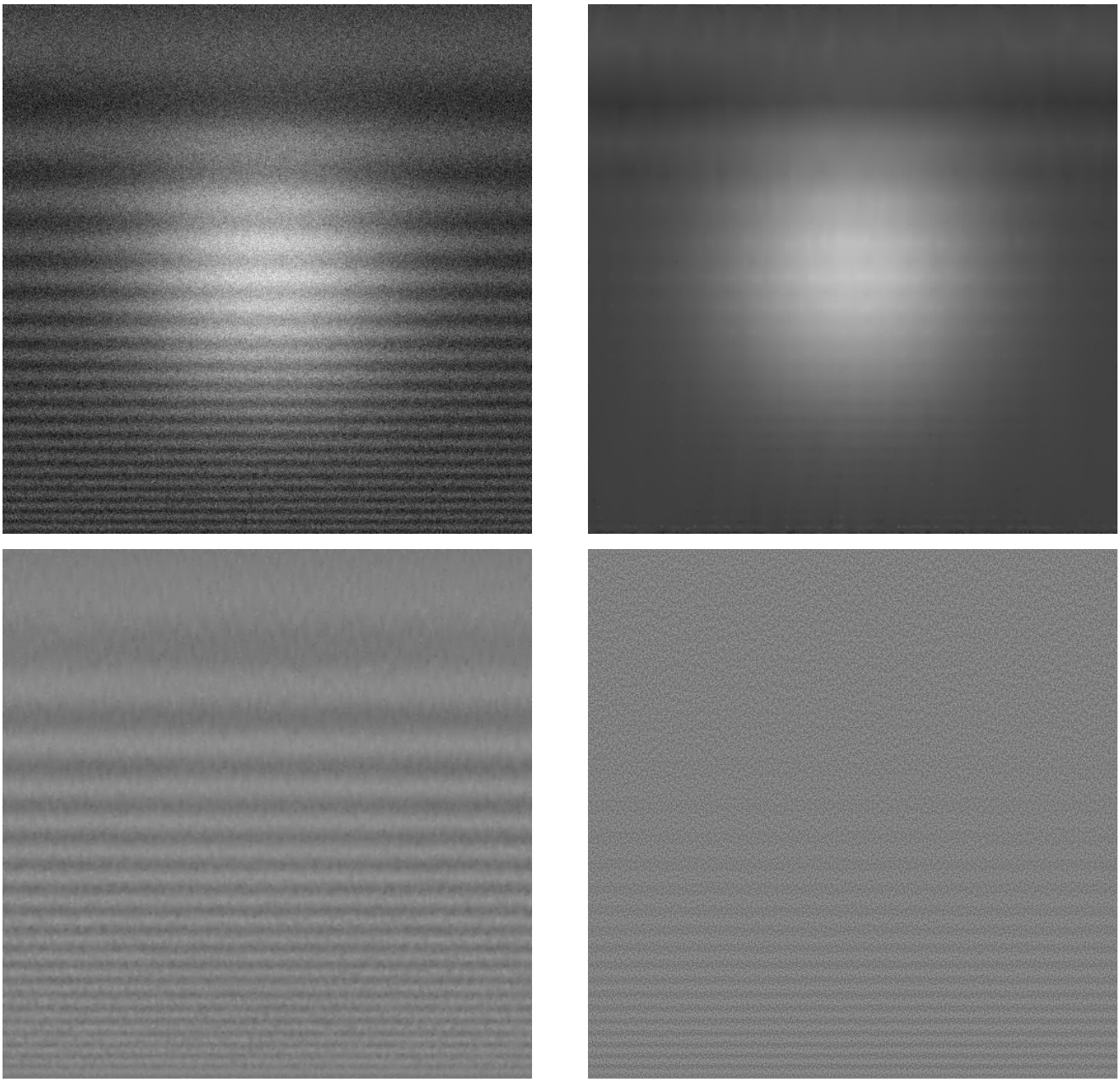

Fig. 5 Top left: composite 512×512 image; top right: smooth part u. Bottom left: texture $f - u - v$ plus 128; bottom right: noise v plus 128.

less smoothness still, see [13]. So, mathematically, the notion of the value of the original image, or even of the smooth part u, along a line is subtle, and practically it means that the value along one vertical line may have little relationship with the value along an adjacent or nearby line.

The value of an image f along a slice is related the boundary values along $\partial\Omega$ of a function in a smoothness space defined on Ω itself, or so-called "trace theorems"—each vertical line divides I into two subdomains, and one-sided "traces" may exist, determined by the value of f on each of the subdomains. One may consider "the" value along a vertical line to be the average of these two one-sided traces. Given an arbitrary function u in $B^1_\infty(L_1(I))$, the value of u along a vertical slice S, the trace of the function of that slice, is not in general in $L_1(S)$, but only in $L_p(S)$ for $p < 1$. This can be proved using techniques similar to those in [22]; for $u \in \mathrm{BV}(I)$, the trace of u is in $L_1(S)$—see [1].

The situation for noise ϵ on two adjacent vertical slices is even worse—the expected value of their inner product is zero (they're orthogonal, in expectation). We

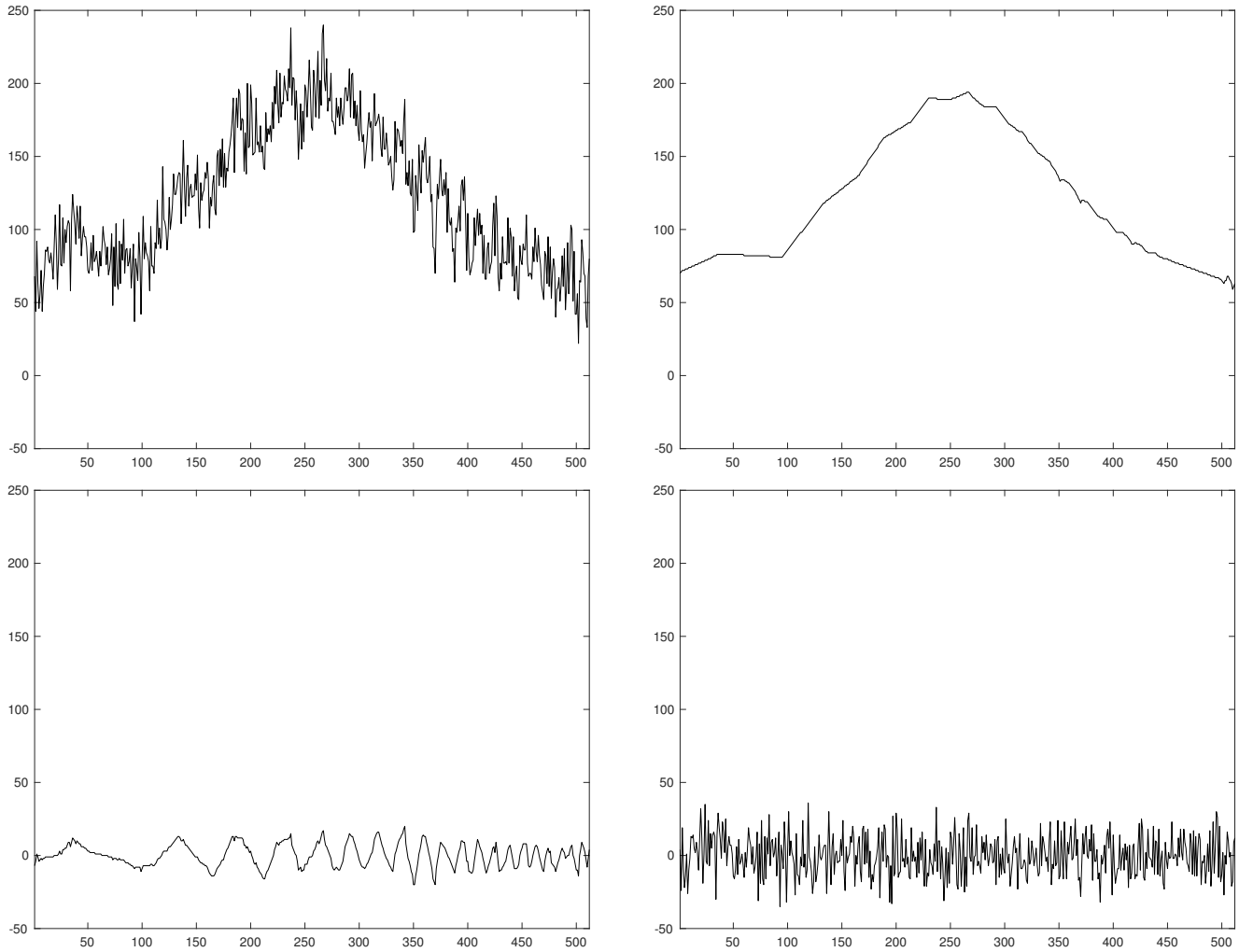

Fig. 6 Cross-section of the middle column of the images in Figure 5. From left to right, top to bottom: composite 512×512 image; smooth part u, periodic texture $f - u - v$; noise v.

can expect no relationship between the noise on two difference slices. The relationship between the texture part on two adjacent slices is between that of the smooth part u and the extracted noise—it's not as strong as the "smooth" part, but certainly stronger than noise.

So we chose a vertical line near the center of the image where the values of the components of f along that line illustrated a few points, at least a bit better than some of the nearby vertical lines.

First, we note that because the Human Visual System is very good at seeing patterns in images, one can see "ghosting" of the high-frequency oscillatory texture part in the smooth part of the image, but the slice of the smooth part u shows that the actual high-frequency oscillations in u are very small—one hardly notices them in the slice of u.

Second, one can get an idea of the magnitude of the low-frequency oscillation that is included in the smooth part u.

And finally, one can see that the amplitude of the oscillatory texture $f - u - v$ is largest in the middle of the slice and smaller at the ends, where the texture has partially "bled" into the smooth part u and the noise v.

Appendix

Here we describe a family of self-similar functions with fractal-like properties that illustrate some of the differences between membership in the Besov space $B^1_\infty(L_1(I))$ and the space of functions of bounded variation $\mathrm{BV}(I)$.

We begin with the function Φ that is continuous on $\mathbb{R}^2$, zero outside of I, with $\Phi((\frac{1}{2}, \frac{1}{2})) = 1$, and a linear polynomial on the triangles in I delineated by the boundary of I, the line $x_2 = x_1$, and the line $x_1 + x_2 = 1$. The graph of Φ is a pyramid with base I and height 1.

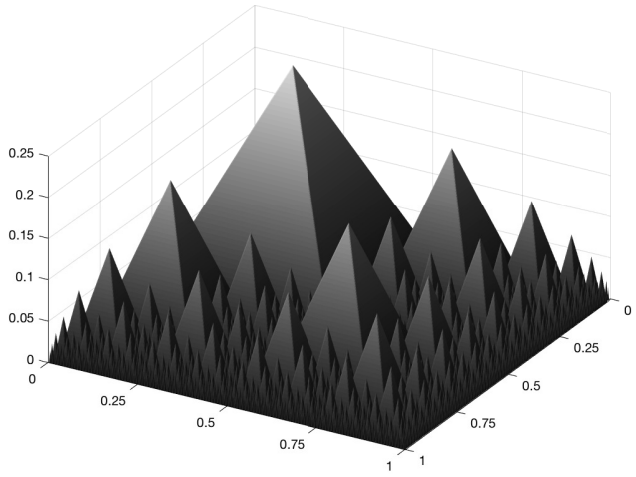

Fig. 7 Graph of the function f in (28) that is in $B^1_\infty(L_1(I))$ but not in $\mathrm{BV}(I)$.

We define $\phi(x) = \frac{1}{4}\Phi(2x)$, so ϕ is a scaled dyadic dilate of Φ with support $[0,1/2]^2$.

We construct the sequence of functions $f_0(x) = \phi(x)$ and

$$f_{k+1}(x) = \phi(x) + \frac{2}{3}[f_k(2x-(1,0)) + f_k(2x-(0,1)) + f_k(2x-(1,1))]. \quad (27)$$

We see that f_k consists of f_{k-1} plus 3^k dyadically dilated translates $\phi_{j,k}(x) = \phi(2^k x - j)$ for some multi-indices $j = (j_1, j_2)$ with coefficients $(2/3)^k$. The supports of all the $\phi_{j,k}$ are essentially disjoint.

Finally, we let $f = \lim_{k\to\infty} f_k$. We can write

$$f = \sum_{k=0}^{\infty}\sum_j c_{j,k}\phi_{j,k}(x), \quad c_{j,k} = \Big(\frac{2}{3}\Big)^k. \quad (28)$$

Thus, f is an infinite sum of scaled, dilated, and translated versions of the single pyramid Φ; Figure 7 illustrates the graph of f.

The arguments in [12][1] show that for any $0 < p, q \le \infty$ and $\alpha < \min(2, 1+1/p)$ we have

$$\|f\|_{B^\alpha_q(L_p(I))} \asymp \left(\sum_k \left(\sum_j \left[2^{\alpha k}\|c_{j,k}\phi_{j,k}\|_{L_p(I)}\right]^p\right)^{\frac{q}{p}}\right)^{\frac{1}{q}},$$

with the usual changes when p or q are ∞. In our case, we have

$$\|f\|_{B^1_\infty(L_1(I))} \asymp \sup_k \sum_j 2^k\Big(\frac{2}{3}\Big)^k \|\phi_{j,k}\|_{L_1(I)},$$

where for each k there are 3^k different offsets j.

We note that $\|\phi_{j,k}\|_{L_1(I)} = 4^{-k}\|\phi\|_{L_1(I)}$; because there are 3^k terms in the sum for each k,

$$\sum_j 2^k\Big(\frac{2}{3}\Big)^k \|\phi_{j,k}\|_{L_1(I)} = \|\phi\|_{L_1(I)},$$

and we have $\|f\|_{B^1_\infty(L_1(I))} \asymp \|\phi\|_{L_1(I)} < \infty$.

We'll now see that f is *not* in $\mathrm{BV}(I)$. Denoting the variation of f by $V(f)$, a simple scaling argument shows that $V(\phi_{j,k}) = 2^{-k}V(\phi)$. Since the supports of all the $\phi_{j,k}$ in the definition of f are essentially disjoint, the co-area formula shows that

$$\begin{aligned} V(f) &= \sum_k\sum_j \Big(\frac{2}{3}\Big)^k V(\phi_{j,k}) \\ &= \sum_k \Big(\frac{2}{3}\Big)^k \times 3^k \times 2^{-k}V(\phi) \\ &= \sum_k V(\phi) = \infty. \end{aligned}$$

In other words, there is a constant C such that for all $h \in \mathbb{R}^2$,

$$\|f(\cdot+2h) - 2f(\cdot+h) + f\|_{L_1(I_{2h})} \le C|h|,$$

but there is *no* constant C such that for all $h \in \mathbb{R}^2$,

$$\|f(\cdot+h) - f\|_{L_1(I_h)} \le C|h|.$$

Note that by replacing $2/3$ in (27) with any $0 < r < 2/3$, the resulting limit f is in both $B^1_\infty(L_1(I))$ and $\mathrm{BV}(I)$ (indeed, it's in $B^1_1(L_1(I)) \subset \mathrm{BV}(I)$). In this case, we have

$$V(f) = \frac{1}{1-(3/2)r}V(\phi),$$

so the variation of f tends to ∞ as $r \to 2/3$ (as one might expect), while $\|f\|_{B^1_\infty(L_1(I))}$ remains bounded.

And if $r > 2/3$, the function f is in neither $\mathrm{BV}(I)$ nor $B^1_\infty(L_1(I))$.

Thus both $\mathrm{BV}(I)$ and $B^1_\infty(L_1(I))$ contain fractal-like functions, but their norms in $B^1_\infty(L_1(I))$ can be arbitrarily smaller than their norms in $\mathrm{BV}(I)$.

Acknowledgments

The authors would like to thank Kristian Bredies for providing the penguin images in Figure 1 and for confirming the TGV experiments related to Figure 4.

[1] The crucial point is to see that we can take the local polynomial approximation in formula (4.18) of [12] to be identically zero on dyadic subsquares that are not entirely contained in the support of a single $\phi_{j,k}$—the self-similarity of f means that the local error with approximation zero is a fixed multiple of the error with the best linear polynomial approximation on each subsquare, and hence is near optimal with a constant that doesn't depend on the scale.

References

1. Anzellotti G, Giaquinta M (1978) BV functions and traces. Rend Sem Mat Univ Padova 60:1–21 (1979), URL http://www.numdam.org/item?id=RSMUP_1978__60__1_0
2. Aujol JF, Chambolle A (2005) Dual norms and image decomposition models. International Journal of Computer Vision 63(1):85–104, DOI 10.1007/s11263-005-4948-3, URL http://dx.doi.org/10.1007/s11263-005-4948-3
3. Aujol JF, Aubert G, Blanc-Fraud L, Chambolle A (2005) Image decomposition into a bounded variation component and an oscillating component. Journal of Mathematical Imaging and Vision 22(1):71–88, DOI 10.1007/s10851-005-4783-8, URL http://dx.doi.org/10.1007/s10851-005-4783-8
4. Bredies K, Kunisch K, Pock T (2010) Total generalized variation. SIAM J Imaging Sci 3(3):492–526, DOI 10.1137/090769521, URL http://dx.doi.org/10.1137/090769521
5. Chambolle A (2005) Total variation minimization and a class of binary MRF models. In: Energy Minimization Methods in Computer Vision and Pattern Recognition: 5th International Workshop, EMMCVPR 2005, St. Augustine, FL, USA, November 9-11, 2005: proceedings/ Anand Rangarajan, Baba Vemuri, Alan L. Yuille (eds.)., LNCS 3757, Springer-Verlag, pp 136–152
6. Chambolle A, Pock T (2011) A first-order primal-dual algorithm for convex problems with applications to imaging. J Math Imaging Vision 40(1):120–145, DOI 10.1007/s10851-010-0251-1, URL http://dx.doi.org/10.1007/s10851-010-0251-1
7. Chambolle A, DeVore RA, Lee NY, Lucier BJ (1998) Nonlinear wavelet image processing: variational problems, compression, and noise removal through wavelet shrinkage. IEEE Trans Image Process 7(3):319–335, DOI 10.1109/83.661182, URL http://dx.doi.org/10.1109/83.661182
8. Chambolle A, Levine SE, Lucier BJ (2011) An upwind finite-difference method for total variation-based image smoothing. SIAM J Imaging Sci 4(1):277–299, DOI 10.1137/090752754, URL https://doi.org/10.1137/090752754
9. Cramér H (1946) Mathematical Methods of Statistics. Princeton Mathematical Series, vol. 9, Princeton University Press, Princeton, N. J.
10. DeVore RA, Lorentz GG (1993) Constructive Approximation, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol 303. Springer-Verlag, Berlin
11. DeVore RA, Lucier BJ (1992) Fast wavelet techniques for near-optimal image processing. In: Military Communications Conference, 1992. MILCOM '92, Conference Record. Communications - Fusing Command, Control and Intelligence., IEEE, pp 1129–1135 vol.3, DOI 10.1109/MILCOM.1992.244110
12. DeVore RA, Popov VA (1988) Interpolation of Besov spaces. Trans Amer Math Soc 305(1):397–414
13. DeVore RA, Jawerth B, Lucier BJ (1992) Image compression through wavelet transform coding. IEEE Trans Inform Theory 38(2, part 2):719–746, URL https://doi.org/10.1109/18.119733
14. Ditzian Z, Ivanov KG (1993) Minimal number of significant directional moduli of smoothness. Anal Math 19(1):13–27, DOI 10.1007/BF01904036, URL http://dx.doi.org/10.1007/BF01904036
15. Donoho DL, Johnstone IM (1995) Adapting to unknown smoothness via wavelet shrinkage. J Amer Statist Assoc 90(432):1200–1224, URL http://links.jstor.org/sici?sici=0162-1459(199512)90:432<1200:ATUSVW>2.0.CO;2-K&origin=MSN
16. Haddad A, Meyer Y (2007) Variational methods in image processing. In: Perspectives in nonlinear partial differential equations, Contemp. Math., vol 446, Amer. Math. Soc., Providence, RI, pp 273–295, DOI 10.1090/conm/446/08636, URL http://dx.doi.org/10.1090/conm/446/08636
17. Papafitsoros K, Schönlieb C (2014) A combined first and second order variational approach for image reconstruction. Journal of Mathematical Imaging and Vision 48(2):308–338, DOI 10.1007/s10851-013-0445-4, URL http://dx.doi.org/10.1007/s10851-013-0445-4
18. Ridders C (1979) A new algorithm for computing a single root of a real continuous function. Circuits and Systems, IEEE Transactions on 26(11):979–980, DOI 10.1109/TCS.1979.1084580
19. Rockafellar RT (1997) Convex Analysis. Princeton Landmarks in Mathematics, Princeton University Press, Princeton, NJ, reprint of the 1970 original, Princeton Paperbacks
20. Rudin LI, Osher S, Fatemi E (1992) Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60(14):259–268, DOI http://dx.doi.org/10.1016/0167-2789(92)90242-F, URL http://www.sciencedirect.com/science/article/pii/016727899290242F
21. Sauer K, Bouman C (1992) Bayesian estimation of transmission tomograms using segmentation based optimization. Nuclear Science, IEEE Transactions

on 39(4):1144–1152, DOI 10.1109/23.159774
22. Schneider C (2010) Trace operators in Besov and Triebel-Lizorkin spaces. Z Anal Anwend 29(3):275–302, URL https://doi.org/10.4171/ZAA/1409
23. Wang Z (2009) SSIM Index with automatic downsampling, Version 1.0. https://www.mathworks.com/matlabcentral/answers/uploaded_files/29995/ssim.m, [Online; accessed 27-July-2017]
24. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP (2004) Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4):600–612, DOI 10.1109/TIP.2003.819861