Interpreting Translation-Invariant Wavelet Shrinkage as A New Image Smoothing Scale Space

Antonin Chambolle[1] and Bradley J. Lucier[2] (*Senior Member, IEEE*)

Abstract

Coifman and Donoho suggested translation-invariant wavelet shrinkage as a way to remove noise from images. Basically, their technique applies wavelet shrinkage to a two-dimensional version of the semi-discrete wavelet representation of Mallat and Zhong. Coifman and Donoho also showed how the method could be implemented in $O(N \log N)$ operations, where there are N pixels. In this paper, we provide a mathematical framework for iterated translation-invariant wavelet shrinkage, and show, using a theorem of Kato and Masuda, that with orthogonal wavelets it is equivalent to gradient descent in $L_2(I)$ along the semi-norm for the Besov space $B_1^1(L_1(I))$, which, in turn, can be interpreted as a new nonlinear wavelet-based image smoothing scale space. Unlike many other scale spaces, the characterization is not in terms of a nonlinear partial differential equation.

Keywords: Wavelets, wavelet shrinkage, gradient descent, image smoothing scale space.

EDICS Category: 2-NFLT (Nonlinear Filtering and Enhancement). Also 2-WAVP (Wavelets and Multiresolution Processing).

1. Introduction

Ronald Coifman and David Donoho [4] suggested translation-invariant wavelet shrinkage as a way to remove noise from images. Basically, their technique applies wavelet shrinkage to a two-dimensional version of the semi-discrete wavelet representation of Mallat and Zhong [15]. Coifman and Donoho also showed how the method could be implemented in $O(N \log N)$ operations, where there are N pixels, which compares to $O(N)$ operations for ordinary wavelet shrinkage, and $O(N \log N)$ operations for the Fast Fourier Transform. In this paper, we provide a mathematical framework for iterated translation-invariant wavelet shrinkage, and show, using a theorem of Kato and Masuda [13], that with orthogonal wavelets it is equivalent to gradient descent in $L_2(I)$ along the semi-norm for the Besov space $B_1^1(L_1(I))$, which, in turn, can be interpreted as a new nonlinear wavelet-based image smoothing scale space.

---

[1] CEREMADE (CNRS UMR 7534), Université de Paris–Dauphine, 75775 Paris CEDEX 16, France, `Antonin.Chambolle@ceremade.dauphine.fr`. Supported by the CNRS.

[2] Department of Mathematics, Purdue University, West Lafayette, IN 47907-1395, +1 (765) 494-1979 (office), +1 (765) 494-0548 (fax), `lucier@math.purdue.edu`. Supported in part by the Office of Naval Research, Contract N00014-91-J-1152. Part of this work was done while the author was a visiting scholar at CEREMADE, Université de Paris–Dauphine, Paris, France.

The rest of the paper is organized as follows. Section 2 notes the Besov spaces we use and recalls the equivalence between certain Besov space norms of functions and sequence norms of wavelet coefficients. In Section 3 we briefly describe the periodic wavelet constructions we use. Section 4 gives an overview of image smoothing scale spaces, while Section 5 relates several scale spaces to variational problems, and Section 6 shows how to interpret usual wavelet shrinkage as an image smoothing scale space. In Section 7, we discuss various redundant wavelet representations of images, of which the translation-invariant wavelet representation is of most interest to us. Section 8 formalizes the notion of iterated translation-invariant wavelet shrinkage, and uses a theorem of Kato and Masuda to show that, in the limit, it is mathematically equivalent (when using orthogonal wavelets) to gradient descent in $L_2(I)$ along the $B_1^1(L_1(I))$ semi-norm. Finally, in Section 9 we present various computational examples of interest.

2. Orthogonal Wavelets and Besov Spaces

We begin with a function f defined on the unit square $I := [0, 1)^2$, which is extended periodically to all of $\mathbb{R}^2$ by

$$f(x + j) = f(x), \quad x \in I, \ j \in \mathbb{Z}^2,$$

where $\mathbb{Z}^2 := \{(j_1, j_2) \mid j_1, j_2 \in \mathbb{Z}\}$. This function f represents our image on I.

Besov spaces have been used with great success to both explain and predict the performance of wavelet-based image processing algorithms, e.g., in compression [6], noise removal [2] [9], and tomography [8] [14]. Here, we work with the Besov space $B_1^1(L_1(I))$, which contains, roughly speaking, functions with one derivative in $L_1(I)$.

The precise definition of the Besov spaces $B_q^\alpha(L_p(I))$, which can be found, e.g., in [7], does not concern us here; what is important, both theoretically and algorithmically, is the equivalence between the $B_q^\alpha(L_p(I))$ norm and certain norms of wavelet coefficients.

We consider compactly supported orthogonal wavelets, specifically Daubechies' orthogonal wavelets [5]. Assume we have a scaling function ϕ and wavelets $\Psi := \{\psi^{(i)} \mid i = 1, 2, 3\}$ constructed by tensor products of a one-dimensional orthogonal wavelet system.

If we define
$$\psi_{j,k}(x) := 2^k \psi(2^k x - j)$$

for $\psi \in \Psi$, $k \geq 0$, and $j \in \mathbb{Z}^2$, then any function f in $L_2(\mathbb{R}^2)$ can be written as

$$f = \sum_j \langle f, \phi(\cdot - j) \rangle \, \phi(\cdot - j) + \sum_{j,k,\psi} \langle f, \psi_{j,k} \rangle \, \psi_{j,k},$$

where the sums range over all $j \in \mathbb{Z}^2$, all $k \geq 0$, and all $\psi \in \Psi$.

If the functions in Ψ are smooth enough, then one can often determine whether a function f is in the Besov space $B_q^\alpha(L_p(\mathbb{R}^2))$ by examining its wavelet coefficients. We need this equivalence only when $p = q$, in which case

$$|f|_{B_q^\alpha(L_q(\mathbb{R}^2))} \asymp \left(\sum_{j,k,\psi} [2^{\alpha k} |\langle f, \psi_{j,k} \rangle| \, \|\psi_{j,k}\|_{L_q}]^q \right)^{1/q} = \left(\sum_{j,k,\psi} [2^{\alpha k} 2^{(1-2/q)k} |\langle f, \psi_{j,k} \rangle|]^q \right)^{1/q}.$$

$(A(f) \asymp B(f)$ means that there exist positive constants C_1 and C_2 such that for all f, $C_1 A(f) \leq B(f) \leq C_2 A(f)$.) See, e.g., [10].

We also have the equivalent norms

$$\|f\|_{B_q^\alpha(L_q(\mathbb{R}^2))} \asymp \left(\sum_{j} |\langle f, \phi(\cdot - j) \rangle|^q \right)^{1/q} + \left(\sum_{j,k,\psi} [2^{\alpha k} 2^{(1-2/q)k} |\langle f, \psi_{j,k} \rangle|]^q \right)^{1/q}.$$

3. Periodic Wavelets

One can easily construct periodic wavelets on $L_2(I)$ that can be used to decompose periodic functions f on $L_2(I)$. For example, for $\psi \in \Psi$, $k \geq 0$ and $j \in \{0, 1, \dots, 2^k - 1\}^2$, one sets [7]

$$\psi_{j,k}^\circ(x) := \sum_{\ell \in \mathbb{Z}^2} \psi_{j,k}(x + \ell), \quad x \in I.$$

One constructs ϕ° in the same way; we have $\phi^\circ(x) = 1$ for all x, since $\{\phi(\cdot - j) \mid j \in \mathbb{Z}^2\}$ forms a partition of unity.

Again, it is not hard to show [7] that one now has a periodic orthogonal wavelet system on $L_2(I)$ such that

$$f(x) = \langle f, \phi^\circ \rangle + \sum_{j,k,\psi} \langle f, \psi_{j,k}^\circ \rangle \, \psi_{j,k}^\circ(x)$$

and

$$|f|_{B_q^\alpha(L_q(I))} \asymp \left(\sum_{j,k,\psi} [2^{\alpha k} 2^{(1-2/q)k} |\langle f, \psi_{j,k}^\circ \rangle|]^q \right)^{1/q}.$$

Since we deal only with periodic wavelets in the rest of the paper, we drop the superscripts and write for $f \in L_p(I)$, $1 \leq p < \infty$,

(1) $$f = \langle f, 1 \rangle + \sum_{j,k,\psi} \langle f, \psi_{j,k} \rangle \, \psi_{j,k}$$

and

(2) $$|f|_{B_q^\alpha(L_q(I))} \asymp \left(\sum_{j,k,\psi} [2^{(\alpha+1-2/q)k} |\langle f, \psi_{j,k} \rangle|]^q \right)^{1/q}.$$

3

4. Image Smoothing Scale Spaces

The concept of an image smoothing scale space seems to us rather vague, although there are many examples; see, for example [11]. We think of it as a contraction semigroup on $L_2(I)$. That is, it is an operator S_λ that depends on a parameter λ that satisfies for nonnegative λ and μ and all f and g in $L_2(I)$

$$S_\lambda S_\mu f = S_{\lambda+\mu} f, \tag{3}$$

$$\|S_\lambda f - S_\lambda g\|_{L_2(I)} \le \|f - g\|_{L_2(I)}, \text{ and} \tag{4}$$

$$S_0 f = f. \tag{5}$$

This definition does not require that $S_\lambda f$ be continuous in λ for each f; an example of such a discontinuous smoothing scale space S_λ is the projection onto a closed subspace (or any closed convex subset) of $L_2(I)$ for $\lambda > 0$ and the identity otherwise. We define a continuous contraction semigroup as satisfying (3), (4), and (5) as well as

$$\lim_{\lambda \to 0, \ \lambda > 0} \|S_\lambda f - f\|_{L_2(I)} = 0. \tag{6}$$

Perhaps the best-known example of an image smoothing scale space is the space generated by the heat equation; i.e., $S_\lambda f(x) = u(x, \lambda)$ where $u(x, t)$ satisfies

$$\begin{aligned} \frac{\partial u}{\partial t} - \Delta u &= 0, \quad x \in I, \ t > 0, \\ u(x, 0) &= f(x), \quad x \in I, \end{aligned} \tag{7}$$

with periodic boundary conditions or Neumann boundary conditions.

5. Variational Problems

It is argued in [16] that many reasonable image smoothing scale spaces can be realized as descent along the subgradient of a lower-semicontinuous, (proper) convex functional on $L_2(I)$. We give definitions of these terms below. An excellent reference for this section is [1].

A (proper) convex functional is a function ϕ that maps $L_2(I)$ to $(-\infty, \infty]$, not identically equal to ∞, such that for all f and g in $L_2(I)$ and all $t \in (0, 1)$,

$$\phi((1 - t)f + tg) \le (1 - t)\phi(f) + t\phi(g).$$

ϕ is lower-semicontinuous (l.s.c.) if and only if for all values $a \in (-\infty, \infty)$, the set of all f such that $\phi(f) \le a$ is closed in $L_2(I)$. The domain of ϕ, denoted $\operatorname{dom} \phi$, is the set of all f such that $\phi(f)$ is finite.

4

An example of a l.s.c. convex functional is

$$
(8) \qquad \phi_1(f) := \begin{cases} \dfrac{1}{2} \displaystyle\int_I |\nabla f(x)|^2 \, dx, & f \in H^1(I), \\ \infty, & \text{otherwise}, \end{cases}
$$

where $H^1(I)$ is the space of functions in $L_2(I)$ with square-integrable gradients. This functional is associated with the heat equation. If a subset S of $L_2(I)$ is closed and convex, then

$$
(9) \qquad \phi_2(f) = \chi_S(f) := \begin{cases} 0, & f \in S, \\ \infty, & \text{otherwise}, \end{cases}
$$

is also a l.s.c. convex functional, whose gradient flow is just given by the projection onto S at any time $t > 0$.

The subgradient of a l.s.c. convex function ϕ, denoted $\partial\phi$, is defined by

$$
g \in \partial\phi(f) \iff (\forall h \in L_2(I)) \quad \phi(h) \geq \phi(f) + \langle g, h - f \rangle.
$$

Note that $\partial\phi$ is generally multi-valued. If ϕ is differentiable at f then the subgradient of ϕ at f is single-valued, consisting of the derivative of ϕ at f.

The domain of $\partial\phi$ is the set of all f for which $\partial\phi(f)$ is not empty. Note that if $\phi(f) = \infty$, then $\partial\phi(f)$ is empty, since there is at least one h for which $\phi(h) < \infty$. Thus, the domain of $\partial\phi$ is a subset of the domain of ϕ.

The subgradient of ϕ_1 is equal to $\{-\Delta f\}$ for $f \in H^2(I)$, the space of functions with square-integrable second derivatives, and is empty otherwise. Thus, the domain of $\partial\phi$ can be strictly smaller than the domain of ϕ; in all cases, however, $\operatorname{dom}\partial\phi$ is dense in $\operatorname{dom}\phi$. In (9), $\partial\phi_2$ is harder to describe. One knows, however, that if f is in the interior of S then $\partial\phi_2(f) = \{0\}$; if S has a well-defined exterior normal at a point f on the boundary of S, $\partial\phi_2(f)$ consists of all positive multiples of that exterior normal.

We consider the formal problem: Find $u \colon [0, \infty) \to L_2(I)$ that satisfies

$$
(10) \qquad \begin{aligned} \frac{\partial u}{\partial t} + \partial\phi(u) &\ni 0, \quad t \geq 0, \\ u(0) &= u_0 \in L_2(I). \end{aligned}
$$

The classical approach to solving this equation involves the Yosida approximation, which corresponds to using backward differences [1]. One chooses a timestep Δt and obtains approximations $u_{\Delta t}^n \approx u(n\Delta t)$ that satisfy: $u_{\Delta t}^0 = u_0$ and

$$
(11) \qquad \frac{u_{\Delta t}^n - u_{\Delta t}^{n-1}}{\Delta t} + \partial\phi(u_{\Delta t}^n) \ni 0, \quad n = 1, 2, \ldots.
$$

For any $u_{\Delta t}^{n-1} \in L_2(I)$, this problem always has a unique solution, which can be characterized as the function that minimizes over all $g \in L_2(I)$ the functional

(12) $$\frac{1}{2\Delta t}\|g - u_{\Delta t}^{n-1}\|_{L_2(I)}^2 + \phi(g).$$

We define

$$u(t) := \lim_{\substack{n \to \infty \\ n\Delta t = t}} u_{\Delta t}^n;$$

it is well known that this limit exists, and $S_t u_0 := u(t)$ defines a contraction semigroup on $L_2(I)$ that satisfies (10) [1].

With the previous examples, the solution of (10) with $\phi = \phi_1$ is the same as the solution of (7) with initial data $u_0 = f$. In this case, we have a continuous contraction semigroup on $L_2(I)$. The solution of (10) with $\phi = \phi_2$ is

$$u(t) = \begin{cases} u_0, & t = 0, \\ P_S u_0, & t > 0, \end{cases}$$

where P_S is the projection operator onto the set S. Here, the solution operator $S_t u_0 := u(t)$ is a contraction semigroup, but not a continuous contraction semigroup for $u_0 \notin S$.

6. Wavelet Shrinkage as an Image Smoothing Scale Space

This section repeats and isolates some arguments in Section C of [2]; more discussion of our motivations can be found there.

We consider now the specific case of gradient descent along the functional $|f|_{B_1^1(L_1(I))}$ in $\mathbb{R}^2$. Given a suitably smooth periodic orthogonal wavelet basis, we have by (2),

$$|f|_{B_1^1(L_1(I))} \asymp \phi(f) := \sum_{j,k,\psi} |\langle f, \psi_{j,k} \rangle|.$$

We now take the position that it doesn't really matter what form of equivalent functionals we take, so we consider gradient descent along the functional $\phi(f)$. The domain of ϕ is the set of all functions in $B_1^1(L_1(I))$, which is dense in $L_2(I)$; the subgradient of ϕ is

$$\partial\phi(f) = L_2(I) \cap \sum_{j,k,\psi} \operatorname{sign}(\langle f, \psi_{j,k} \rangle) \, \psi_{j,k},$$

where

$$\partial|x| = \operatorname{sign} x := \begin{cases} \{1\}, & x > 0, \\ [-1, 1], & x = 0, \\ \{-1\}, & x < 0. \end{cases}$$

6

As usual, the domain of $\partial\phi$ consists of all f for which $\partial\phi(f)$ is nonempty. Because the wavelets are orthogonal, and for each nonzero term in f the corresponding term in $\partial\phi(f)$ has size $\|\psi_{j,k}\|_{L_2(I)}$, the domain of $\partial\phi$ consists of all f of the form (1) with only finitely many nonzero terms in the sum.

We now note the equality

$$(13) \qquad \|f\|^2_{L_2(I)} = |\langle f, 1\rangle|^2 + \sum_{j,k,\psi} |\langle f, \psi_{j,k}\rangle|^2.$$

With these definitions of $\|f\|_{L_2(I)}$ and $\phi(f)$, we now propose to solve (10) along the lines of Section C in [2]. To this end, we solve at each time step (11), by minimizing (12). Calculus shows that

$$\langle u^n_{\Delta t}, 1\rangle = \langle u_0, 1\rangle$$

for all n and Δt, and

$$\langle u^n_{\Delta t}, \psi_{j,k}\rangle = S_{n\Delta t}(\langle u_0, \psi_{j,k}\rangle),$$

where

$$S_\lambda(x) := \begin{cases} x - \lambda, & \lambda \le x, \\ 0, & -\lambda \le x \le \lambda, \\ x + \lambda, & x \le -\lambda. \end{cases}$$

Since $u^{T/\Delta t}_{\Delta t}$ doesn't depend on Δt, we see that the limit $u(T) = u^{T/\Delta t}_{\Delta t}$ for any Δt for which $T/\Delta t$ is an integer. Thus, the solution of (10) is just

$$u(t) = \langle u_0, 1\rangle + \sum_{j,k,\psi} S_t(\langle u_0, \psi_{j,k}\rangle)\psi_{j,k},$$

which is *wavelet shrinkage* as introduced by David Donoho and Iain Johnstone [9]. Thus, wavelet shrinkage induces a continuous contraction semigroup on $L_2(I)$ that can be interpreted as gradient descent along the $B^1_1(L_1(I))$ seminorm, which in turn can be interpreted as a nonlinear image smoothing scale space.

7. REDUNDANT WAVELET REPRESENTATION OF IMAGES

The mapping between $L_2(I)$ and the sequence space ℓ_2 given by

$$f \in L_2(I) \leftrightarrow \{\langle f, 1\rangle, \ \langle f, \psi_{j,k}\rangle\} \in \ell_2$$

is one-to-one and onto; indeed, it is an isometry. So, for every $f \in L_2(I)$ we have a corresponding sequence in ℓ_2, but, more importantly, for every sequence in ℓ_2 there is a corresponding $f \in L_2(I)$.

There are some practical difficulties with using this representation of images. For instance, the seminorm on the right hand sides of (2) is *wavelet dependent*: if you change the wavelets used in the representation, you change the semi-norm, in contrast to the intrinsic definition of the Besov space norm.

More disquieting is the fact that while usual definition of the Besov space norm is invariant under translation of the function f, (2) is not. Coifman and Donoho proposed to use translation-invariant wavelet smoothing to get around this problem, but translation-invariant representations of images were used before this (e.g., representations using the continuous wavelet transform [5, Chapter 2]). We propose to make this idea precise in this section.

We achieve translation invariance as follows. For h in I, we define the translation operator $T_h f(x) := f(x - h)$, remembering that our functions are doubly periodic on the plane. One can find the wavelet representation of $T_{-h} f$, and then apply T_h to this wavelet representation. Finally, we average over all $h \in I$.

Putting these words into formulae goes as follows.

$$
T_{-h} f(x) = \langle T_{-h} f, 1 \rangle + \sum_{j,k,\psi} \langle T_{-h} f, \psi_{j,k} \rangle \psi_{j,k}(x)
$$

$$
= \int_I f(y + h)\, dy + \sum_{j,k,\psi} \int_I f(y + h)\, 2^k \psi(2^k y - j)\, dy \times 2^k \psi(2^k x - j)
$$

$$
= \int_I f(y)\, dy + \sum_{j,k,\psi} \int_I f(y)\, 2^k \psi(2^k(y - h) - j)\, dy \times 2^k \psi(2^k x - j)
$$

(The third equality comes from a change of variables $y + h \to y$.) Thus,

$$
f(x) = \int_I T_h(T_{-h} f(x))\, dh
$$

$$
= \int_I f(y)\, dy + \int_I \sum_{j,k,\psi} \int_I f(y)\, 2^k \psi(2^k(y - h) - j)\, dy \times 2^k \psi(2^k(x - h) - j)\, dh
$$

$$
= \int_I f(y)\, dy + \sum_{j,k,\psi} \int_I \int_I f(y)\, 2^k \psi(2^k(y - h))\, dy \times 2^k \psi(2^k(x - h))\, dh
$$

The last equality comes from a change in the order of integration and summation and the change of variables $2^k h + j \to 2^k h$.

Note that the terms in the sum no longer depend on the index j, and for each k there are 2^{2k} terms with different j. By defining $\psi_k(x) := 2^k \psi(2^k x)$, we get the final representation formula

$$
f(x) = \int_I f(y)\, dy + \sum_{k,\psi} 2^{2k} \int_I \int_I f(y)\, \psi_k(y - h)\, dy\, \psi_k(x - h)\, dh.
$$

8

This representation is a two-dimensional version of the semi-discrete, translation-invariant representation of Mallat and Zhong [15]. We invite the reader to compare this formula with the representation formula for the continuous wavelet transform [5, p. 33].

We also have the equality

$$
\begin{aligned}
\|f\|_{L_2(I)}^2 &= \|T_{-h}f\|_{L_2(I)}^2 \\
&= \left(\int_I f(y+h)\,dy\right)^2 + \sum_{j,k,\psi} |\langle T_{-h}f, \psi_{j,k}\rangle|^2 \\
&= \left(\int_I f(y+h)\,dy\right)^2 + \sum_{j,k,\psi} \left(\int_I f(y+h)\,2^k\psi(2^k y - j)\,dy\right)^2 \\
&= \left(\int_I f(y)\,dy\right)^2 + \sum_{j,k,\psi} \left(\int_I f(y)\,2^k\psi(2^k(y-h)-j)\,dy\right)^2.
\end{aligned}
$$

By similar arguments,

$$
\begin{aligned}
\|f\|_{L_2(I)}^2 &= \int_I \|T_{-h}f\|_{L_2(I)}^2\,dh \\
&= \left(\int_I f(y)\,dy\right)^2 + \int_I \sum_{j,k,\psi} \left(\int_I f(y)\,2^k\psi(2^k(y-h)-j)\,dy\right)^2 dh \\
&= \left(\int_I f(y)\,dy\right)^2 + \sum_{k,\psi} 2^{2k} \int_I \left(\int_I f(y)\,\psi_k(y-h)\,dy\right)^2 dh.
\end{aligned}
$$

Finally,

$$
\begin{aligned}
|f|_{B_1^1(L_1(I))} &= \int_I |T_{-h}f|_{B_1^1(L_1(I))}\,dh \\
&\asymp \int_I \sum_{j,k,\psi} \left|\int_I f(y)\,2^k\psi(2^k(y-h)-j)\,dy\right| dh \\
&= \sum_{k,\psi} 2^{2k} \int_I \left|\int_I f(y)\,\psi_k(y-h)\,dy\right| dh
\end{aligned}
$$

(14)

We note that this representation of f is *redundant*: the mapping

(15)
$$
f \to \left\{\int_I f(y)\,\psi_k(y-h)\,dy \mid h \in I, k \geq 0, \psi \in \Psi\right\}
$$

is a one-to-one mapping (modulo constants) from $L_2(I)$ to $L_2(\Omega)$, where $\Omega = I \times \mathbb{N} \times \Psi$ with measure $dh \otimes 2^{2k}\delta_k \otimes \delta_\psi$, where dh is Lebesgue measure and δ is point mass. However,

9

it is not *onto*—the image of $L_2(I)$ is a closed subspace of $L_2(\Omega)$, but it is *not* all of $L_2(\Omega)$. This is true also for the continuous wavelet transform. So, generally speaking, an arbitrary element of $L_2(\Omega)$ is not the image of any f under the mapping (15).

In practice, we are given only a finite amount of data, so we cannot calculate the above formulae for all $k \geq 0$ and all translations $h \in I$. Let's assume now that we are given 2^m rows of 2^m pixels, each of which is the average of f on a square of size $2^{-m} \times 2^{-m}$. Then using the orthogonal wavelets constructed by Daubechies [5], we can calculate these formulae for $k < m$ and average over 2^{2m} different pixel translations $j/2^m$, $j = (j_1, j_2)$, $0 \leq j_1, j_2 < 2^m$, instead of averaging over $h \in I$. With these changes, we get

$$f(x) = \int_I f(y)\,dy + \sum_{\substack{0 \leq k < m \\ j, \psi}} 2^{2(k-m)} \int_I f(y)\,\psi_k(y - \frac{j}{2^m})\,dy\,\psi_k(x - \frac{j}{2^m}),$$

$$\|f\|_{L_2(I)}^2 = \left(\int_I f(y)\,dy\right)^2 + \sum_{\substack{0 \leq k < m \\ j, \psi}} 2^{2(k-m)} \left(\int_I f(y)\,\psi_k(y - \frac{j}{2^m})\,dy\right)^2, \text{ and}$$

$$|f|_{B_1^1(L_1(I))} \asymp \sum_{\substack{0 \leq k < m \\ j, \psi}} 2^{2(k-m)} \left|\int_I f(y)\,\psi_k(y - \frac{j}{2^m})\,dy\right|.$$

(The fact that we begin with pixel values that are not, typically, the exact value of $\langle f, \phi_{j,m} \rangle$ introduces an error that is bounded in Section D of [2].)

Note that for each dyadic level k we need to compute 3×2^{2m} terms, one for each pixel and one for each $\psi \in \Psi$. A simple argument shows that we can use the wavelet rewrite rule to calculate each term in constant time (see [4]), so the entire calculation takes $O(m2^{2m})$ operations, as compared with $O(2^{2m})$ operations for the discrete wavelet transform.

8. Translation-Invariant Denoising

To simplify notation in this section, we assume without loss of generality that f has integral zero; if it does not, then we subtract the mean of f from f before proceeding.

Coifman and Donoho proposed applying wavelet shrinkage to the most recent representation of f written above in order to remove noise from a noisy image f. That is, they proposed to calculate

$$\sum_{\substack{0 \leq k < m \\ j, \psi}} 2^{2(k-m)} S_\lambda \left(\int_I f(y)\,\psi_k(y - \frac{j}{2^m})\,dy\right) \psi_k(x - \frac{j}{2^m})$$

as a smoothing transform for f. In the idealized case where one averages over all $h \in I$

10

rather than over all pixel translations j, the operator would be

$$\sum_{k,\psi} 2^{2k} \int_I S_\lambda (\int_I f(y)\,\psi_k(y-h)\,dy)\,\psi_k(x-h)\,dh$$

The question is now how to interpret these transforms.

If we denote this transform by $S_\lambda f$, then we see immediately that S_λ is *not* a semigroup, as opposed to wavelet shrinkage with the discrete wavelet transform. In particular, $S_\lambda S_\mu f \neq S_{\lambda+\mu} f$ for general f.

This poses a practical problem. Let's say that we shrink the wavelet coefficients by λ, but find that the smoothing is not enough. If we then shrink by μ, we find that the result is not the same as if we shrink in one step by $\lambda + \mu$. The semigroup property is important because we would like the image after total "shrinkage" of λ, say, not to depend on the substeps we take of size $\lambda_1, \dots, \lambda_n$ such that $\lambda = \lambda_1 + \cdots + \lambda_n$.

If we denote the above smoothing operators by $\tilde{S}_\lambda$, one can ask whether

$$\lim_{n\to\infty} \tilde{S}^n_{\lambda/n} f$$

exists. We show below that the answer is yes, and we denote the resulting operator by $\mathbf{S}_\lambda$. $\mathbf{S}_\lambda$ is a continuous contraction semigroup on $L_2(I)$ that defines a wavelet-based image smoothing scale space. If we denote

$$\mathcal{W}f(k,\psi,h) := \int_I f(y)\,\psi_k(y-h)\,dy$$

then $\mathbf{S}_\lambda$ formally satisfies

$$\frac{d\mathbf{S}_\lambda f}{d\lambda} = -\sum_{\substack{0\le k<m \\ j,\psi}} 2^{2(k-m)} \operatorname{sign}(\mathcal{W}\mathbf{S}_\lambda f(k,\psi,\tfrac{j}{2^m}))\,\psi_k(x-\tfrac{j}{2^m}).$$

Exactly the same issues arise when we average over all $h \in I$ instead of over all pixels; in this case, $\mathbf{S}_\lambda$ formally satisfies

$$\frac{d\mathbf{S}_\lambda f}{d\lambda} = -\sum_{k,\psi} 2^{2k} \int_I \operatorname{sign}(\mathcal{W}\mathbf{S}_\lambda f(k,\psi,h))\,\psi_k(x-h)\,dh.$$

We now show how a theorem of Kato and Masuda [13] can be used when averaging over all $h \in I$; a similar argument works when averaging over all j.

11

To do this, we note that the mapping $f \to \mathcal{W}f$ maps $L_2(I)$ to a closed subspace of $L_2(\Omega)$, which we call V. Conversely, for any $W(k, \psi, h)$ in $L_2(\Omega)$,

$$\tilde{P}W(x) := \sum_{k,\psi} 2^{2k} \int_I W(k, \psi, h) \psi_k(x - h) \, dh$$

is a function in $L_2(I)$. $\tilde{P}$ is a bounded linear projection from $L_2(\Omega)$ to $L_2(I)$. Thus, we can define an orthogonal projection P from $L_2(\Omega)$ to V by

$$PW(k, \psi, h) := \mathcal{W}\tilde{P}W(k, \psi, h).$$

We now recall that for the discrete wavelet transform, wavelet shrinkage is gradient descent along the discrete L_1 norm of the wavelet coefficients. In the same way, the operator $W(k, \psi, h) \to S_\lambda(W(k, \psi, h))$ is gradient descent in $L_2(\Omega)$ along the $L_1(\Omega)$ norm. That is, it is the solution operator of the problem

$$\frac{du}{dt} + \partial \|u\|_{L_1(\Omega)} \ni 0, \quad t \geq 0,$$
$$u(0) = W.$$

Furthermore, orthogonal projection $W \to P_V W$ onto a closed subspace V of $L_2(\Omega)$ is the solution operator of the problem

$$\frac{du}{dt} + \partial \chi_V(u) \ni 0, \quad t \geq 0,$$
$$u(0) = W,$$

for any positive t. Thus, our original operator can be written

$$\tilde{S}_\lambda f = \mathcal{W}^{-1} P_V S_\lambda(\mathcal{W}f),$$

and

$$\tilde{S}_{\lambda/n}^n f = \mathcal{W}^{-1} (P_V S_{\lambda/n})^n (\mathcal{W}f).$$

Example 4.1 in Kato and Masuda [13] shows that

$$\lim_{n \to \infty} (P_V S_{t/n})^n \mathcal{W}f$$

is the solution operator of

$$\frac{du}{dt} + \partial[\chi_V(u) + \|u\|_{L_1(\Omega)}] \ni 0, \quad t \geq 0,$$
$$u(0) = \mathcal{W}f.$$

12

This type of result is known as a Trotter product formula. Since, by §7, we can *define* $|u|_{B_1^1(L_1(I))}$ to be

$$|u|_{B_1^1(L_1(I))} := \|\mathcal{W}u\|_{L_1(\Omega)} + \chi_V(\mathcal{W}u)$$

(since $\chi_V(\mathcal{W}u) = 0$), we have that $\mathbf{S}_t f$ is the solution operator of

$$\frac{du}{dt} + \partial |u|_{B_1^1(L_1(I))} \ni 0, \quad t \geq 0,$$
$$u(0) = f.$$

9. Computations

First, we compute with the Haar wavelet. Although the Haar wavelet is orthogonal, it is *not* smooth enough for the equivalence between the $B_1^1(L_1(I))$ norm and the sequence norm given in (14) to hold. Thus, although the result given above shows that iterated translation-invariant wavelet shrinkage using the Haar wavelet can be interpreted as gradient descent along a norm given by the right-hand side of (14), that norm is not equivalent to the $B_1^1(L_1(I))$ norm.

Our second set of images are computed with the formally second-order accurate orthogonal wavelets introduced by Daubechies and described by [5]. Here, we *do* have the equivalence (14) for $B_1^1(L_1(I))$, so the theory of the previous section applies.

In the left of Figure 1, we show the effects of smoothing on a simple geometrical image. Note that with the Haar wavelet, horizontal and vertical edges and the corners between them are well preserved, while diagonal edges and other corners are smoothed more drastically. With the smooth, orthogonal, wavelets, all edges and corners are smoothed more uniformly. Thus, it appears that translation-invariant smoothing using the Haar wavelet does not correspond to gradient descent of a norm that is equivalent to a *rotationally-invariant* norm, as is the usual $B_1^1(L_1(I))$ norm.

In the right of Figure 1, we show the effects of smoothing on a "natural" image. Note that while the Haar wavelets achieve more of a "segmentation" of the image, horizontal and vertical lines are strongly preferred.

In Figure 2, we show a fingerprint image used in [2] as an example together with the same image after adding Gaussian noise with a standard deviation of 32 grey scales. In the left part of Figure 3, we show the noisy image after applying translation-invariant wavelet shrinkage with orthogonal wavelets with the (experimentally determined) optimal value of $t = 0.071188$. The optimal shrinkage parameter using plain wavelet shrinkage (in this case, applied to the biorthogonal wavelets of [3] and [12] as illustrated on page 272 of [5]) was estimated in [2] to be $t = 41.834798/512 = 0.081709$; the right half of Figure 3 repeats Figure 4 of [2], which shows the image after wavelet shrinkage by what is called the "critical" parameter in that paper, of $43.516416/512 = .084993$. (Note that
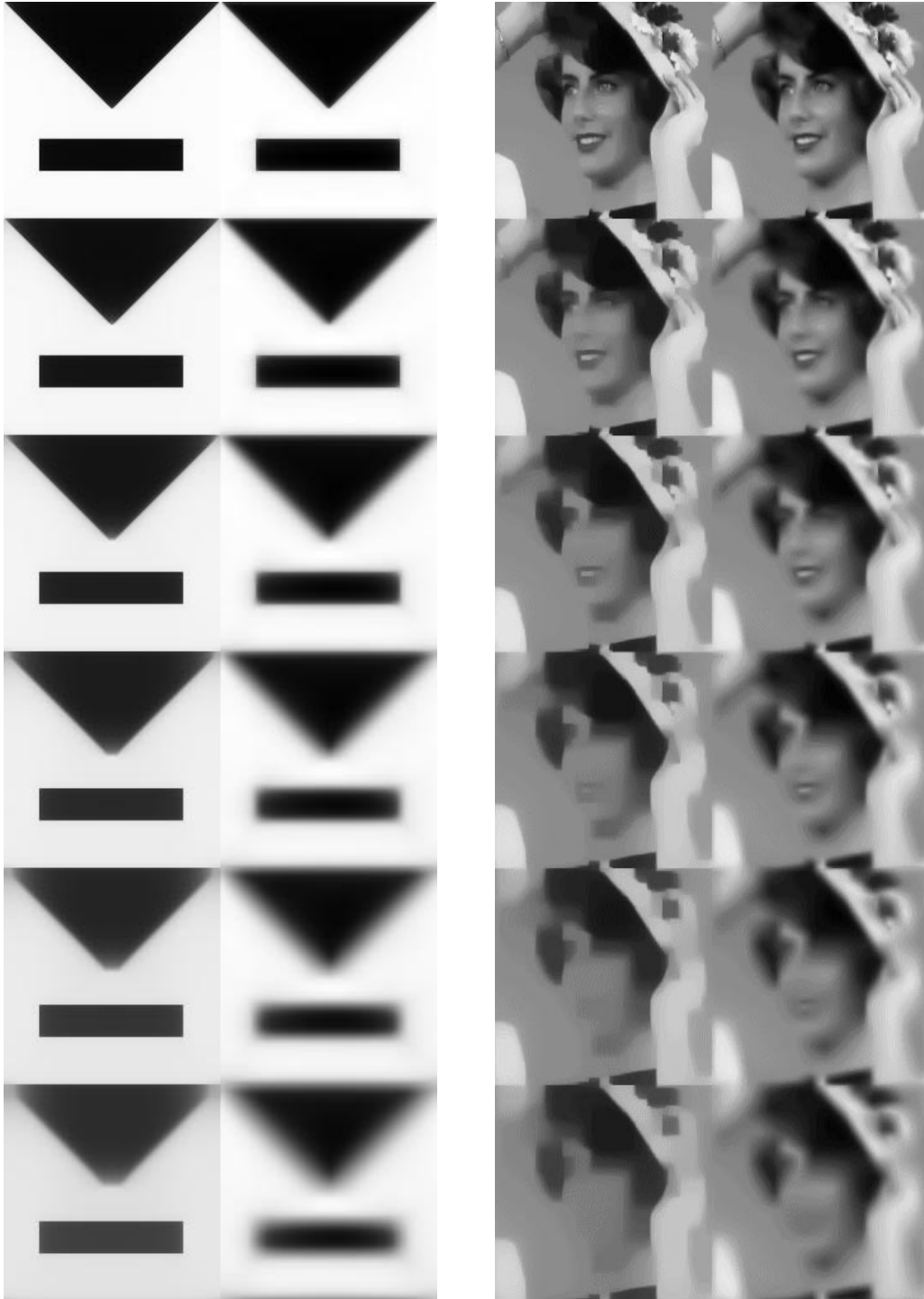
13

FIG. 1. Scale space smoothing. For each set of images, Haar is on the left and orthogonal Daubechies is on the right. Maximum $t = 3.0$ on the left, maximum $t = 0.75$ on the right.

FIG. 2. Original image on left, noisy image on right.

FIG. 3. Translation-invariant wavelet denoising on left, $t = 0.0711878$; wavelet shrinkage with "critical" shrinkage on right.

the shrinkage parameters reported in [2] must be divided by 512, the square root of the number of pixels, to compare to this paper.) It would seem, heuristically, that the process of averaging over all pixel translates of the image after wavelet shrinkage would provide some measure of smoothing beyond that effected by the wavelet shrinkage itself, so the smaller optimal shrinkage parameter in this paper does not surprise us; on the other hand, we can offer no argument to quantify this effect.

REFERENCES

[1] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Mathematics Studies, No. 5., North-Holland Publishing Co., Amsterdam-London, 1973.
[2] A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier, *Nonlinear wavelet image processing: Variational problems, compression, and noise removal through wavelet shrinkage*, IEEE Trans. Image

Processing, 7 (1998), pp. 319–335, Special Issue on Partial Differential Equations and Geometry-Driven Diffusion in Image Processing and Analysis.

[3] A. Cohen, I. Daubechies, and J.-C. Feauveau, *Biorthogonal bases of compactly supported wavelets*, Comm. Pure Appl. Math., 45 (1992), pp. 485–560.

[4] R. R. Coifman and D. L. Donoho, *Translation-invariant de-noising*, in Wavelets and Statistics, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York, 1995, pp. 125–150.

[5] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conference Series in Applied Mathematics 91, SIAM, Philadelphia, 1992.

[6] R. A. DeVore, B. Jawerth, and B. J. Lucier, *Image compression through wavelet transform coding*, IEEE Trans. Information Theory, 38, 2 (1992), pp. 719–746, Special issue on Wavelet Transforms and Multiresolution Analysis.

[7] R. A. DeVore and B. J. Lucier, *Wavelets*, Acta Numerica, 1 (1992), pp. 1–56.

[8] D. L. Donoho, *Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 101–126.

[9] D. L. Donoho and I. Johnstone, *Adapting to unknown smoothness via wavelet shrinkage*, J. Amer. Statist. Assoc., 90 (1995), pp. 1200–1224.

[10] M. Frazier and B. Jawerth, *A discrete transform and decompositions of distribution spaces*, J. of Functional Anal., 93 (1990), pp. 34–170.

[11] B. M. ter Haar Romeny, ed., *Geometry-driven Diffusion in Computer Vision*, Kluwer Academic Publishers, Dordrecht, 1994.

[12] C. Herley and M. Vetterli, *Biorthogonal bases of symmetric compactly supported wavelets*, in Wavelets, Fractals, and Fourier Transforms, Oxford Univ. Press, New York, 1993, pp. 91–1008.

[13] T. Kato and K. Masuda, *Trotter's product formula for nonlinear semigroups generated by the subdifferentials of convex functionals*, J. Math. Soc. Japan, 30 (1978), pp. 169–178.

[14] N.-Y. Lee and B. J. Lucier, *Wavelet methods for inverting the Radon transform with noisy data*, `http://www.math.purdue.edu/~lucier/692/tomography.pdf`.

[15] S. Mallat and S. Zhong, *Characterization of signals from multiscale edges*, IEEE Trans. on Patt. Anal. and Mach. Intell., 14 (1992), pp. 710–732.

[16] J.-M. Morel and S. Solimini, *Variational Methods in Image Segmentation*, Progress in Nonlinear Differential Equations and Their Applications—Volume 14, Birkhauser, Boston, MA, 1994.