Wavelet Footprints: Theory, Algorithms, and Applications

Pier Luigi Dragotti, *Member, IEEE,* and Martin Vetterli, *Fellow, IEEE*

Abstract—In recent years, wavelet-based algorithms have been successful in different signal processing tasks. The wavelet transform is a powerful tool because it manages to represent both transient and stationary behaviors of a signal with few transform coefficients. Discontinuities often carry relevant signal information, and therefore, they represent a critical part to analyze. In this paper, we study the dependency across scales of the wavelet coefficients generated by discontinuities. We start by showing that any piecewise smooth signal can be expressed as a sum of a piecewise polynomial signal and a uniformly smooth residual (see Theorem 1 in Section II). We then introduce the notion of footprints, which are scale space vectors that model discontinuities in piecewise polynomial signals exactly. We show that footprints form an overcomplete dictionary and develop efficient and robust algorithms to find the exact representation of a piecewise polynomial function in terms of footprints. This also leads to efficient approximation of piecewise smooth functions. Finally, we focus on applications and show that algorithms based on footprints outperform standard wavelet methods in different applications such as denoising, compression, and (nonblind) deconvolution. In the case of compression, we also prove that at high rates, footprint-based algorithms attain optimal performance (see Theorem 3 in Section V).

Index Terms—Compression, denoising, matching pursuit, nonlinear approximation, wavelets.

I. INTRODUCTION

THE design of a complete or overcomplete expansion that allows for compact representation of certain relevant classes of signals is a central problem in signal processing and its applications. Parsimonious representation of data is important for compression [14]. Furthermore, achieving a compact representation of a signal also means intimate knowledge of the signal features and this can be useful for many other tasks including denoising, classification, and interpolation. From a computational analysis point of view, one can say that the problem is to build a dictionary $\mathcal{D} = \{f_i\}_{i \in I}$ of elementary functions that can well approximate any signal in a given functional class $\mathcal{F}$ with the superposition of few of its elements.

The design of a dictionary with good approximation properties, however, is not the only important element. Together with $\mathcal{D}$, one also needs to develop fast algorithms that can efficiently find the sparsest representation of any signal $g \in \mathcal{F}$ in terms of the elements of $\mathcal{D}$. When $\mathcal{D} = \{f_i\}_{i \in I}$ is a basis, there is a unique way to express g as a linear combination of the f_is, and this representation can be easily found computing the inner products between g and the duals of f_i's.[1] Despite this nice property, overcomplete dictionaries are often preferred to basis expansions. Overcomplete dictionaries are more flexible; they can better adapt to the characteristics of the signal under consideration, and this allows for sparser signal representations. Examples of overcomplete dictionaries include best basis methods or adaptive wavelet packets [6], [26]. In the case of overcomplete bases, however, it is more difficult to develop fast algorithms that find the right sparse representation of a signal in $\mathcal{F}$. Because the elements of $\mathcal{D}$ are linearly dependent, there are infinitely many ways to express g as a linear combination of the f_i's. In a few cases, it is possible to arrive at sparse signal representations with linear complexity algorithms [12], [16], but in general, the search for the sparsest signal representation is an NP-complete problem [10]. Note that techniques based on singular value decomposition (SVD) and pseudo-inverse do not yield compact signal representations [17]. Other methods like basis pursuit [3] are usually computationally intensive; matching pursuit [21], which is a greedy iterative algorithm, is computationally efficient but does not converge in a finite number of iterations in general.

In this paper, we focus on the class of piecewise smooth signals. In particular, we will mostly consider piecewise polynomial signals. We propose a new representation of these functions in terms of objects we call *footprints* and make up an overcomplete dictionary of atoms. The footprints dictionary is built from the wavelet transform. Given a signal of interest, we first perform the wavelet transform of this signal, and then, the wavelet coefficients are expressed in terms of footprints. Together with the scaling coefficients, footprints can represent any piecewise polynomial signal. The main property of footprints is that they characterize efficiently the singular structures of the signal, which usually carry important information. Wavelets are also efficient at representing singularities [20]; however, the wavelet coefficients generated by a singularity are dependent across scales. By constructing the footprint expansion on the wavelet transform, we remove this dependency completely. Thus, by representing any discontinuity with the combination

P. L. Dragotti is with the Electrical and Electronic Engineering Department, Imperial College, London, UK (e-mail: p.dragotti@ic.ac.uk).

M. Vetterli is with the Audio-Visual Communications Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland, and with the Department of Electrical Engineering and Computer Science, University of California, Berkeley CA 94720 USA (e-mail: Martin.Vetterli@epfl.ch).

[1]The dual elements coincide with f_i if the dictionary is orthonormal.

of a few footprints, we can get a sparser representation of the signal under consideration.

Even though the footprint expansion is overcomplete, it can be made locally orthogonal, and this allows us to use fast algorithms to find the right sparse decomposition of the signal in terms of footprints. Alternatively, it is also possible to use matching pursuit. We show that there are situations in which matching pursuit with footprints can attain the sparsest signal representation with a finite number of iterations. Finally, we will see that the use of this dictionary leads to efficient algorithms for compression, denoising, and nonblind deconvolution of piecewise smooth signals.

The paper is organized as follows. Section II is meant to build up intuition about footprints. We analyze the dependency across scales of the wavelet coefficients generated by discontinuities and demonstrate a decomposition of a piecewise smooth signal into a piecewise polynomial signal and a regular residual (Theorem 1). This theorem will be invoked each time we will move from piecewise polynomial to piecewise smooth signals. In Section III, we present the footprint expansion, and in Section IV, we develop algorithms to efficiently represent piecewise polynomial signals in terms of footprints. In both sections, we study the case of piecewise constant signals in detail and then extend the analysis to the case of piecewise polynomial signals. Section V focuses on applications, namely, denoising, deconvolution, and compression. Traditional wavelet-based algorithms are reviewed, new algorithms based on footprints are presented, and their performance is analyzed. In Section VI, numerical simulations showing interesting improvements over traditional methods are presented, and conclusions are given in Section VII.

II. DEPENDENCY OF THE WAVELET COEFFICIENTS ACROSS SCALES

In wavelet-based signal processing, it is usually important to exploit the dependency across scales of the wavelet coefficients, and several efforts have been made in this direction recently; see, for instance, [2], [9], [22], and [28]. The singular structures of a signal often carry critical information, and thus, their efficient characterization is crucial in many signal processing tasks.

In this section, we review some of the properties of the wavelet transform, namely, its ability to characterize the local regularity of a function, and then, we focus on the analysis of the dependency of the wavelet coefficients generated by discontinuities.

Our interest is in piecewise smooth signals, that is, in signals that are made of regular pieces. The regularity of a function is usually measured with the Lipschitz exponent [23].[2] We say that the restriction of $f(t)$ to $[a, b]$ is uniformly Lipschitz $\alpha \geq 0$ over $[a, b]$ if there exists $K > 0$ such that for all $\nu \in [a, b]$, there exists a polynomial $p_\nu(t)$ of degree $m = \lfloor \alpha \rfloor$ such that

$$\forall t \in (a, b), \ |f(t) - p_\nu(t)| \leq K|t - \nu|^\alpha. \tag{1}$$

Therefore, we define a piecewise smooth function $f(t)$, $t \in [0, T[$ with $K + 1$ pieces, as follows:

$$f(t) = \sum_{i=0}^{K} f_i(t) \mathbf{1}_{[t_i, t_{i+1}[}(t) \tag{2}$$

[2]The so-defined Lipschitz exponent is sometimes called Hölder exponent.

where $t_0 = 0$, $t_{K+1} = T$, and $f_i(t)$ is uniformly Lipschitz α over $[t_i, t_{i+1}]$. Such signals are interesting because many signals encountered in practice can be modeled as piecewise smooth.

Consider now an orthonormal wavelet series with scale and shift parameters m and n, respectively. We use the convention that small scales correspond to large, negative m, that is

$$\psi_{m,n}(t) = \frac{1}{2^{m/2}} \psi\left(2^{-m} t - n\right) \ m, n \in \mathbb{Z}$$

where $\psi(t)$ is the wavelet basis function. Moreover, assume that the wavelet has k vanishing moments, that is

$$\int_{-\infty}^{\infty} t^d \psi(t) dt = 0, \ d = 0, 1, \ldots, k - 1.$$

Then, it follows that the wavelet coefficients of a function that is uniformly Lipschitz $\alpha \leq k$ on an interval $[a, b]$ decay across scales as $2^{m(\alpha+1/2)}$ [23]. The (local) decay property of the wavelet coefficients is at the heart of the success of the wavelet transform in several applications. Now, because of this decay property, larger wavelet coefficients tend to be around the singular parts of a signal, that is, around points with small Lipschitz coefficients. These wavelet coefficients gather most of the energy of the original signal, and for this reason, we are interested in modeling their behavior across scales. For instance, given a signal as in (2), we are interested in studying the wavelet coefficients related to the break points $t_i \ i = 1, 2, \ldots, K$.

To begin our analysis, we start by considering a particular subclass of piecewise smooth signals, namely, piecewise polynomial signals. A function $p(t) \ t \in [0, T[n$ is piecewise polynomial with $K + 1$ pieces if

$$p(t) = \sum_{i=0}^{K} p_i(t) \mathbf{1}_{[t_i, t_{i+1}[}(t) \tag{3}$$

where $t_0 = 0$, $t_{K+1} = T$, and $p_i(t) = \sum_{d=0}^{D} a_i^{(d)} t^d \ i = 0, 1, \ldots, K$ are polynomials of maximum degree D. Piecewise polynomial signals have a finite number of degrees of freedom and are easier to analyze. However, despite their simplicity, they can be used to efficiently approximate piecewise smooth functions. In fact, if the piecewise polynomial approximation is chosen properly, the approximation error shows interesting regularity properties.

Theorem 1: Given a piecewise smooth signal $f(t)$ defined as in (2), that is, with pieces of Lipschitz regularity α. Then, there exists a piecewise polynomial signal $p(t)$ with pieces of maximum degree $d = \lfloor \alpha \rfloor$ such that the difference signal $r_\alpha(t) = f(t) - p(t)$ is uniformly Lipschitz α over $[0, T[$.

Proof: See Appendix A. □

Theorem 1 indicates a practical way to deal with piecewise smooth signals. It shows that any piecewise smooth signal $f(t)$ can be expressed as the sum of a piecewise polynomial signal and a residual that is uniformly Lipschitz α. That is, $f(t) = p(t) + r_\alpha(t)$. Now, since the residual is regular, it can be well represented with wavelets (the wavelet decomposition of $r_\alpha(t)$ results in small coefficients with fast decay across scales). Therefore, the only elements we need to analyze are discontinuities

in the piecewise polynomial function and, in particular, the dependency across scales of the wavelet coefficients generated by these piecewise polynomial discontinuities.[3]

We start by considering the simple case of piecewise constant functions with only one discontinuity at location t_1 (i.e., $p(t) = a_0^{(0)} \mathbf{1}_{[0,t_1[}(t) + a_1^{(0)} \mathbf{1}_{[t_1,T[}(t))$ and a wavelet series with one vanishing moment and compact support. The decomposition of this signal in the wavelet basis results in zero wavelet coefficients, except for the coefficients in the cone of influence of t_1. Recall that the cone of influence of t_1 in the scale-space plane is the set of points (m, n) such that t_1 is included in the support of $\psi_{m,n}(t)$. In this case, the wavelet coefficients in this cone of influence are dependent: They have only one degree of freedom. This can be easily shown recalling that a wavelet with k vanishing moments and fast decay can be written as the kth-order derivative of a function θ, which also has a fast decay [23]. Thus, the following conditions are true: $\psi(t) = (-1)^k (d^k \theta(t)/dt^k)$ and $\psi_{m,n}(t) = (-1)^k 2^{km} (d^k \theta_{m,n}(t)/dt^k)$, where $\theta_{m,n}(t) = (1/2^{m/2}) \theta(2^{-m} t - n)$. Since the kth derivative of a function is well defined in the sense of distributions, it follows that

$$
\begin{aligned}
\langle p(t), \psi_{m,n}(t) \rangle &= 2^m \int_{-\infty}^{\infty} \frac{dp(t)}{dt} \theta_{m,n}(t) dt \\
&= 2^m \int_{-\infty}^{\infty} \left(a_1^{(0)} - a_0^{(0)} \right) \delta(t - t_1) \theta_{m,n}(t) dt
\end{aligned}
$$

where we used integration by parts to move the derivative from $\theta(t)$ to $p(t)$. That is, $\langle p(t), (d\theta(t)/dt) \rangle = -\langle (dp(t)/dt), \theta(t) \rangle$. Thus, if the wavelet has compact support, $\langle p(t), \psi_{m,n}(t) \rangle$ is equal to zero if $\psi_{m,n}(t)$ does not overlap t_1, and $\langle p(t), \psi_{m,n}(t) \rangle$ depends only on the difference $a_1^{(0)} - a_0^{(0)}$ otherwise. This means that the wavelet behavior across scales is deterministic. If one knows the value of a single nonzero wavelet coefficient in the cone of influence of t_1, one can derive from it all the other wavelet coefficients in that cone of influence.

This discussion generalizes to the case of piecewise polynomial signals with polynomials of maximum degree D. Consider the case of a piecewise polynomial function with one discontinuity at t_1 and polynomials $p_i(t) = \sum_{d=0}^{D} a_i^{(d)} t^d$, $i = 0, 1$. Compute the wavelet decomposition with a wavelet having $D + 1$ vanishing moments and compact support. Again, the nonzero wavelet coefficients are only in the cone of influence of t_1, and we have

$$
\begin{aligned}
\langle p(t), \psi_{m,n}(t) \rangle &= 2^{mk} \int_{-\infty}^{\infty} \frac{d^k p(t)}{dt^k} \theta_{m,n}(t) dt \bigg|_{k=D+1} \\
&= 2^{m(D+1)} \int_{-\infty}^{\infty} \sum_{d=0}^{D} c_d \delta^{(d)} \\
&\quad \cdot (t - t_1) \theta_{m,n}(t) dt
\end{aligned} \tag{4}
$$

where $\delta^{(d)}(t)$ is the dth derivative of the Dirac δ-function, and the coefficients c_d depend on the differences $\left(a_1^{(d)} - a_0^{(d)} \right)$, $d = 0, 1, \ldots, D$.[4] Thus, in the more general case, the wavelet

coefficients in the cone of influence of t_1 have only $D + 1$ degrees of freedom, and one can determine all these wavelet coefficients by knowing only $D + 1$ nonzero coefficients in that cone of influence.

In summary, the above analysis indicates that piecewise polynomial signals are well represented by wavelets but that it is possible to model piecewise polynomial discontinuities in a more efficient way. In the next section, we present a new way to express discontinuities in piecewise polynomial signals. Together, with Theorem 1, this will lead to efficient algorithms to represent piecewise smooth signals. Although, we could perform this analysis in continuous time, we concentrate on the discrete-time case. This is because our final target is to develop efficient algorithms that act on discrete-time signals.

Before concluding this section, we want to analyze the border effects. Since our signals are defined on a finite interval $[0, T]$, we need to extend them outside this interval in order to perform a wavelet decomposition. Several extensions are possible [23]. In our formulation, we make a periodic extension, that is, we assume that signals are T-periodic and that, on the period $[0, T]$, they are given by (2) and (3). Now, this extension creates an artificial discontinuity at $t = m \cdot T$, $m \in \mathbb{Z}$, and Theorem 1 does not guarantee that the periodic extension of $r_\alpha(t)$ is regular in $t = m \cdot T$. However, using higher order polynomials (i.e., polynomials of maximum degree $d = \lfloor \alpha \rfloor + 1$), one can easily generalize the result of Theorem 1 and guarantee regularity of $r_\alpha(t)$ over all $\mathbb{R}$ [15].

III. FOOTPRINT DICTIONARIES

We move from continuous-time to discrete-time signals and introduce the notion of footprints, which are scale-space vectors containing all the wavelet coefficients generated by particular polynomial discontinuities.[5] We show that any piecewise polynomial discontinuity is specified by the linear combination of a few footprints and that footprints can be interpreted as an overcomplete expansion with good approximation properties.

A. Preliminaries

For our discussion, we need to introduce two discrete-time wavelet operators. The first one is an orthonormal discrete-time wavelet decomposition with J levels.[6] This decomposition can be efficiently implemented with a critically subsampled octave-band filterbank [30]. Let $\psi_{jl}[n]$ denote the wavelet function at scale j and shift l and $\phi_{Jl}[n]$ the scaling function at shift l. This wavelet operator is linear and periodically shift-variant with period 2^J. The other operator is the wavelet frame obtained by shifting out (with corresponding equivalent filters) the subsamplers in the filterbank [30]. In this case, we denote the wavelet functions at scale j and shift l with $\tilde{\psi}_{jl}[n]$ and the scaling function at shift l with $\tilde{\phi}_{Jl}[n]$. This frame is shift invariant.

The discrete-time signals we consider are N-dimensional vectors defined over the interval $[0, N - 1]$. Now, the wavelet

---

[3]For simplicity, we call *piecewise polynomial discontinuity* a singularity between two polynomials.

[4]To be more precise, $c_d = \sum_{i=0}^{d} t_1^{d-i} \left(a_1^{(D-i)} - a_0^{(D-i)} \right)$.

[5]In continuous time, one can define footprints equivalently, but they are of infinite dimension and are of little computational value.

[6]For simplicity, we study only the orthogonal case. However, the notion of footprints easily generalizes to the case of biorthogonal wavelets.

operators defined above act in $l_2(\mathbb{Z})$; therefore, we need to modify them to act on $[0, N-1]$. As anticipated in the previous section, we use a periodic extension [23]; therefore, the wavelet basis becomes $\psi_{jl}^{per}[n] = \sum_{k=-\infty}^{\infty} \psi_{jl}[n + kN]$ and $\phi_{Jl}^{per}[n] = \sum_{k=-\infty}^{\infty} \phi_{Jl}[n + kN]$. Recall that for any $J \leq \log_2 N$, this set of periodic wavelets forms an orthogonal basis in $l_2([0, N-1])$[23]. The same extension applies to the wavelet frame, and in this case, we get a frame in $l_2([0, N-1])$.

Finally, our interest is in the class of piecewise smooth and piecewise polynomial signals, and the previous definitions of piecewise smooth and piecewise polynomial signals can be naturally extended to the discrete-time case. In particular, a discrete-time piecewise polynomial signal $p[n]$, $n \in [0, N-1]$ is given by $p[n] = \sum_{i=0}^{K} p_i[n]\mathbf{1}_{[k_i, k_{i+1}]}[n]$, where $k_0 = 0, k_{K+1} = N$ and $p_i[n]$ $i = 0, 1, \ldots, K$ is a sampled polynomial of maximum degree D.

Depending on the use of a wavelet basis or a wavelet frame, we have two different footprint dictionaries as analyzed in the next sections.

B. Footprints Built From a Wavelet Basis

In this section, we construct the footprint dictionary from a wavelet basis. First, we study the simple case of piecewise constant signals and Haar wavelets. In this particular setting, the footprint dictionary $\mathcal{D}$ is a biorthogonal basis. Then, we consider the more general case of piecewise polynomial signals and higher order wavelets. We show that in this case $\mathcal{D}$ is always overcomplete.

1) Piecewise Constant Signals: Consider a piecewise constant signal $x[n]$, $n \in [0, N-1]$ with only one discontinuity at position k, and consider a J-level wavelet decomposition of this signal with a Haar wavelet:

$$x[n] = \sum_{l=0}^{N/2^J-1} c_l \phi_{Jl}[n] + \sum_{j=1}^{J} \sum_{l=0}^{N/2^j-1} y_{jl} \psi_{jl}[n] \quad (5)$$

where $y_{jl} = \langle x, \psi_{jl} \rangle$, and $c_l = \langle x, \phi_{Jl} \rangle$.[7] Since the Haar wavelet has one vanishing moment and finite support, the nonzero wavelet coefficients of this decomposition are only in the cone of influence of k. Thus, (5) becomes

$$x[n] = \sum_{l=0}^{N/2^J-1} c_l \phi_{Jl}[n] + \sum_{j=1}^{J} y_{jk_j} \psi_{jk_j}[n]$$

where $k_j = \lfloor k/2^j \rfloor$. Moreover, as in the continuous-time case, all these coefficients depend only on the amplitude of the discontinuity at k. Thus, if one defines a vector that contains all of them, one can specify any other step discontinuity at k by multiplying this vector by the right factor. This consideration leads to the following definition (see also Fig. 1).

Definition 1: Given a piecewise constant signal x with only one discontinuity at position k, we call footprint $f_k^{(0)}$ the scale-space vector obtained by gathering together all the wavelet coefficients in the cone of influence of k and then imposing $\|f_k^{(0)}\| = 1$. Expressed in the wavelet basis, this

---

[7]Note that we are assuming N to be a power of 2, and this way, a wavelet decomposition with a Haar wavelet does not suffer from border effects.
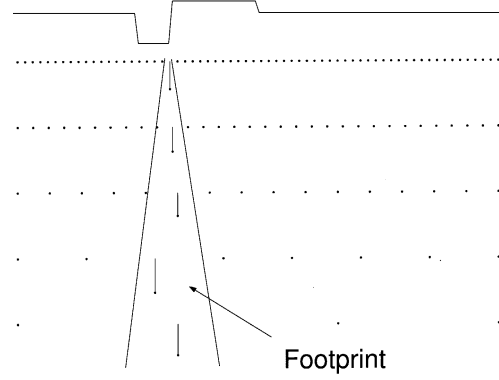

Fig. 1. Time domain (top) and wavelet domain (bottom) representation of the footprint $f_k^{(0)}$ with $N = 128$, $J = 5$ and $k = 41$. Notice that, except for the case $J = \log_2 N$ (N being a power of 2), a footprint does not look like a pure step edge function since the footprint definition does not include the scaling coefficients.

footprint can be written as $f_k^{(0)}[n] = \sum_{j=1}^{J} d_{jk_j} \psi_{jk_j}[n]$, where $d_{jk_j} = y_{jk_j}/\sqrt{\sum_{j=1}^{J} y_{jk_j}^2}$.

Now, any piecewise constant signal $x[n]$ with a step discontinuity at k can be represented in terms of the scaling functions $\phi_{Jl}[n]$ and of $f_k^{(0)}$. For instance, the signal $x[n]$ in (5) becomes

$$x[n] = \sum_{l=0}^{N/2^J-1} c_l \phi_{Jl}[n] + \alpha f_k^{(0)}[n] \quad (6)$$

where $\alpha = \langle x, f_k^{(0)} \rangle = \sum_{j=1}^{J} y_{jk_j} d_{jk_j}$. The above discussion can be repeated for any other step discontinuity at different locations, and for each location l, we have a different footprint $f_l^{(0)}$. Call $\mathcal{D} = \{f_k^{(0)}, k = 0, 1, \ldots, N-1\}$ the complete dictionary of footprints. Some of the properties of this dictionary depend on the number J of wavelet decomposition levels. For instance, just like the wavelet basis, footprints are shift variant unless the shift is equal to $m \cdot 2^J$, $m \in \mathbb{Z}$. That is

$$f_l^{(0)}[n] = f_k^{(0)}[n + k - l], if\ l - k = m \cdot 2^J, m \in \mathbb{Z}. \quad (7)$$

In addition, footprints are orthogonal to the scaling functions, but the orthogonality condition between footprints depends on the number J of wavelet decomposition levels. Assume $k = k' + m \cdot 2^J$, $l = l' + n \cdot 2^J$, and $l \geq k$. We have

$$\left\langle f_k^{(0)}, f_l^{(0)} \right\rangle = 0, \qquad \text{if } m \neq n$$
$$\left\langle f_k^{(0)}, f_l^{(0)} \right\rangle = \sqrt{\frac{k'(2^J - l')}{l'(2^J - k')}}, \quad \text{otherwise.} \quad (8)$$

Therefore, footprints related to neighboring discontinuities are biorthogonal. Finally, consider again (7). Since $f_0^{(0)}[n] = 0$, it follows that $f_{m \cdot 2^J}^{(0)}[n] = 0$. Thus, $\mathcal{D}$ contains only $N - N/2^J$ elements. Moreover, we have the following proposition.

Proposition 1: The elements of $\mathcal{D}$ together with the $N/2^J$ scaling functions $\phi_{Jl}[n]$, $l = 0, 1, \ldots, N/2^J - 1$ form a biorthogonal basis for $l_2([0, N-1])$.

Proof: See Appendix B $\qquad\qquad \square$

Therefore, any signal $x[n]$, $n \in [0, N-1]$ can be expressed in terms of footprints and scaling functions. In particular, if x is piecewise constant with K discontinuities, together with the scaling functions, K footprints are sufficient to represent it. This

can be shown by noticing that a piecewise constant signal with only one discontinuity can be expressed in terms of one footprint [see (6)], and piecewise constant signals with K discontinuities are given by the superposition of K piecewise constant signals with only one discontinuity. Therefore, the footprint representation of a signal x with K discontinuities at positions $k_1, k_2, \ldots, k_K$ is given by

$$x[n] = \sum_{l=0}^{N/2^J - 1} c_l \phi_{Jl}[n] + \sum_{i=1}^{K} \alpha_i f_{k_i}^{(0)}[n]. \tag{9}$$

Note how this representation is sparser than the corresponding representation in a wavelet basis that requires J times more wavelets than footprints if the cones of influence do not overlap. The problem of finding the discontinuity locations and the correct values α_i in (9) will be treated in detail in Section IV.

Finally, one may wonder if any object generated with the superposition of K footprints is piecewise constant with a number of discontinuities equal to the number of footprints. That is, are $\mathcal{D}$ and the scaling functions an unconditional basis for the class of piecewise constant signals? It turns out that this property is satisfied only when $J = \log_2 N$ (where N is a power of 2).

Proposition 2: For $N = 2^J$, the scaling function $\phi_{J0}[n]$ and the $N - 1$ footprints $f_k^{(0)}$, $k = 1, 2, \ldots, N - 1$ represent a biorthogonal basis that is unconditional for the class of piecewise constant signals defined over $[0, N - 1]$.

Proof: The biorthogonality comes from Proposition 1. We only need to show that this basis is unconditional. That is, assume that x is a piecewise constant signal with discontinuities at $k_1, k_2, \ldots, k_K$, and consider its representation in terms of footprints: $x[n] = c_0 \phi_{J0}[n] + \sum_{i=1}^{K} \alpha_i f_{k_i}^{(0)}$. We need to show that for any set of coefficients $\hat{\alpha}_i$ satisfying $|\hat{\alpha}_i| \leq |\alpha_i|$, the signal $\hat{x}[n] = c_0 \phi_{J0}[n] + \sum_{i=1}^{K} \hat{\alpha}_i f_{k_i}^{(0)}$ is still piecewise constant with discontinuity locations $k_1, k_2, \ldots, k_K$. This can be seen by noticing that for $J = \log_2 N$, $\phi_{J0}[n]$ is a constant function, and $f_k^{(0)}[n]$ is piecewise constant with one discontinuity at k ($k = 1, 2, \ldots, N - 1$). Therefore, any linear combination of f_{k_i} $k_i \in \{k_1, k_2, \ldots, k_K\}$ gives a piecewise constant signal with discontinuity locations $k_1, k_2, \ldots, k_K$. $\square$

1) Piecewise Polynomial Signals: We now generalize the above discussion to the case of discrete-time piecewise polynomial signals with polynomials of maximum degree D. We show that in this context, each discontinuity is represented by $D + 1$ footprints rather than one footprint.

Consider orthogonal wavelets with at least $D + 1$ vanishing moments and compact support L, and consider a piecewise polynomial signal $x[n]$ with only one discontinuity at k. Its J-level wavelet decomposition with periodic wavelets is

$$x[n] = \sum_{l=0}^{N/2^J - 1} c_l \phi_{Jl}^{per}[n] + \sum_{j=1}^{J} \sum_{l=0}^{N/2^j - 1} y_{jl} \psi_{jl}^{per}[n]. \tag{10}$$

First, notice that the periodic extension of the wavelet basis creates a second discontinuity at location zero and that this is a polynomial discontinuity. Thus, the nonzero wavelet coefficients of this expansion are only in the cone of influence of k and in the cone of influence of zero. Assume, for now, that

$0 \ll k \ll N$ and $2^J \ll N$ so that there are no wavelet coefficients in common between these two cones of influence. We can write

$$x[n] = \sum_{l=0}^{N/2^J - 1} c_l \phi_{Jl}^{per}[n] + \sum_{j,l \in I_0} y_{jl} \psi_{jl}^{per}[n] + \sum_{j,l \in I_k} y_{jl} \psi_{jl}^{per}[n] \tag{11}$$

where I_k is the set of indices (j, l), which are in the cone of influence of k, and I_0 is the set of indices (j, l), which are in the cone of influence of zero. It is easy to verify that there are $J \times (L - 1)$ wavelet coefficients in each cone of influence. From (4), we know that the wavelet coefficients in each of these cones of influence have only $D + 1$ degrees of freedom. Thus, we want to find a set of $D + 1$ footprints that can characterize these coefficients. To build this set of footprints, we resort to time-domain analysis.

The class of piecewise polynomial signals with one discontinuity at a fixed position $k \in [0, N - 1]$ forms a linear space of dimension $2(D + 1)$, and a possible basis for that space is represented by the following vectors:

$$P^{(d)}[n] = n^d, d = 0, 1, \ldots, D; n \in [0, N - 1]$$
$$T_k^{(d)}[n] = \mathbf{1}_{[k, N-1]}(n - k + 1)^d, d = 0, 1, \ldots, D, n \in [0, N - 1].$$

We can express these signals in a wavelet basis, and we have

$$P^{(d)}[n] = \sum_{l=0}^{N/2^J - 1} b_l^{(d)} \phi_{Jl}^{per}[n] + \sum_{j,l \in I_0} p_{jl}^{(d)} \psi_{jl}^{per}[n]$$

$$T_k^{(d)}[n] = \sum_{l=0}^{N/2^J - 1} c_{kl}^{(d)} \phi_{Jl}^{per}[n] + \sum_{j,l \in I_0} t_{jl}^{(d)} \psi_{jl}^{per}[n]$$
$$+ \sum_{j,l \in I_k} t_{jl}^{(d)} \psi_{jl}^{per}[n] \tag{12}$$

where we have used the fact that the nonzero wavelet coefficients of $P^{(d)}[n]$ are only in I_0, whereas the nonzero wavelet coefficients of $T_k^{(d)}$ are in the cones of influence of k and zero. Now, any signal $x[n]$ in this class can be written as

$$x[n] = \sum_{d=0}^{D} \alpha_0^{(d)} P^{(d)}[n] + \sum_{d=0}^{D} \alpha_k^{(d)} T_k^{(d)}[n]. \tag{13}$$

Therefore, combining (11)–(13) and considering only the elements in I_k, we have

$$\sum_{j,l \in I_k} y_{jl} \psi_{jl}^{per}[n] = \sum_{d=0}^{D} \alpha_k^{(d)} \sum_{j,l \in I_k} t_{jl}^{(d)} \psi_{jl}^{per}[n]. \tag{14}$$

Call $f_k^{(d)}[n] = \sum_{j,l \in I_k} t_{jl}^{(d)} \psi_{jl}^{per}[n]$ the scale-space vector gathering the $J \times (L - 1)$ wavelet coefficients generated by the discontinuity in $T_k^{(d)}$. Equation (14) shows that the wavelet coefficients generated by any polynomial discontinuity at k are characterized by a linear combination of $f_k^{(d)}$. This indicates that the wavelet coefficients in the cone of influence of a polynomial discontinuity have only $D + 1$ degrees of freedom and proves that these coefficients lie on a subspace of dimension $D + 1$. The vectors $f_k^{(d)}$, $d = 0, 1, \ldots, D$ span that subspace and can represent the set of footprints for which we are looking. However, it is

always better to have orthogonal bases; therefore, the footprints we will consider are obtained by applying a Gram–Schmidt orthogonalization process to $f_k^{(d)}$. Thus, from the above discussion, we have the following proposition.

Proposition 3: Given a piecewise polynomial signal with polynomials of maximum degree D and with one discontinuity at position k, the $J \times (L-1)$ nonzero wavelet coefficients in the cone of influence of that discontinuity lie on a sub-space of dimension $D+1$.

Definition 2: We call footprints $f_k^{(d)} d = 0, 1, \ldots, D$ the elements of an orthogonal vector basis that spans the subspace of dimension $D+1$ generated by a polynomial discontinuity at k. Footprints are obtained by gathering together all the nonzero wavelet coefficients generated by the discontinuity in $T_k^{(d)} d = 0, 1, \ldots, D$ and then imposing the following two conditions:

$$\left\| f_k^{(d)} \right\| = 1 \; d = 0, 1, \ldots, D$$
$$\left\langle f_k^{(i)}, f_k^{(j)} \right\rangle = \delta_{ij} i = 0, 1, \ldots, D; j = 0, 1, \ldots, D.$$

With this set of footprints, we can characterize any polynomial discontinuity at position k. In particular, (10) can be written as

$$x[n] = \sum_{l=0}^{N/2^J - 1} c_l \phi_{Jl}^{per}[n] + \sum_{j,l \in I_0} y_{jl} \psi_{jl}^{per}[n] + \sum_{d=0}^{D} \alpha^{(d)} f_k^{(d)}[n]$$

where $\alpha^{(d)} = \langle x, f_k^{(d)} \rangle$, $d = 0, 1, \ldots, D$.[8] With a similar analysis, we can create a different set of $D+1$ footprints to characterize a polynomial discontinuity at a different location. To characterize any polynomial discontinuity (including the discontinuity in zero), we need a dictionary $\mathcal{D} = \{f_k^{(d)}, d = 0, 1, \ldots, D; k = 0, 1, \ldots, N-1.\}$ of $(D+1)N$ footprints. With this dictionary of footprints and with the scaling functions, we can represent any piecewise polynomial signal. In particular, a signal x with K discontinuities at locations $k_1, k_2, \ldots, k_K$ is given by

$$x[n] = \sum_{l=0}^{N/2^J - 1} c_l \phi_{Jl}^{per}[n] + \sum_{i=0}^{K} \sum_{d=0}^{D} \alpha_i^{(d)} f_{k_i}^{(d)}[n] \quad (15)$$

where $k_0 = 0$ is the discontinuity due to the periodic extension. Note again how this representation is sparser than the corresponding representation in a wavelet basis.

As for the case of piecewise constant signals, footprints are orthogonal to the scaling functions, but footprints related to close discontinuities are biorthogonal. In particular, we have that $\langle f_l^{(d)}, f_k^{(c)} \rangle = 0$ for $|l - k| > (L-1) \cdot 2^J$. Moreover, footprints are periodically shift-invariant of period 2^J, and hence

$$f_l^{(d)}[n] = f_k^{(d)}[n + k - l]$$
$$if \; l - k = m \cdot 2^J, m \in \mathbb{Z}, d = 0, 1, \ldots, D. \quad (16)$$

It is also of interest to note that due to the periodic extension, the coefficients $\alpha_i^{(d)}$ in (15) are not independent. For instance, for $D = 0$, it follows that $\alpha_0^0 = -\sum_{i=1}^{K} w_{k_i}^0 \alpha_{k_i}^0$, where the weights

$w_{k_i}^0$ depend on the normalization in Definition 2 (without normalization, it would be $w_{k_i}^0 = 1$). In general, we have that

$$\alpha_0^0 = -\sum_{i=1}^{K} \sum_{d=0}^{D} w_{k_i}^d \alpha_{k_i}^d \quad (17)$$

where the weights $w_{k_i}^d$ depend on the orthogonalization process in Definition 2.[9] For this reason, to extend Proposition 2 to the piecewise polynomial case, we need to consider the constraint in (17). Thus, we have the following proposition.

Proposition 4: For $J = \log_2 N$, any linear combination of $\phi_{J0}^{per}[n]$ and of the $(D+1)N$ footprints $f_k^{(d)}$, which verifies (17), gives a piecewise polynomial signal.

Proof: We want to show that given a piecewise polynomial signal $x[n]$ represented as in (15), for any set of coefficients $\hat{\alpha}_i^{(d)}$ satisfying $|\hat{\alpha}_i^{(d)}| \leq |\alpha_i^{(d)}|$ and (17), the signal $\hat{x}[n] = c_0 \phi_{J0}^{per}[n] + \sum_{i=0}^{K} \sum_{d=0}^{D} \hat{\alpha}_i^{(d)} f_{k_i}^{(d)}[n]$ is still piecewise polynomial with discontinuity locations $k_1, k_2, \ldots, k_K$. This can be proved using arguments similar to that of Proposition 2. The scaling function $\phi_{J0}[n]$ is constant. Moreover, any pair of footprints $\hat{\alpha}_0^{(0)} f_0^{(0)}[n] + f_{k_i}^{(d)}[n]$, with $\hat{\alpha}_0^0$ satisfying (17), represent a piecewise polynomial signal with one discontinuity at $k_i \in \{k_1, k_2, \ldots, k_K\}$. Therefore, any linear combination of these pairs of footprints and of $\phi_{J0}[n]$ gives a piecewise polynomial signal with discontinuities at $k_1, k_2, \ldots, k_K$. $\square$

Strictly speaking, Proposition 4 shows conditions under which any linear combination of footprints leads to piecewise polynomial signals, but it does not prove that footprints are an unconditional expansion for the class of piecewise polynomial signals. However, in the rest of the paper, for simplicity, we will say that dictionaries of footprints satisfying the hypotheses of Proposition 4 are unconditional for the class of piecewise polynomial signals.

C. Footprints Built From a Wavelet Frame

We have constructed a dictionary of $(D+1)N$ footprints that can efficiently represent piecewise polynomial signals. However, this representation, like the wavelet transform, is not shift-invariant. In some settings, it is useful to have a shift-invariant dictionary. Such a dictionary can be constructed by simply replacing the wavelet basis with the wavelet frame. In particular, let $x[n]$ be a piecewise constant signal with only one discontinuity at k. We have

$$x[n] = \sum_{l=0}^{N/2^J - 1} c_l \tilde{\phi}_{Jl}^{per}[n] + \sum_{j,l \in \tilde{I}_0} y_{jl} \tilde{\psi}_{jl}^{per}[n]$$
$$+ \sum_{j,l \in \tilde{I}_k} y_{jl} \tilde{\psi}_{jl}^{per}[n]$$

where we have again used the fact that the nonzero coefficients are only in the cones of influence of k and 0. In this case, the cone of influence of k contains $J \times (L_j - 1)$ coefficients, where L_j is the length of the equivalent filter at level j. Moreover, y_{jl} is given by $y_{jl} = \langle x, \hat{\psi}_{jl}^{per} \rangle$, where $\{\hat{\phi}, \hat{\psi}\}$ is the dual

---

[8]In the case of biorthogonal wavelets, it would be $\alpha^{(d)} = \langle x, \hat{f}_k^{(d)} \rangle$ with $\hat{f}_k^{(d)} = \sum_{j,l \in I_k} d_{jl} \hat{\psi}_{jl}^{per}[n]$, where $\hat{\psi}_{jl}^{per}$ is the dual of ψ_{jl}^{per}.

[9]The easiest way to verify this property is by noticing that if we take the $(D+1)$th-order derivative of a periodic discrete-time piecewise polynomial signal, the sum of the resulting nonzero coefficients is always zero.

TABLE I
FOOTPRINT DICTIONARIES

	piecewise constant signals and Haar wavelet	piecewise polynomial signals and wavelet basis	piecewise polynomial signals and wavelet frame
dictionary properties	complete, shift variant, unconditional if $J = \log_2 N$.	overcomplete, shift variant, unconditional if $J = \log_2 N$ and Eq. (17) is verified.	overcomplete, shift invariant, unconditional if $J = \log_2 N$ and Eq. (17) is verified.

frame of $\{\tilde{\phi}, \tilde{\psi}\}$. Now, the shift-invariant footprint related to location k is given by $\tilde{f}_k^{(0)}[n] = \sum_{j,l \in \tilde{I}_k} d_{jl} \tilde{\psi}_{jl}^{per}[n],$, where $d_{jl} = y_{jl}/\sqrt{\sum_{j,l \in \tilde{I}_k} y_{jl}^2}$. The other footprints can be designed in the same way, and it follows that $\tilde{f}_k^{(0)}[n] = \tilde{f}_0^{(0)}[n-k]$. That is, all footprints are shifted versions of one footprint. If J is chosen such that $\langle \tilde{f}_0^{(0)}, \tilde{f}_k^{(0)} \rangle = 0$, $x[n]$ can be expressed as

$$x[n] = \sum_{l=0}^{N/2^J-1} c_l \tilde{\phi}_{Jl}^{per}[n] + \alpha_0 \tilde{f}_0^{(0)} + \alpha_k \tilde{f}_k^{(0)}$$

where $\alpha_k = \langle x, \hat{f}_k^{(0)} \rangle$, and $\hat{f}_k^{(0)}[n] = \sum_{j,l \in \tilde{I}_k} d_{jl} \hat{\psi}_{jl}^{per}[n]$. In the same way, we can design the footprint dictionary related to higher order polynomials. In this case, one has to consider the signals $T_k^{(d)}$ and their transforms with a wavelet frame. The footprints $\tilde{f}_k^{(d)}$ at location k are obtained following the same procedure as illustrated in Definition 2. Finally, given the dictionary $\mathcal{D} = \{\tilde{f}_k^{(d)}, d = 0, 1 \ldots, D; k = 0, 1, \ldots, N-1\}$, we have that $\tilde{f}_k^{(d)}[n] = \tilde{f}_0^{(d)}[n+k], d = 0, 1, \ldots, D$. As in the previous case, any piecewise polynomial signal can be expressed in terms of this dictionary, and we have

$$x[n] = \sum_{l=0}^{N/2^J-1} c_l \tilde{\phi}_{Jl}^{per}[n] + \sum_{i=0}^{K} \sum_{d=0}^{D} \alpha_i^{(d)} \tilde{f}_{k_i}^{(d)}[n]. \qquad (18)$$

IV. REPRESENTATION ALGORITHMS

In the previous sections, we have constructed different dictionaries of footprints according to the kind of wavelets involved (i.e., wavelet bases or wavelet frames) and to the class of signals considered (i.e., piecewise constant or piecewise polynomial signals). The main characteristics of these dictionaries are summarized in Table I. Before focusing on the representation algorithms, we want to mention that the space required to store these footprints dictionaries is not high since it grows only linearly with the size N of the signal. In particular, in the case of shift-variant footprints, the required storage space is of the order of $(L-1) \cdot J \cdot (D+1) \cdot 2^J$ coefficients, where $(L-1) \cdot J$ are the wavelet coefficients contained in each footprint, and $(D+1) \cdot 2^J$ are the number of footprints one has to store since the others are shifted version of those [see (16)]. Therefore, when $J = \log_2 N$ (worst case), we have that the required memory space grows like $N \log_2 N$. Similar results apply to the case of shift invariant footprints.

Now, we need to develop a fast and robust algorithm that can find the right representation of piecewise polynomial signals in terms of footprints. The algorithms that we present are valid for any of the families of footprints in Table I. However, for simplicity, we study only the case of footprints built from a wavelet basis, the extension to the wavelet frame being straightforward.

Consider a piecewise polynomial signal x with polynomials of degree D and with K discontinuities at $k_1, k_2, \ldots, k_K$. We have seen that this signal can be written as in (15). Thus, our target is to develop an algorithm that can find this representation of x. In our analysis, we do not consider the scaling functions since coefficients c_l in (15) are always given by $c_l = \langle x, \phi_{Jl}^{per} \rangle$.

We present two different approaches. The first one is a variation of the traditional matching pursuit algorithm. We show that in particular situations, this method can arrive at the correct representation of $x[n]$ in a finite number of iterations. The second approach is in spirit similar to matching pursuit, but it uses the property that the orthogonality condition between footprints depends on the number J of decomposition levels. We show that with a slight increase in complexity, this second algorithm always attains the correct signal representation with $\lceil K/2 \rceil$ iterations, where K is the number of discontinuities in the signal.

A. Matching Pursuit With Footprints

Matching pursuit [21] is a greedy iterative algorithm that derives sparse approximated representations of a signal x in terms of a given dictionary $\mathcal{D}$ of unit norm vectors.

Assume that $\mathcal{D}$ is the footprint dictionary and that $x[n]$ is a piecewise polynomial signal. Matching pursuit can be used to approximate x with $\mathcal{D}$. We know that the wavelet coefficients generated by a single polynomial discontinuity at k lie on a subspace of size $D+1$ and that this subspace is spanned by the footprints $f_k^{(d)}, d = 0, 1, \ldots, D$ (Proposition 3). Hence, instead of using the usual matching pursuit that projects the signal on single vectors, we employ a subspace pursuit, where the signal is projected on different subspaces.

In the first iteration, for each possible discontinuity location $k \in [0, N-1]$, the algorithm computes the $D+1$ inner products $\langle x, f_k^{(d)} \rangle, d = 0, 1, \ldots, D$ and chooses the location k_0 such that $\sum_{d=0}^{D} |\langle x, f_{k_0}^{(d)} \rangle|^2$ is maximum. Then, x can be written as its projection onto $f_{k_0}^{(d)}, d = 0, 1, \ldots, D$ and a residual R_x^1: $x = \sum_{d=0}^{D} \langle x, f_{k_0}^{(d)} \rangle f_{k_0}^{(d)} + R_x^1$. Since footprints related to the same discontinuity location are orthogonal (Definition 2), we can write $\|x\|^2 = \|R_x^1\|^2 + \sum_{d=0}^{D} |\langle x, f_{k_0}^{(d)} \rangle|^2$. Therefore, by

choosing k_0 such that $\sum_{d=0}^{D} |\langle x, f_{k_0}^{(d)} \rangle|^2$ is maximum, we minimize the norm of the error R_x^1. The algorithm is then iterated on the residual.

Note that for $D = 0$ (piecewise constant signals), the subspace pursuit reduces to the traditional matching pursuit. The subspace pursuit with the footprint dictionary has the same drawback as a typical matching pursuit algorithm, that is, it is not guaranteed to converge in a finite number of steps. However, there exist situations in which it obtains the exact representation of x in a finite number of iterations. In fact, one can easily verify the following theorem [15].

Theorem 2: Given a piecewise polynomial signal with K discontinuities at $k_1, k_2, \ldots, k_K$. If the distance between the two closest discontinuities is larger than $(L - 1) \cdot 2^J$, subspace pursuit with footprints obtains the exact footprint representation of x in K iterations.

B. Adaptive Depth Footprint Pursuit

The basic intuition behind Theorem 2 is that the number of decomposition levels J should be chosen according to the distance between discontinuities. If J is chosen properly, one can get the correct representation of x in a few iterations with a very simple method like matching pursuit. The problem is that we do not know *a priori* the discontinuity locations. Therefore, we propose a new algorithm, where we first find the discontinuity locations and then estimate the footprint coefficients related to those discontinuities. For simplicity, we concentrate on the case of piecewise constant signals and Haar wavelets. Assume, for instance, that x has K discontinuities at positions $k_1, k_2, \ldots, k_K$:

$$x[n] = \sum_{l=0}^{N/2^J - 1} c_l \phi_{Jl}[n] + \sum_{i=1}^{K} \alpha_i f_{k_i}^{(0)}[n] \quad (19)$$

and that the footprint dictionary is chosen with $J = \log_2 N$: $\mathcal{D} = \{f_k^{(0)} = \sum_{j=1}^{J} d_{jk_j} \psi_{jk_j}; k = 0, 1, \ldots, N-1\}$.[10] The discontinuity locations $k_1, k_2, \ldots, k_K$ are found in the following way.

```
Algorithm IV.1: (Location Estimation)
1) Compute the dual basis of D and Call
f̃_k^(0) k = 1,2,...,n - 1 the elements of this
dual basis. ¹¹
2) Compute the inner products ⟨x, f̃_k^(0)⟩,
k = 1,2,...,N - 1. The discontinuity loca-
tions correspond to the indexes of the
basis' elements which have nonzero inner
products with x.
```

Now that $k_1, k_2, \ldots, k_K$ are known, we need to evaluate the coefficients α_i. The footprint coefficients are evaluated with an iterative method that is, in spirit, similar to matching pursuit. At each iteration, we choose J such that the footprints related to the

[10]It is worth pointing out that in this case, $\mathcal{D}$ is a biorthogonal basis, and therefore, the exact representation of x can be found using the dual basis of $\mathcal{D}$. However, this solution is not robust to noise and does not generalize to piecewise polynomial signals. Therefore, it will not be considered here.

[11]It is of interest to note that this dual basis turns out to be a first-order derivative.

TABLE II
DENOISING OF PIECEWISE LINEAR SIGNALS WITH NO MORE THAN THREE DISCONTINUITIES

N	64	128	256	512
Footprints	17.4dB	19.9dB	21.9dB	24.5dB
Hard thresholding	15.1dB	17.1dB	18.7dB	21.4dB
Cycle spinning	17.9dB	20dB	22dB	24.5dB
Cycle-spin footprints	18.4dB	20.4dB	22.5dB	24.9dB

two closest discontinuities are orthogonal, we estimate the footprints coefficients of these two discontinuities, and iterate the process on the residual. At each iteration, we do not project the signal directly on the two closest footprints; instead, we compute the two dual footprints and project the signals on these two dual elements. The complete algorithm operates as follows.

```
Algorithm IV.2: (coefficient Estimaton)
1) Call K = {k_1,k_2,...,k_k} the set of esti-
mated discontinuity locations.
2) Assume that k_{m-1} and k_m are the two
closest discontinuities in K. Choose
J_1 = ⌊log_2(k_m - k_{m-1} )⌋.
3) Call f̂_{k_m}^(0) the sub-footprint obtained by
considering only the first J_1 elements of
f_{k_m}^(0). That is f̂_{k_m}^(0) = ∑_{j=1}^{J_1} d_{jk_{mj}} ψ_{jk_{mj}}. Define,
in the same way, the sub-footprint f̂_{k_{m-1}}^(0).
4) The sub-footprint f̂_{k_{m-1}}^(0) is orthog-
onal to f_k^(0), k ∈ K - {k_{m-1}} and veri-
fies ⟨f̂_{k_{m-1}}^(0), f_{k_{m-1}}^(0)⟩ = ||f̂_{k_{m-1}}^(0)||². Likewise, the
sub-footprint f̂_{k_m}^(0) is orthogonal to f_k^(0),
k ∈ K - {k_m} and ⟨f̂_{k_m}^(0), f_{k_m}^(0)⟩ = ||f̂_{k_m}^(0)||². Thus, the
contributions α_{m-1} and α_m are given by
```

$$\alpha_{k_{m-1}} = \frac{1}{\|\hat{f}_{k_{m-1}}^{(0)}\|} \left\langle x, \frac{\hat{f}_{k_{m-1}}^{(0)}}{\|\hat{f}_{k_{m-1}}^{(0)}\|} \right\rangle$$

$$\alpha_{k_m} = \frac{1}{\|\hat{f}_{k_m}^{(0)}\|} \left\langle x, \frac{\hat{f}_{k_m}^{(0)}}{\|\hat{f}_{k_m}^{(0)}\|} \right\rangle. \quad (20)$$

```
5) Remove k_{m-1},k_m from K and subtract the
two estimated contributions from the orig-
inal signal: R_x^1 = x - α_{k_{m-1}}f_{k_{m-1}}^(0) - α_{k_m}f_{k_m}^(0).
6) If K is not empty, iterate the process
on the residual; otherwise, stop.
```

Notice that since, at each iteration, we estimate two footprint coefficients, the algorithm ends after $\lceil K/2 \rceil$ iterations. Therefore, we are guaranteed that the algorithm converges after a finite number of steps. The interesting point of this algorithm is that, at each iteration, it is very easy to find the pair of dual footprints related to the footprints under consideration. There are two other advantages of this algorithm compared with matching pursuit. First, at each iteration, we choose the largest possible J_1 such that the footprints related to the two closest discontinuities
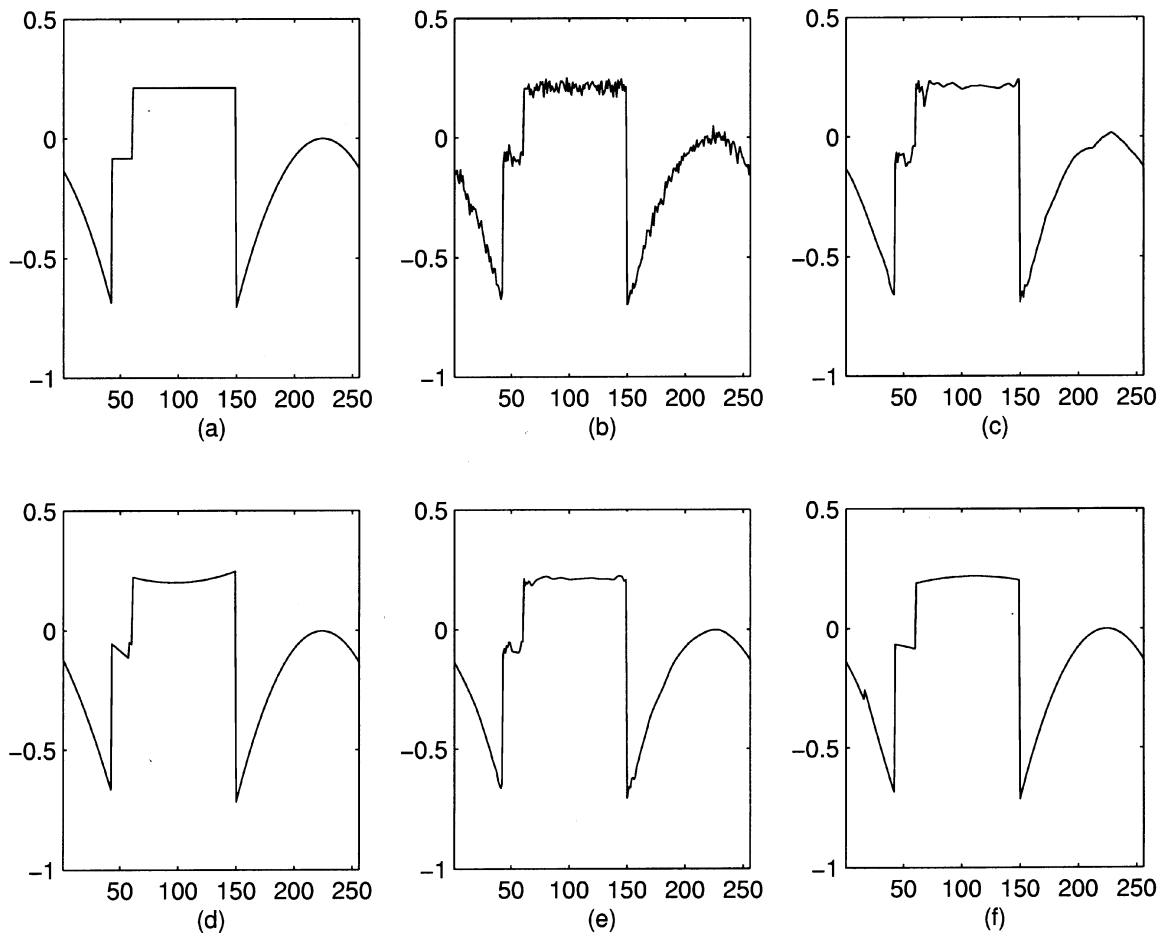
Fig. 2.   SNR results for denoising. (a) Original signal. (b) Noisy signal (22.5 dB). (c) Hard thresholding (25.3 dB). (d) Hard thresholding with footprints (28.5 dB). (e) Cycle spinning with wavelet transform (29.8 dB). (f) Cycle spinning with footprints (30.8 dB).

are orthogonal. Since multiscale operators like footprints are robust to noise, by choosing J_1 as large as possible, we increase this robustness. Second, the signal is reconstructed in terms of the footprint dictionary with $J = \log_2 N$, and this dictionary is unconditional for the class of piecewise constant signals (Proposition 2). Thus, we are sure that the reconstructed signal is still piecewise constant. This is a useful property when the signal to estimate has been corrupted by noise.

The algorithm generalizes to the piecewise polynomial case. The discontinuities are estimated with a $D+1$-order derivative, whereas the coefficients $\alpha_i^{(d)}$ are evaluated with a procedure similar to the one presented above. That is, at each iteration, we choose J such that the footprints related to the two closest discontinuities are orthogonal, we estimate the footprint coefficients of these two discontinuities, and we iterate the process on the residual. Finally, the coefficient α_0^0 is computed using (17). As for the previous case, since $J = \log_2 N$, Proposition 4 guarantees that the reconstructed signal is always piecewise polynomial.

V. APPLICATIONS

In this section, we focus on three main applications for which wavelets are successful, namely denoising, deconvolution, and compression. We present alternative algorithms based on the

footprint expansion and show that these methods can further improve wavelet-based algorithms. The main characteristic of the footprint methods is that they can deal more efficiently with discontinuities.

A. Denoising

The term denoising usually refers to the removal of noise from a corrupted signal. In the typical problem formulation, the original signal x has been corrupted by additive noise. One observes $z[n] = x[n] + e[n]$, where $e[n]$ are independent and identically distributed (i.i.d.) zero mean Gaussian variables with variance σ^2, and the original signal is deterministic and independent of the noise. The goal of the denoising algorithm is to obtain an estimate $\hat{x}$ of the original signal which minimizes a risk function, usually the mean square error $E[\| x - \hat{x} \|^2]$. The wavelet-based denoising algorithm introduced by Donoho and Johnstone [11] simply shrinks the wavelet coefficients. That is, it sets all wavelet coefficients smaller than a threshold to zero and keeps the coefficients above the threshold (hard thresholding) or shrinks them by a fixed amount (soft thresholding). The threshold is usually set to $T = \sigma\sqrt{2 \ln N}$, where N is the size of the signal [11]. A limit of this approach is that it does not exploit the dependency across scales of the wavelet coefficients. Thus, to overcome this limit, we apply a threshold in the footprint domain rather than in the wavelet domain. Doing so, we
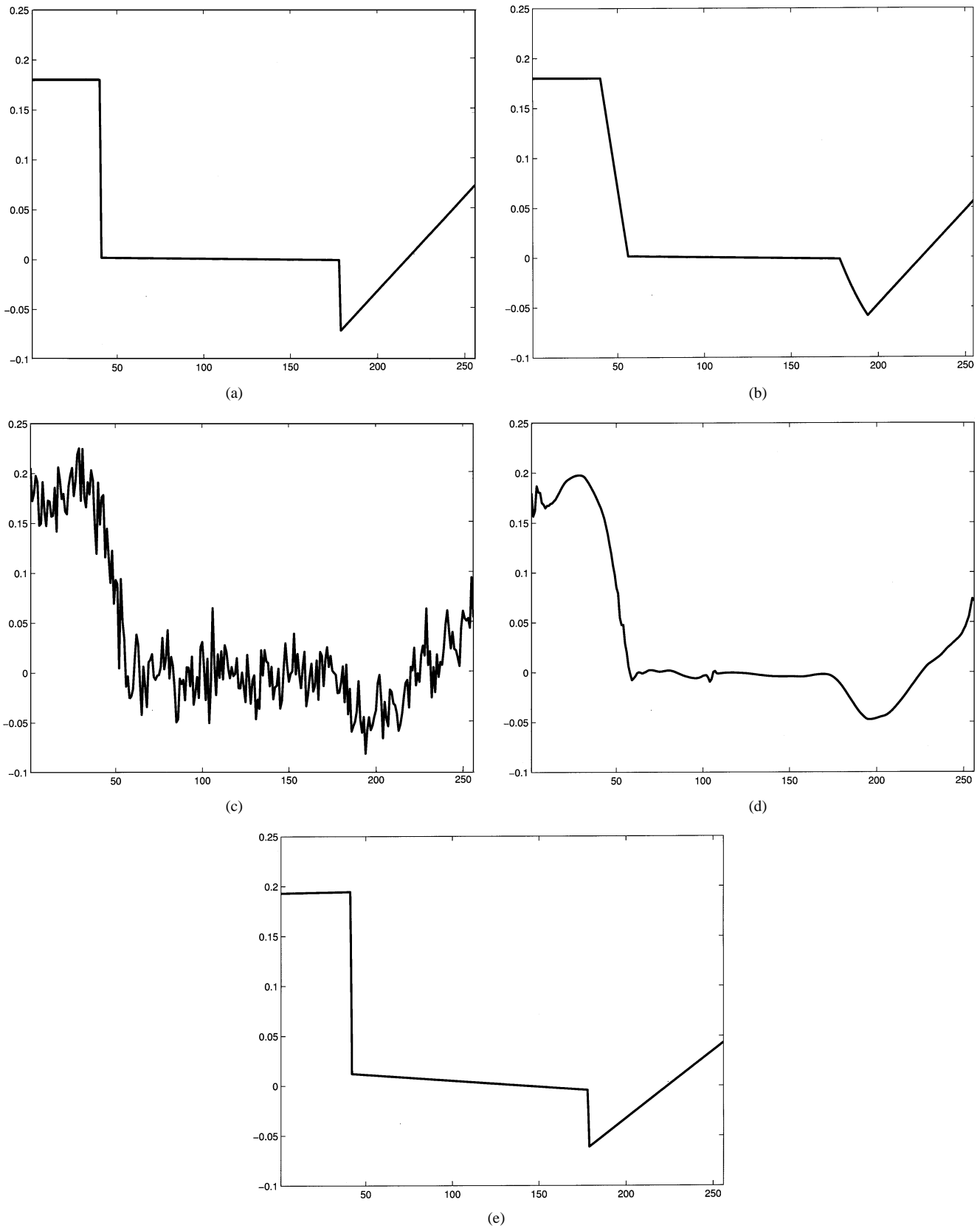
Fig. 3. Deconvolution of a piecewise linear signal. (a) Test signal ($N = 256$). (b) Signal convolved with a box filter. (c) Observed signal.(SNR $= 6.5$ dB) (d) Deconvolution with WaRD (SNR $= 8.8$ dB). (e) Deconvolution with footprints (SNR $= 13.4$ dB).

better exploit the dependency of the wavelet coefficients across scales. As a matter of fact, denoising in the footprint domain is equivalent to applying a vector threshold in the wavelet domain rather than a scalar threshold as in the usual methods.

Assume that $x[n]$ is piecewise polynomial. We can express piecewise polynomial signals in terms of footprints; thus, our denoising system attempts to estimate this footprint representation from the observed noisy version $z[n]$. The estimation procedure follows the same steps as Algorithms IV.1 and IV.2. That is, one first estimates the discontinuity locations and then evaluates the footprint coefficients. Since we only observe a noisy version of the signal, we need to slightly modify these two steps to make them more robust to noise. Again, for simplicity, we focus on piecewise constant signals. The discontinuity locations are estimated in the following way

```
Algorithm V.1: (location Estimation, Noisy
Case)
1) Choose a dictionary
```
$\mathcal{D} = \{f_k^{(0)} = \sum_{j=1}^{j} d_{jk_j}\psi_{jk_j}; k = 0, 1, \ldots, n - 1\}$ of
```
footprints with
```
$j = \log_2 N$. This dictionary
```
represents a biorthogonal basis.
2) Compute the dual basis of
```
$\mathcal{D}$ and call
$\tilde{f}_k^{(0)}$ $k = 1, 2, \ldots, N - 1$ the elements of this
```
dual basis.
3) Compute the
```
$N - 1$ inner products
$\langle z, \tilde{f}_k^{(0)} \rangle$ $k = 1, 2, \ldots, N - 1$.
```
4) Consider as discontinuity locations the
ones related to the inner products larger
than the threshold
```
$T_k = \|\tilde{f}_k^{(0)}\|T$. That is,
```
if
```
$|\langle z, \tilde{f}_k^{(0)}\rangle| \geq T_k$, then assume that there is
```
a discontinuity at location
```
k. T is the
```
universal threshold equal to
```
$\sigma\sqrt{2\ln N}$ [11].

We have a set of estimated discontinuity locations: $\hat{k}_1, \hat{k}_2, \ldots, \hat{k}_{\hat{K}}$. The problem is that due to the noise, this estimation can have errors. Thus, this possibility must be considered in the next step, where the footprints coefficients are evaluated.

```
Algorithm V.2: (coefficient Estimation,
Noisy Case)
1) Call
```
$\hat{\mathcal{K}}$ the set of estimated disconti-
```
nuity locations.
2) Choose
```
$J_1 = \lfloor \log_2(\hat{k}_m - \hat{k}_{m-1}) \rfloor$, where $\hat{k}_{m-1}, \hat{k}_m$
```
are the two closest discontinuity loca-
tions in
```
$\hat{\mathcal{K}}$.
```
3) For each possible location
```
$k \in [\hat{k}_{m-1}, \hat{k}_m]$,
```
compute the inner product
```
$\langle z, (\hat{f}_k^{(0)}/\|\hat{f}_k^{(0)}\|)\rangle$,
```
where
```
$\hat{f}_k^{(0)}$ is the sub-footprint obtained
```
by considering only the first
```
J_1 wavelet
```
coefficients of
```
$f_k^{(0)}$.
```
4) Choose the location
```
k_1 such that
$|\langle z, (\hat{f}_{k_1}^{(0)}/\|\hat{f}_{k_1}^{(0)}\|)\rangle|$ is maximum.
```
5) If

$$|\langle z, \frac{\hat{f}_{k_1}^{(0)}}{\|\hat{f}_{k_1}^{(0)}\|}\rangle| \geq T \qquad (21)$$

```
then compute the residual
```
$R_z^1 = z - 1/\|\hat{f}_{k_1}^{(0)}\|\langle z, \hat{f}_{k_1}^{(0)}/\|\hat{f}_{k_1}^{(0)}\|\rangle\hat{f}_{k_1}^{(0)}$.
```
6) Iterate step 4-5 on the residual until
condition (21) is no longer verified.
7) Once (21) is no longer verified, remove
the two discontinuity locations
```
$\hat{k}_{m-1}, \hat{k}_m$
```
from
```
$\hat{\mathcal{K}}$.
```
8) If
```
$\hat{\mathcal{K}}$ is empty, stop. Otherwise, go to
```
step 2.
```

Finally, the estimated signal $\hat{x}$ is

$$\hat{x} = \langle z, \phi_{J0}\rangle\phi_{J0}[n] + \sum_{m=0}^{M-1}\frac{1}{\|\hat{f}_{k_m}^{(0)}\|}\left\langle R_z^m, \frac{\hat{f}_{k_m}^{(0)}}{\|\hat{f}_{k_m}^{(0)}\|}\right\rangle f_{k_m}^{(0)}[n] \qquad (22)$$

where $M$ is the total number of iterations, $R_z^m$ is the residual after $m$ iterations, and $R_z^0 = z$.

First, notice that since the footprints $f_{k_m}^{(0)}$ in (22) are obtained taking a wavelet transform with $J = \log_2 N$ decomposition levels, we are sure that the estimated signal $\hat{x}$ is piecewise constant like $x$ (Proposition 2). This is an important property because traditional denoising algorithms suffer from the presence of artifacts around discontinuities (pseudo-Gibbs effects). The advantage of denoising in the footprints basis is that these artifacts are automatically eliminated.

Notice that at each iteration, given the two closest discontinuity locations $\hat{k}_{m-1}, \hat{k}_m$, we run a complete matching pursuit algorithm on the interval $[\hat{k}_{m-1}, \hat{k}_m]$ (steps 3–6 of the algorithm). In this way, if there is a discontinuity that has not been detected in the discontinuity estimation step, it can be found in this step. This is the main difference between the noiseless and noisy version of the algorithm.

The proposed denoising algorithm generalizes to piecewise polynomial signals. In this case, given the interval $[\hat{k}_{m-1}, \hat{k}_m]$, instead of running matching pursuit on this interval, we run the subspace pursuit presented in Section IV-A. That is, for each $k \in [\hat{k}_{m-1}, \hat{k}_m]$, we project the set of corresponding noisy wavelet coefficients on the right subspace, we choose the largest projection, and if this projection is larger than the threshold, we keep it. All the other previous considerations apply also to the piecewise polynomial case.

Denoising in the wavelet domain suffers from the lack of shift invariance of the wavelet basis. One way to overcome this limitation is to use a denoising method called cycle-spinning [5]. For a range of shifts, cycle spinning shifts the noisy signal, denoises each shifted version and finally, unshifts and averages the denoised signals. Since footprints suffer from the same lack of shift invariance as wavelets, one can use the idea of cycle spinning to reduce this shift dependency. The only difference between cycle spinning with wavelets and cycle spinning with footprints is that in this second case, each shifted version of the signal is denoised with footprints (Algorithms V.1 and V.2) rather than wavelets. The only limit of this approach is that we can no longer guarantee that the denoised signal is piecewise polynomial. That is, Propositions 2 and 4 do not apply to this case. In Section VI, we consider both methods (denoising with

Fig. 4. Deconvolution of a piecewise smooth signal. (a) Test signal ($N = 256$). (b) Observed signal (SNR $= 16.7$ dB). (c) Piecewise polynomial estimation (SNR $= 21.1$ dB). (d) Residual: $\bar{r} = y - h * \hat{p}$. (e) Deconvolution of the residual with a Wiener filter. (f) Complete deconvolved signal (SNR $= 21.8$ dB).

footprints and cycle spinning with footprints) and compare them with the equivalent wavelet-based algorithms.

### B. Deconvolution

In its simplest form, the deconvolution problem can be stated as follows. The original unknown signal $x[n]$ is blurred by a convolution operator $h[n]$ and corrupted by additive white Gaussian noise. One searches for a good estimate of $x[n]$ from the observed signal

$$y[n] = h[n]*x[n] + e[n]. \qquad (23)$$

Either $h[n]$ is known, or it has to be estimated (blind deconvolution). In most cases, $h[n]$ behaves as a lowpass filter and does not have a bounded inverse, and for this reason, such a deconvolution problem is usually called ill-posed.

There is a large number of methods that provide possible solutions to the deconvolution problem [1], [2], [7], [14], [18], [19], [24], [29]. Among them, wavelet-based methods have become popular recently [1], [2], [14], [18], [24]. This is mostly because these methods deal well with discontinuities and are computationally simple. In our approach, we use the footprint expansion to further improve wavelet-based techniques. We assume that $h[n]$ is known.[12]

Consider the case where $x[n]$ is piecewise polynomial. We know that it can be written as a linear combination of footprints [see (15)]. Thus, by replacing $x[n]$ with its footprints representation, (23) becomes

$$y[n] = \sum_{l=0}^{N/2^J-1} c_l \phi_{Jl}^{per}[n] + \sum_{i=0}^{K} \sum_{d=0}^{D} \alpha_i^{(d)} h[n] * f_{k_i}^{(d)}[n] + e[n]$$

$$= \sum_{l=0}^{N/2^J-1} c_l \phi_{Jl}^{per}[n] + \sum_{i=0}^{K} \sum_{d=0}^{D} \alpha_i^{(d)} \bar{f}_{k_i}^{(d)}[n] + e[n]$$

where, in the last equality, we assumed $\bar{f}_{k_i}^{(d)}[n] = h[n] * f_{k_i}^{(d)}[n]$.[13] That is, $y[n]$ is given by a linear combination of blurred footprints $\bar{f}_{k_i}^{(d)}[n]$ plus the additive white noise $e[n]$. In our deconvolution algorithm, we first attempt to remove the white noise and then the blurring effect. The noise is removed using the denoising Algorithms V.1 and V.2, but we use the blurred dictionary $\bar{\mathcal{D}}$ to perform denoising rather than $\mathcal{D}$. The deblurring process then simply consists in replacing the $\bar{f}_{k_i}^{(d)}$'s with the corresponding nonblurred footprints. The complete algorithm can be summarized as follows.

```
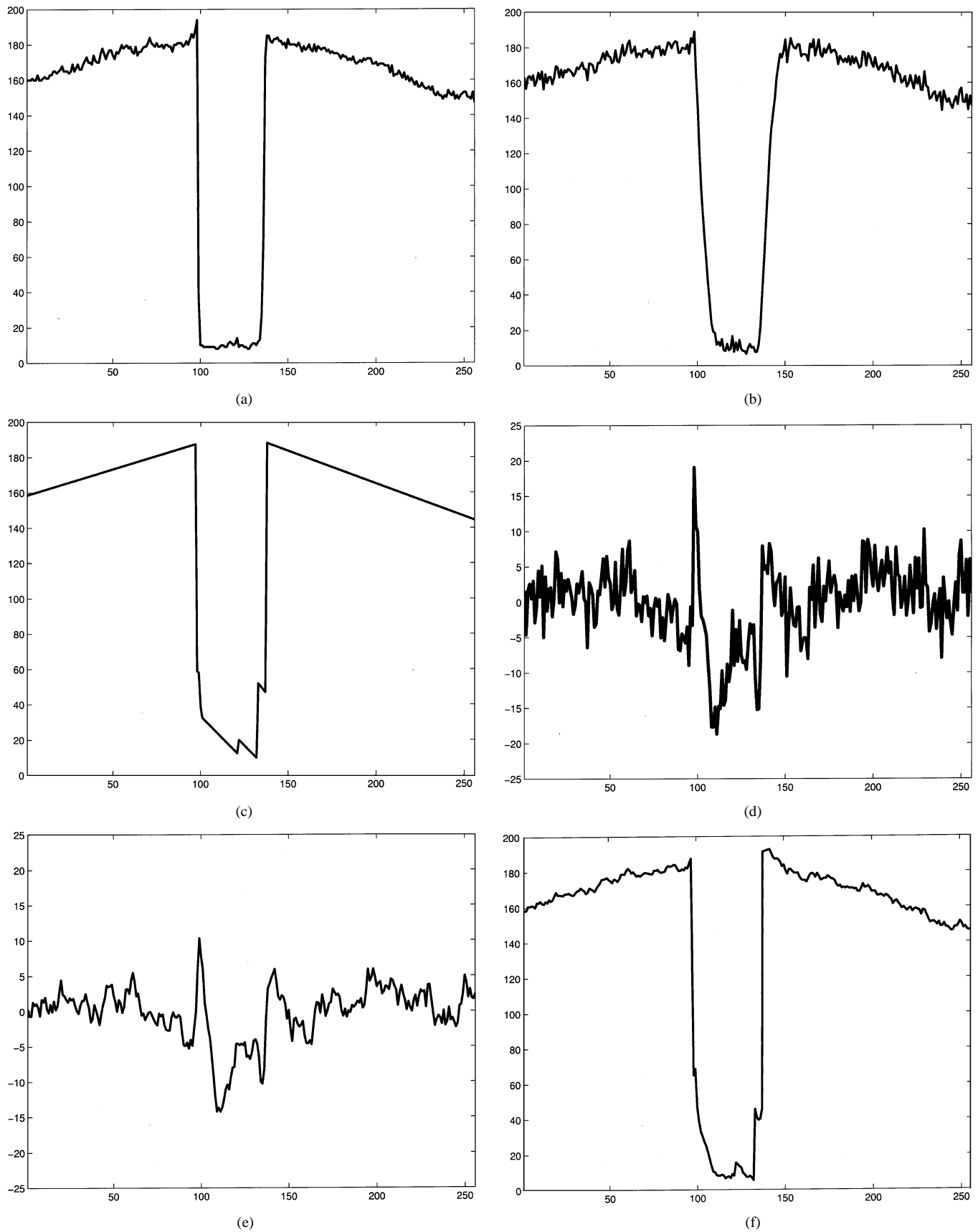Algorithm V.3: (Deconvolution of Piecewise
Polynomial Signals)
1) Consider the Dictionary of Blurred
Footprints D̄ = {f̄_k^(d); k = 0,1,...,n − 1; D =
0,1,...,D}. Remove the noise in y[n] using
Algorithms V.1 and V.2 and assuming D̄ as
the reference footprints dictionary.
```

[12]It is worth pointing out that in some cases, footprints can be used for blind deconvolution as well [15].

[13]In practice, the convolution filter has little effect on the low-frequency scaling coefficients. Therefore, in our formulation, we do not consider this effect and use the original scaling functions.

```
2) Call the denoised signal
ŷ[n] = ∑_{l=0}^{N/2^J-1} ĉ_l φ_{Jl}^{per}[n] + ∑_{i=0}^{K̂} ∑_{d=0}^{D} α̂_i^(d) f̄_{k̂_i}^(d)[n].
The deconvolved signal x̂[n] is given by
x̂[n] = ∑_{l=0}^{N/2^J-1} ĉ_l φ_{Jl}^{per}[n] + ∑_{i=0}^{K̂} ∑_{d=0}^{D} α̂_i^(d) f_{k̂_i}^(d)[n],
where we have simply replaced f̄_{k̂_i}^(d)[n] with
f_{k̂_i}^(d)[n].
```

If $x[n]$ is piecewise smooth, we use a *two-step* deconvolution algorithm. The procedure of this algorithm is based on the result of Theorem 1, which says that $x[n]$ can be written as the sum of a piecewise polynomial signal $p[n]$ and a regular function $r[n]$. That is, $x[n] = p[n] + r[n]$. Therefore, the observed signal $y[n]$ can be written as $y[n] = h[n]*p[n] + h[n]*r[n] + e[n]$. The aim of the algorithm is to estimate the two contributions $p[n]$ and $r[n]$ in two different phases. The complete algorithm operates in the following way (we assume that $h[n]$ is known).

```
Algorithm V.4: (Two-Step Deconvolution)
1) Estimate the piecewise polynomial be-
havior underlying y[n] with the deconvo-
lution Algorithm V.3. Call the estimated
signal p̂[n].
2) Compute the residual r̃[n] = y[n] − h[n]*p̂[n].
3) Deconvolve the residual with a Wiener
filter g[n]: r̂[n] = g[n]*r̃[n]
4) The estimated signal is x̂[n] = p̂[n] + r̂[n].
```

C. Compression

Wavelets are widely used in compression. The reason is that wavelets have very good approximation properties for representing certain classes of signals like piecewise smooth signals. While good approximation properties are necessary for good compression, it might not be enough. In compression, one has to consider the costs corresponding to indexing, and compressing the retained elements in the approximation and independent coding of these coefficients might be inefficient [31].

Consider a piecewise smooth signal defined as in (2), that is, a function with pieces that are α-Lipschitz regular and with a finite number of discontinuities. It was shown in [4] that standard wavelet-based schemes such as zerotrees [28] can achieve the following distortion-rate performance:

$$D(R) \le c_1 R_s^{-2\alpha} + c_2 \sqrt{R_e} 2^{-c_3 \sqrt{R_e}} \qquad (24)$$

where $R = R_s + R_e$ and R_e are the bits used to quantize the wavelet coefficients generated by the discontinuities, whereas R_s are the bits used to code the wavelet coefficients corresponding to the smooth parts of the signal. Now, suppose that the signal is piecewise polynomial. Then, the wavelet coefficients related to the smooth parts of the signal are exactly zero; therefore, there is no need to use any bits to code them. The distortion of a wavelet-based scheme becomes

$$D(R) \le c_2 \sqrt{R} 2^{-c_3 \sqrt{R}}. \qquad (25)$$

However, a direct approach to compression of piecewise polynomial signals, based on an oracle telling us where discontinuities are, will lead to $D(R) \le c_4 2^{-c_5 R}$ [25], and such behavior

is achievable using dynamic programming [25]. This large gap between ideal performance given by the scheme based on dynamic programming and wavelet performance is mainly due to the independent coding of the wavelet coefficients across scales. Statistical modeling [9] of such dependencies can improve the constants in (25), but going from $\sqrt{R}$ to R in the exponent requires taking the deterministic behavior of wavelet coefficients across scales at singularities into account. This is well done using footprints, which thus close the gap with the ideal performance.

Theorem 3: Consider piecewise polynomial signals with polynomials of maximum degree D and no more than K discontinuities. A coder, which represents these signals in the footprints basis and which scalar quantizes the discontinuity locations and the footprint coefficients, achieves

$$D(R) \leq c_6 2^{-c_7 R}. \tag{26}$$

Proof: See Appendix C. □

Thus, this theorem shows that in case of piecewise polynomial signals, footprints significantly improve performance of wavelet coders. Footprints can be used for piecewise smooth signals as well. Theorem 1 shows that a piecewise smooth signal can be separated into two contributions: a piecewise polynomial part (call it $p[n]$) and a residual $r[n]$, which is regular (α-Lipschitz over $\mathbb{R}$). Now, $p[n]$ can be compressed with footprints, and this coder achieves (26). The residual $r[n]$ can be compressed with any other coder that achieves [4]

$$D(R) \leq c_8 R^{-2\alpha}. \tag{27}$$

It is worth noticing that because of the regularity of $r[n]$, the performance in (27) can be achieved with a simple coder based on linear approximation of $r[n]$ in a wavelet or Fourier basis [4]. Combining (26) and (27) shows that a two-stage compression algorithm based on footprints and on linear approximation of the residual achieves

$$D(R) \leq c_8 R_s^{-2\alpha} + c_6 2^{-c_7 R_e}. \tag{28}$$

Comparing (24) and (28), we can see that this coder does not change the asymptotics of the distortion-rate function of wavelet coders ($\sim c_8 R_s^{-2\alpha}$). However, by coding the discontinuities efficiently, this coder reaches the asymptotic behavior more rapidly. Finally, notice that for this last performance, the underlying assumption is that the encoder knows in advance the signal to code, and this way, it can separate the polynomial and the smooth parts of the signal. In the experimental results, we will show that a realistic encoder can obtain similar performance without knowing the signal characteristics in advance.

VI. NUMERICAL EXPERIMENTS

In this section, we compare footprints with wavelet-based methods on several examples. Our purpose is to show that footprints are a versatile tool and that we can get good results in a variety of applications.

Fig. 5. Theoretical and experimental D/R curves. Dashed-dotted: theoretical wavelet performance. Dashed: empirical footprint performance. Line: ideal performance.

A. Denoising

For denoising, we consider only piecewise polynomial signals. In Table II, we compare the performance of our denoising systems with a classical hard thresholding algorithm [11] and cycle-spinning [5]. In this experiment, we consider piecewise linear signals with no more than three discontinuities. The performance is analyzed in function of the size N of the signal. The table clearly shows that denoising with footprints outperforms the hard thresholding system, whereas cycle-spinning with footprints outperforms traditional cycle-spinning. In Fig. 2, we show an example of the denoising algorithm on piecewise quadratic signals. We can see that signals denoised with footprints present better visual quality since they do not suffer from pseudo-Gibbs effects.

B. Deconvolution

In this case, we consider two different signals. One is a piecewise linear signal, and the other one is a line of the image "Cameraman," which represents a possible example of piecewise smooth signals. We first consider the case of a piecewise linear signal and compare the performance of our system with WaRD [24]. In this simulation, the original signal is first convolved with a box filter, and then, white noise is added. The noise variance is set to $\sigma^2 = 0.02$. Fig. 3 shows that our system outperforms WaRD in both visual quality and SNR. It is of interest to note that the signal reconstructed with footprints does not present artifacts around discontinuities and that it manages to efficiently sharpen the discontinuities. Of course, one of the reasons why footprints perform so well is because the considered signal perfectly fits the model since it is piecewise polynomial.

In Fig. 4, we consider the case where the signal is piecewise smooth. Again, the original signal is convolved with a box filter, and then, white noise is added. In this case, we use the *two-step deconvolution* algorithm. The estimated piecewise polynomial behavior $\hat{p}[n]$ underlying the signal is shown in Fig. 4(c). The estimated residual $\tilde{r}[n]$ and the deconvolved

Fig. 6.   Compression of a piecewise smooth function.

residual $\hat{r}[n]$ are shown in Fig. 4(d) and (e), respectively. Finally, the reconstructed signal is shown in Fig. 4(f).

C. Compression

In Theorem 3, we have shown that in case of piecewise polynomial signals, a footprint-based coder can achieve the ideal rate-distortion performance. That is, it has the correct rate of decay of the R-D function. Now, we are interested in a numerical confirmation of this theorem. We consider piecewise constant signals with no more than five discontinuities. The signal has size $N = 2^{10}$, and the discontinuity locations are uniformly distributed over the interval $[0, N - 1]$. The footprint coder operates as in Theorem 3, that is, it scalar quantizes the footprint coefficients and the discontinuity locations. Bits are allocated with a reverse waterfilling strategy. In Fig. 5, we compare the rate-distortion performance of this footprint coder against the ideal bound and the ideal performance of a wavelet-based coder. We can see that the behavior of the footprint coder is consistent with the theory since it has the same rate of decay as the ideal distortion function.

Finally, we consider a piecewise smooth signal. The compression operates in the following way. With a denoising-like algorithm, we estimate the piecewise polynomial behavior underlying the signal and compress it with footprints. The residual is assumed regular, and it is compressed in a wavelet basis. That is, the first k coefficients of the wavelet decomposition are quantized, whereas the others are set to zero (linear approximation).

The allocation of the bits between the piecewise polynomial signal and the residual and the number k of wavelet coefficients that are quantized is chosen off-line, using some *a priori* knowledge of the signal. In Fig. 6, we show an example of the performance of the proposed compression scheme and compare it with a 1-D version of SPIHT [27]. The signal to compress is given by the union of smooth pieces. In this example, our system outperforms SPIHT by more than 4 dB. Since SPIHT is more suited to compress 2-D signals, this comparison is only indicative. However, it shows that a compression system based on footprints can outperform traditional wavelet methods also in the case of piecewise smooth signals.

VII. CONCLUSIONS

In this paper, we have presented a new way of modeling the dependency across scales of wavelet coefficients with elements we called footprints. Footprints form an overcomplete dictionary and are efficient at representing the singular structures of a signal. With footprints, it is possible to get a sparser representation of piecewise smooth signals than with wavelet bases, and this is useful in several signal processing tasks. Numerical simulations confirm that footprints outperform wavelet methods in several applications. In short, wavelets have been very successful on signals with discontinuities, be it for denoising, deconvolution, or compression. Wavelet footprints pursue this program further by explicitly using the structure of discontinuities across scales. The results, both theoretical and experimental,

confirm the potential of this approach. Together with the simplicity of the algorithms involved, this indicates the power of this new data structure.

APPENDIX A

A. Proof of Theorem 1

Consider first, a piecewise smooth signal $f(t)$ $t \in [0, T[$ with only two pieces. That is $f(t) = f_1(t)\mathbf{1}_{[0,t_1[} + f_2(t)\mathbf{1}_{[t_1,T[}$ and $f_1(t), f_2(t)$ are uniformly Lipschitz α over $[0, t_1]$ and $[t_1, T]$, respectively. Recall that if a function f is uniformly Lipschitz $\alpha > d$ in the neighborhood of ν, then it is necessarily d times continuously differentiable in that neighborhood.[14] Moreover, the polynomial $p_\nu(t)$ in (1) is the Taylor expansion of f at ν. Now, since $f_1(t), f_2(t)$ are uniformly Lipschitz α over $[0, t_1]$ and $[t_1, T]$, they are necessarily $d = \lfloor \alpha \rfloor$ times continuously differentiable on these intervals. Call $p(t) = p_{t_1^-}(t)\mathbf{1}_{[0,t_1[} + p_{t_1^+}(t)\mathbf{1}_{[t_1,T[}, t \in [0, T[$ the piecewise polynomial signal whose two pieces $p_{t_1^-}(t)$ and $p_{t_1^+}(t)$ are given by $p_{t_1^-}(t) = f_1(t_1) + f_1'(t_1)(t-t_1) + \cdots + f_1^{(d)}(t_1)/d!(t-t_1)^d$ and $p_{t_1^+}(t) = f_2(t_1) + f_2'(t_1)(t-t_1) + \cdots + f_2^{(d)}(t_1)/d!(t-t_1)^d$. That is, $p_{t_1^-}(t)$ and $p_{t_1^+}(t)$ are the Taylor expansions of $f(t)$ about t_1 taken from the left and from the right of t_1. Now, the signal $r_\alpha(t) = f(t) - p(t)$ is d times continuously differentiable in $[0, T[- \{t_1\}$, and in t_1, it verifies

$$\lim_{t \to t_1^-} r_\alpha(t) = \lim_{t \to t_1} f_1(t) - p_{t_1^-}(t)$$
$$= \lim_{t \to t_1^+} r_\alpha(t)$$
$$= \lim_{t \to t_1} f_2(t) - p_{t_1^+}(t) = 0$$
$$\lim_{t \to t_1^-} r_\alpha^{(l)}(t) = \lim_{t \to t_1} f_1^{(l)}(t) - p_{t_1^-}^{(l)}(t)$$
$$= \lim_{t \to t_1^+} r_\alpha^{(l)}(t)$$
$$= \lim_{t \to t_1} f_2^{(l)}(t) - p_{t_1^+}^{(l)}(t) = 0, l = 1, 2, \ldots, d.$$

Therefore, $r_\alpha(t)$ is d times continuously differentiable on the entire interval $[0, T]$. Therefore, it is uniformly Lipschitz $\alpha' > d$ on that interval. The remaining step is to prove that $\alpha' = \alpha$. This is clearly true for all points away from t_1, and we only need to prove that $r_\alpha(t)$ is α-Lipschitz in t_1. Using the definition of Lipschitz regularity, we have that $|f(t) - p_{t_1^-}(t)| \le K_1|t - t_1|^\alpha$, for $t < t_1$ and $|f(t) - p_{t_1^+}(t)| \le K_2|t - t_1|^\alpha$, for $t \ge t_1$. Now, since $r_\alpha(t) = f(t) - p(t)$, we can write $|r_\alpha(t)| \le K_1|t - t_1|^\alpha$, for $t < t_1$ and $|r_\alpha(t)| \le K_2|t - t_1|^\alpha$ for $t \ge t_1$. Thus, if we call $K = \max\{K_1, K_2\}$, then in the neighborhood of t_1, we have $|r_\alpha(t)| \le K|t - t_1|^\alpha$, which proves that $r_\alpha(t)$ is Lipschitz α in t_1. This completes the proof.

The generalization of this result to the case of a piecewise smooth signal $f(t)$ with K discontinuities at locations $t_1, t_2, \ldots, t_K$ is straightforward. Call $p_i(t) = p_{t_i^-}(t)\mathbf{1}_{[0,t_i[} + p_{t_i^+}(t)\mathbf{1}_{[t_i,T[}, i = 1, 2, \ldots, K$; the piecewise polynomial signal with two polynomials $p_{t_i^-}(t)$ and

[14]The converse is also true. That is, a function that is d times continuously differentiable in the neighborhood of ν is Lipschitz $\alpha' > d$ at ν.

$p_{t_i^+}(t)$ and only one discontinuity at location t_i. $p_{t_i^-}(t)$ and $p_{t_i^+}(t)$ are the Taylor expansions of $f(t)$ about t_i taken from the left and from the right of t_i. Then, the piecewise polynomial signal $p(t) = \sum_{i=1}^K p_i(t)$ is such that $r_\alpha(t) = f(t) - p(t)$ is d times continuously differentiable in $t_1, t_2, \ldots, t_K$. Therefore, $r_\alpha(t)$ is d times continuously differentiable on $[0, T]$ and uniformly Lipschitz $\alpha' > d$ on that interval. Finally, as in the previous example, one can show that $\alpha' = \alpha$.

APPENDIX B

A. Proof of Proposition 1

We are considering signals in $\mathbb{R}^N$ and the union of footprints and scaling functions gives N elements. We need to show that this set of elements is complete. This is equivalent to showing that there exists no $x[n]$ with $\|x\| > 0$ such that it has a zero expansion, that is, such that

$$\sum_k \left| \left\langle x, f_k^{(0)} \right\rangle \right|^2 + \sum_l |\langle x, \phi_{Jl} \rangle|^2 = 0. \quad (29)$$

We prove this for the case $J = \log_2 N$, noting that with the same method, one can prove it for any J. Consider the representation of x in terms of the wavelet basis: $x[n] = c_0\phi_{J0}[n] + \sum_{j=1}^J \sum_{l=0}^{N/2^j - 1} y_{jl}\psi_{jl}[n]$. Equation (29) already implies that the scaling coefficient $c_0 = \langle x, \phi_{J0} \rangle$ is zero. We will show that if (29) is true, then also all the wavelet coefficients of x are zero, and therefore, it must be $\|x\| = 0$. Recall that since $J = \log_2 N$, there is only one wavelet coefficient at level J, two wavelet coefficients at level $J - 1$, and so on. First, consider the footprint $f_k^{(0)}[n] = \sum_{j=1}^J d_{jk_j}\psi_{jk_j}[n]$ with $k = 2^J/2$ and the corresponding inner product $\langle x, f_k^{(0)} \rangle$. One can easily verify that the only nonzero coefficient d_{jk_j} of $f_k^{(0)}[n]$ is the one at scale J. That is, $f_k^{(0)}[n] = d_{Jk_J}\psi_{Jk_J}[n] = d_{J0}\psi_{J0}[n]$, where in the last equality, we have used the fact that $k_J = 0$. Thus, we have that $\langle x, f_k^{(0)} \rangle = y_{J0}d_{J0}$, and this inner product is equal to zero only if $y_{J0} = 0$. Consider now the footprint $f_k^{(0)}$ with $k = 2^J/4$. In this case, $f_k^{(0)} = d_{J0}\psi_{J0}[n] + d_{(J-1)0}\psi_{(J-1)0}[n]$, that is, $f_k^{(0)}$ has only two nonzero coefficients d_{jk_J} at scales J and $J - 1$. Therefore, we have that $\langle x, f_k^{(0)} \rangle = y_{J0}d_{J0} + y_{(J-1)0}d_{(J-1)0}$. Since we have seen that $y_{J0} = 0$, this second inner product is zero only if $y_{(J-1)0} = 0$. In the same way, but with the footprint related to position $2^J/4 + 2^J/2$, we can prove that $y_{(J-1)1} = 0$. Therefore, the wavelet coefficients at scales J and $J-1$ are zero. The same analysis can be repeated at each scale, and in conclusion, we have that (29) implies that all the wavelet coefficients of x are zero. Therefore, x must be the zero vector.

APPENDIX C

A. Proof of Theorem 3

Consider a piecewise polynomial signal $x[n] \in [0, N-1]$ of maximum degree D and with no more than K discontinuities. Assume that the signal is bounded in magnitude between $[-A, A]$. We want to prove that a compression scheme based on footprints can achieve

$$D(R) \le c_6 2^{-(2R/(D+3)K+(D+1))} \sim c_6 2^{-c_7 R}.$$

Consider the representation of x in terms of footprints for the case $J = \log_2 N$:

$$x[n] = c_0 \phi_{J0}^{per}[n] + \sum_{i=0}^{K} \sum_{d=0}^{D} \alpha_i^{(d)} f_{k_i}^{(d)}[n]. \qquad (30)$$

The compression algorithm consists in uniform scalar quantizing the discontinuity locations k_i and the footprints coefficients $\alpha_i^{(d)}$. Since x is bounded, the square error relative to the quantization of a single discontinuity location can be upper bounded by $\|x - \hat{x}\|^2 \leq 4A^2|k_i - \hat{k}_i|$, where $\hat{x}$ is the approximated signal. If R_{k_i} bits are used to quantize each discontinuity then $|k_i - \hat{k}_i| \leq (N/2)2^{-R_{k_i}}$ and the distortion related to a single discontinuity is $D_l(R_{k_i}) \leq 2A^2N2^{-R_{k_i}}$. Consider, now, the quantization of the coefficients of the footprints expansion.[15] Since $\|f_k^{(d)}\| = 1$, the square error due to the quantization of a single coefficient $\alpha_i^{(d)}$ is $\|x - \hat{x}\|^2 = (\alpha_i^{(d)} - \hat{\alpha}_i^{(d)})^2$. Now, x is bounded, and therefore, each coefficient $\alpha_i^{(d)}$ is bounded as well: $\alpha_i^{(d)} \in [-B_i^{(d)}, B_i^{(d)}]$. Thus, if $R_i^{(d)}$ bits are allocated to quantize $\alpha_i^{(d)}$, then the distortion due to this quantization can be upperbounded by $D_\alpha(R_i^{(d)}) \leq B^2 2^{-2R_i^{(d)}}$, where $B = \max_{i,d} B_i^{(d)}$. The global distortion bound is obtained by adding all the distortion contributions: $D(R) \leq \sum_{i=1}^{K} D_l(R_{k_i}) + \sum_{j=0}^{K}\sum_{d=0}^{D} D_\alpha(R_i^{(d)})$, where $R = \sum_{i=1}^{K} R_{k_i} + \sum_{i=0}^{K}\sum_{d=0}^{D} R_i^{(d)}$. Finally, by allocating bits over the different distortions with a reverse waterfilling scheme [8], the global distortion becomes

$$D(R) \leq c_6 2^{-(2R/(D+3)K+(D+1))} \sim c_6 2^{-c_7 R}.$$

ACKNOWLEDGMENT

REFERENCES

[1] M. R. Banham and A. K. Katsaggelos, "Spatially adaptive wavelet-based multiscale image restoration," *IEEE Trans. Image Processing*, vol. 5, pp. 619–634, Apr. 1996.

[2] K. Berkner, M. J. Gormish, and E. L. Schwartz, "Multiscale sharpening and smoothing in Besov spaces with applications to image enhancement," *Appl. Comput. Harmon. Anal.*, vol. 11, pp. 2–31, July 2001.

[3] S. Chen and L. Donoho, "Atomic decomposition by basis pursuit," in *Proc. SPIE Int. Conf. Wavelets*, San Diego, CA, July 1995.

[4] A. Cohen, I. Daubechies, O. Guleryuz, and M.T. Orchard, "On the importance of combining wavelet-based nonlinear approximation with coding strategies," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1895–1921, July 2002.

[5] R. R. Coifman and L. Donoho, "Translation Invariant Denoising," Dept. Statist., Stanford Univ., Tech. Rep. 475, 1995.

[6] R. Coifman and M. Wickerhauser, "Entropy-based algorithms for best basis selection," *IEEE Trans. Inform. Theory*, vol. 38, pp. 713–718, Mar. 1992.

[7] P. L. Combettes, "Generalized convex set theoretic image recovery," in *Proc. IEEE Int. Conf. Image Process.*, Lausanne, Switzerland, Sept. 1996.

[8] T. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[9] M. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based signal processing using hidden Markov models," *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, Apr. 1998.

[10] G. Davis, "Adaptive nonlinear approximations," Ph.D. dissertation, New York Univ., New York, NY, 1994.

[11] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, vol. 81, pp. 425–455, Dec. 1994.

[12] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM J. Appl. Math.*, vol. 49/3, pp. 906–931, June 1989.

[13] D. L. Donoho, M. Vetterli, R. A. DeVore, and I. Daubechies, "Data compression and harmonic analysis," *IEEE Trans. Informat. Theory*, vol. 44, pp. 2435–2476, Oct. 1998.

[14] D. L. Donoho, "Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition," *Applied Computat. Harmon. Anal.*, vol. 2, pp. 101–126, Apr. 1995.

[15] P. L. Dragotti, "Wavelet Footprints and Frames for Signal Processing and Communications," Ph.D. dissertation, Swiss Federal Inst. Technol., Lausanne, Switzerland, 2002.

[16] M. Elad and A. M. Bruckstein, "On sparse signal representations," in IEEE Int. Conf. Image Processing, Thessaloniki, Greece, Oct. 2001.

[17] M. M. Goodwin and M. Vetterli, "Matching pursuit and atomic signal models based on recursive filter banks," *IEEE Trans. Signal Processing*, vol. 47, pp. 1890–1902, July 1999.

[18] J. Kalifa, S. Mallat, and B. Rouge, "Image deconvolution in mirror wavelet bases," in IEEE Int. Conf. Image Process., Chicago, IL, Oct. 1998.

[19] A. K. Katsaggelos, "Iterative image restoration algorithm," *Opt. Eng.*, vol. 28, pp. 735–748, 1989.

[20] S. Mallat and W. L. Hwang, "Singularity detection and processing with waveletes," *IEEE Trans. Inform. Theory*, vol. 38, pp. 617–643, Mar. 1992.

[21] S. Mallat and Z. Zhang, "Matching pursuit with time—frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.

[22] S. Mallat and S. Zhong, "Characterization of signals from multiscale edges," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 710–732, July 1992.

[23] S. Mallat, *A Wavelet Tour of Signal Processing*. New York: Academic, 1998.

[24] R. Neelamani, H. Choi, and R. Baraniuk, "Wavelet-based deconvolution for ill-conditioned systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999.

[25] P. Prandoni and M. Vetterli, "Approximation and compression of piecewise smooth functions," *Phil. Trans. Royal Soc. London*, Aug. 1999.

[26] K. Ramchandram and M. Vetterli, "Best wavelet packet bases in a rate-distortion sense," *IEEE Trans. Image Processing*, vol. 2, pp. 160–175, Apr. 1993.

[27] A. Said and W. A. Pearlman, "A new, fast and efficient image codec based on set partitioning in hierarchical trees," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 6, no. 3, pp. 243–249, June 1996.

[28] J. M. Shapiro, "Embedded image coding using zerotrees of wavelets coefficients," *IEEE Trans. Signal Processing*, vol. 41, pp. 3445–3462, Dec. 1993.

[29] H. Stark, *Image Recovery: Theory and Application*. New York: Academic, 1995.

[30] M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*. Englewood Cliffs, NJ: Prentice-Hall, 1995.

[31] M. Vetterli, "Wavelets approximation and compression," *IEEE Signal Processing Mag.*, vol. 18, pp. 59–73, Sept. 2001.

[15]The scaling coefficient c_0 is included in this formulation.

Pier Luigi Dragotti (M'02) received the Laurea degree (summa cum laude) in electrical engineering from the University Federico II, Naples, Italy, in 1997, the Master degree in communications systems from the Swiss Federal Institute of Technology (EPFL), Lausanne, Switzerland, in 1998, and the Ph.D. degree from EPFL in April 2002.

From March to September 1996, he was a visiting student at Stanford University, Stanford, CA, and from July to October 2000, he was an intern at Lucent Technologies Bell Laboratories, Holmdel, NJ. From September 1998 to October 2002, he was a Research Assistant at the Audiovisual Communications Laboratory, EPFL. Since November 2002, he has been a Lecturer with the Electrical and Electronic Engineering Department, Imperial College, London, U.K. His research interests include joint source-channel coding, wavelets, data compression, and multirate signal processing.

Martin Vetterli (F'99) received the Dipl. El.-Ing. degree from ETH Zurich (ETHZ), Zurich, Switzerland, in 1981, the M.S. degree from Stanford University, Stanford, CA, in 1982, and the Dr.-Sci. degree from EPF Lausanne (EPFL), Lausanne, Switzerland, in 1986.

He was a Research Assistant at Stanford and EPFL and has worked for Siemens and AT&T Bell Laboratories. In 1986, he joined Columbia University, New York, NY, where he was last an Associate Professor of electrical engineering and co-director of the Image and Advanced Television Laboratory. In 1993, he joined the University of California at Berkeley, where he was a Professor with the Department of Electrical Engineering and Computer Science until 1997, where he now holds an Adjunct Professor position. Since 1995, he has been a Professor of communication systems at EPFL, where he chaired the Communications Systems Division from 1996 to 1997 and heads the Audiovisual Communications Laboratory. Since 2001, he has directed the National Competence Center in Research on mobile information and communication systems. He held visiting positions at ETHZ in 1990 and at Stanford in 1998. He is on the editorial boards of *Annals of Telecommunications*, *Applied and Computational Harmonic Analysis*, and *The Journal of Fourier Analysis and Application*. He is the coauthor, with J. Kovacevic, of the book *Wavelets and Subband Coding* (Englewood Cliffs, NJ: Prentice-Hall, 1995). He has published about 85 journal papers on a variety of topics in signal/image processing and communications and holds seven patents. His research interests include sampling, wavelets, multirate signal processing, computational complexity, signal processing for communications, digital video processing, and joint source/channel coding.

Dr. Vetterli a member of SIAM and was the Area Editor for Speech, Image, Video and Signal Processing of the IEEE TRANSACTIONS ON COMMUNICATIONS. He received the Best Paper Award from EURASIP in 1984 for his paper on multidimensional subband coding, the Research Prize of the Brown Bovery Corporation (Switzerland) in 1986, for his doctoral thesis, and the IEEE Signal Processing Society's Senior Awards in 1991 and in 1996 (for papers with D. LeGall and K. Ramchandran, respectively). He won the Swiss National Latsis Prize in 1996, the SPIE Presidential award in 1999, and the IEEE Signal Processing Technical Achievement Award in 2001. He is a member of the Swiss Council on Science and Technology. He was a plenary speaker at various conferences (e.g., 1992 IEEE ICASSP).