Applied and Computational Harmonic Analysis **9**, 166–203 (2000)
doi:10.1006/acha.2000.0315, available online at http://www.idealibrary.com on IDEAL®

# Smooth Multiwavelet Duals of Alpert Bases by Moment-Interpolating Refinement

David L. Donoho

*Department of Statistics, Stanford University, Stanford, California*

Nira Dyn and David Levin

*Department of Mathematics, Tel Aviv University, Tel Aviv, Israel*

and

Thomas P. Y. Yu

*Program of Scientific Computing and Computational Mathematics, Stanford University, Stanford, California*

*Communicated by Jianzhong Wang*

Received September 2, 1998; revised November 16, 1999

Using refinement subdivision techniques, we construct smooth multiwavelet bases for $L^2(\mathbb{R})$ and $L^2([0,1])$ which are in an appropriate sense dual to Alpert orthonormal multiwavelets. Our new multiwavelets allow one to easily give smooth reconstructions of a function purely from knowledge of its local moments. At the heart of our construction is the concept of *moment-interpolating* (MI) refinement schemes, which interpolate sequences from coarse scales to finer scales while preserving the underlying local moments on dyadic intervals. We show that MI schemes have smooth refinement limits. Our proof technique exhibits an intimate *intertwining relation* between MI schemes and Hermite schemes. This intertwining relation is then used to infer knowledge about moment-interpolating schemes from knowledge about Hermite schemes. Our MI multiwavelets make Riesz bases for $L^2$ and unconditional bases of a variety of smoothness spaces, so they can efficiently represent smooth functions. We here derive an algorithm which rapidly develops a piecewise polynomial fit to data by recursive dyadic partitioning and then rapidly produces a smooth reconstruction with matching local moments on pieces of the partition. This avoids the blocking effect suffered by piecewise polynomial fitting.



## 1. INTRODUCTION

Alpert [2] and Alpert *et al.* [1] introduced a family of hierarchical orthogonal bases and a class of fast algorithms relying on those. This family of bases is indexed by a parameter



1063-5203/00 $35.00

$m \geq 1$, the *multiplicity*, and the case $m = 1$ is just the usual Haar basis. The bases with $m > 1$ are built from Legendre polynomials of degree $< m$ localized to dyadic intervals. Typical members of these bases have $m$ vanishing moments and are discontinuous.

In this paper we describe a family of smooth multiwavelet bases which are in an appropriate sense dual to Alpert bases. These new bases have the property that their expansion coefficients are linear combinations of a few Alpert moments and the basis functions are smooth. They could be useful in settings where the fast algorithms of Alpert *et al*. [1] had been already programmed, and so one had available a convenient algorithm for rapidly obtaining the Alpert coefficients of an approximate solution, but one desired a smoother reconstruction than the one provided by the Alpert expansions.

### *1.1. Relation to Refinement Schemes*

The smooth multiwavelets arise from the use of a refinement scheme we call *moment interpolation*. This is a generalization of average interpolation (Donoho [11], Harten [19]) to a vector of higher-order moments. It is related to Dubuc–Deslauriers refinement and Hermite refinement as described in this table:

| | Point values and derivatives | Average values and moments |
|---|---|---|
| Scalar | Dubuc–Deslauriers | Average-interpolating |
| Vector | Hermite | Moment-interpolating |

There is a precise mathematical connection between moment-interpolating refinement and Hermite refinement which is described in Section 3 below. Roughly speaking, moment-interpolating refinement of a sequence of local moments gives a result which is the $m$th derivative of Hermite refinement of a specially related sequence obtained by integrating local moment data $m$-times. Hence results about the existence and well-posedness of Hermite refinement schemes imply existence and well-posedness of moment-interpolating refinement schemes, and the results we obtain here about the smoothness properties of moment-interpolating refinement imply parallel smoothness properties of Hermite refinement.

The multiwavelets constructed here are derived from very simple and concrete ideas, and there is no need to introduce sophisticated tools of multichannel filtering, matrix polynomials, and related factorizations in order to obtain the basic algorithms and properties of these multiwavelets. Accordingly the approach described here might be useful for expository purposes and teaching.

### *1.2. Relation to Other Wavelets and Multiwavelets*

We mention a parallel which some readers will find useful. The Haar system was the first system of wavelets, and Cohen *et al*. [5] derived the family ${}_{1,\tilde{N}}\psi$ of smooth wavelets from biorthogonality with the Haar system. The Alpert system was the first system of multiwavelets, and in this paper we derive smooth multiwavelets which are dual to the Alpert system. Hence these are the multiwavelet analogues of the Cohen–Daubechies–Feauveau ${}_{1,\tilde{N}}\psi$ family:

| | Scalar | Vector |
|---|---|---|
| Classical | Haar | Alpert |
| Biorthogonal | CDF | MI multiwavelets |

### *1.3. Smoothness of Refinement*

As mentioned, our approach to analyze the smoothness of moment-interpolating refinement is to establish an intertwining result with the Hermite-interpolatory subdivision scheme. From this result, one can consequently show that the convergence and smoothness of moment-interpolating refinement schemes can be deduced from those of Hermite-interpolating refinement schemes. Hermite schemes have been studied by Dyn and Levin in [16]. In this paper, we carry out the computational procedures suggested in [16] in order to establish the convergence and suboptimal smoothness bounds for Hermite-interpolating refinement (and consequently, moment-interpolating refinement) schemes of various orders and degrees. In [25], we show further that *optimal* Hölder regularity bounds can be computed under the same framework.

### *1.4. Recursive Partitioning*

A potential application for these multiwavelets comes in understanding and extending recursive partitioning techniques. Suppose we are interested in approximating a function $f$ on $[0, 1]$. Recursive partitioning in its simplest form is a method of adaptive nonlinear approximation obtained by adaptively constructing a recursive dyadic partition of the interval and approximating $f$ on each piece of the partition by a constant. The CART book [3] is the standard reference on piecewise-constant recursive partitioning methods. One can generalize recursive partitioning to higher approximation orders by approximating $f$ on each piece of the partition by a polynomial of degree $< m$. One obtains in this way the possibility of much higher-order accuracy.

Recursive partitioning in general gives discontinuous reconstructions. Using the MI multiwavelets constructed here, there is a natural way to construct partitions adaptively and obtain smooth, high-order accurate reconstructions.

To explain this, we first note that there is an intimate connection between Alpert expansions and optimal recursive partitioning algorithms. Roughly speaking, to each partial Alpert expansion of a function $f$ on the interval whose nonzero multicoefficients occur in a tree pattern, there is a corresponding recursive dyadic partition of the interval for which this partial expansion is the best piecewise polynomial fit to the function with given partition. As a result one can find optimal recursive partitions for adaptive piecewise polynomial approximation by polynomials of degree $< m$ by using fast tree-pruning algorithms organized around the multicoefficients of the Alpert system.

An analogous algorithm can be developed around the multicoefficients of the MI multiwavelets constructed here. For a given $f$, one identifies an optimal partition based on a fast tree-pruning algorithm organized around the MI multicoefficients. To obtain a smooth, high accuracy approximation, one utilizes this partition to define a moment-interpolation problem on the resulting possibly inhomogeneous partition. Using the MI multiwavelets, one can obtain a smooth function whose moments on each piece of the partition agree with the specified Alpert moment. This leads to a fast algorithm for approximate recon-

structions with potentially much higher accuracy and smoothness than simple polynomial approximation on recursive dyadic partitions.

Thus we have an affirmative answer to the question: is there a recursive partitioning algorithm which gives smooth reconstructions and high-order accuracy? We also have an interesting interpretation of partial multiwavelet reconstructions where the nonzero multicoefficients occur in a tree pattern — such tree-structured reconstructions have been the object of attention since Shapiro's introduction of EZW [23]. We show here that to each reconstruction with tree-patterned coefficients based on MI multiwavelets, there corresponds naturally a smooth reconstruction based on MI interpolation to moment constraints defined on a possibly inhomogeneous partition. Hence hierarchical constraints in the coefficient domain can be connected rigorously to hierarchical schemes of approximation in the real domain.

### *1.5. Contents*

The paper is organized as follows. In Section 2 we review some facts about Alpert bases. In Section 3 we introduce the notion of moment-interpolating refinement, describe the link with Hermite refinement schemes, and give results on the existence and smoothness of the fundamental solutions to moment-interpolating schemes. In Section 5 we construct biorthogonal bases, and in Section 6 we give norm equivalence properties for those bases. In Section 7 we describe briefly the adaptation of these bases to life on the interval. In Section 8 we describe potential applications in recursive dyadic partition schemes. In Section 9 we briefly describe software for computing with these multiwavelets.

## 2. ALPERT BASES

Let $\pi^\ell(x)$ denote the $\ell$th Legendre polynomial for the interval $[0,1]$, mutilated to have support $[0,1]$. Thus

$$\pi^0(x) = \mathbb{1}_{[0,1]}(x) \tag{2.1}$$

$$\pi^1(x) = \sqrt{12}(x-1/2)\mathbb{1}_{[0,1]}(x) \tag{2.2}$$

$$\pi^2(x) = \sqrt{180}((x-1/2)^2 - 1/12)\mathbb{1}_{[0,1]}(x), \tag{2.3}$$

etc. For $j$ and $k$ integers, let $I_{j,k} = (k/2^j, (k+1)/2^j)$ denote a typical dyadic interval, and define

$$x^\ell_{j,k} = 2^{j/2}\pi^\ell(2^j x - k),$$

the translate and dilate of $\pi^\ell$ that "lives" on $I_{j,k}$ and has $L^2(\mathbb{R})$-norm 1. Any two such functions $\chi^\ell_{j,k}$, $\chi^{\ell'}_{j,k'}$ with the same scale index $j$ and different $k \neq k'$ or $\ell \neq \ell'$ are orthogonal.

Fix $m > 0$ and set

$$\overline{V}_j = \overline{V}^{(m)}_j = \mathrm{Span}(\chi^\ell_{j,k} : k \in \mathbb{Z},\ 0 \le \ell < m).$$

$\overline{V}_j$ is the collection of all piecewise polynomials which are of degree $< m$ on intervals $I_{j,k}$, $k \in \mathbb{Z}$.

The operator $\overline{P}_j f = \sum_{k\in\mathbb{Z}} \sum_{\ell=0}^{m-1} \langle f, \chi_{j,k}^{\ell}\rangle \chi_{j,k}^{\ell}$ gives orthogonal projection onto $\overline{V}_j$. It is convenient below to adopt the convention that $\chi_{j,k}$ denotes the vector of functions $(\chi_{j,k}^{\ell})_{\ell=0}^{m-1}$ and that $\langle f, \chi_{j,k}\rangle$ is a vector of inner products. Then we can write $\overline{P}_j f = \sum_k \langle f, \chi_{j,k}\rangle^T \chi_{j,k}$, avoiding the summation over the $\ell$ index.

The spaces $\overline{V}_j$ are nested: $\overline{V}_j \subset \overline{V}_{j+1}$; indeed a piecewise polynomial with pieces $(I_{j,k})_k$ is also a piecewise polynomial on the finer collection of pieces $(I_{j+1,k})_k$. It follows that the basis elements $(\chi_{j,k}^{\ell})_{k,\ell}$ all have a representation in terms of elements $(\chi_{j+1,k}^{\ell})_{k,\ell}$ for $\overline{V}_{j+1}$. In vector notation we write this as

$$\chi_{j,k} = H_0 \chi_{j+1,2k} + H_1 \chi_{j+1,2k+1}, \tag{2.4}$$

where the $H_i$ are $m$ by $m$ matrices with entries $(H_i)_{\ell,\ell'} = \langle \chi_{0,0}^{\ell}, \chi_{1,i}^{\ell'}\rangle$, $0 \le \ell, \ell' < m$, $i = 0, 1$.

In the scalar case $m = 1$, the $H_i$ are scalars and $H_0 = H_1 = 1/\sqrt{2}$. In the simplest vector case $m = 2$

$$H_0 = \begin{bmatrix} \tau & 0 \\ -\delta & \eta \end{bmatrix}, \qquad H_1 = \begin{bmatrix} \tau & 0 \\ \delta & \eta \end{bmatrix},$$

where $\tau = 1/\sqrt{2}$, $\delta = \sqrt{3/8}$, $\eta = 1/\sqrt{8}$. (Software is available for the general case $m > 2$; see Section 9.)

We consider now the detail space $\overline{W}_j \subset \overline{V}_j$ and associated projection $\overline{Q}_j f = \overline{P}_{j+1} f - \overline{P}_j f$. $\overline{W}_j$ consists of piecewise polynomials with pieces $(I_{j+1,k})_k$ which are orthogonal to piecewise polynomials with coarser pieces $(I_{j,k})_k$. An orthogonal basis $(h_{j,k}^{\ell})_{k,\ell}$ for $\overline{W}_j$ can be constructed as in [1]. As $\overline{W}_j \subset \overline{V}_{j+1}$, each $h_{j,k}^{\ell}$ can be expressed in terms of $(\chi_{j+1,k'}^{\ell'})_{k',\ell'}$:

$$h_{j,k} = G_0 \chi_{j+1,2k} + G_1 \chi_{j+1,2k+1} \tag{2.5}$$

(in vector notation). In the case $m = 1$, the $G_i$ are scalars: $G_0 = 1/\sqrt{2}$, $G_1 = -1/\sqrt{2}$. In the case $m = 2$, the $G_i$ obey

$$G_0 = \begin{bmatrix} \eta & \delta \\ 0 & \tau \end{bmatrix}, \qquad G_1 = \begin{bmatrix} -\eta & \delta \\ 0 & -\tau \end{bmatrix}.$$

The functions $h_{j,k}^{\ell}$ are piecewise polynomials supported in $I_{j,k}$ with knots at the endpoints and midpoint of $I_{j,k}$. For pictures see Alpert *et al.* [1] or Alpert [2].

A key fact about the two-scale matrices we have just defined is that the $2m \times 2m$ matrix

$$U = \begin{bmatrix} H_0 & H_1 \\ G_0 & G_1 \end{bmatrix}$$

is orthogonal. This is equivalent to saying we have two different orthogonal bases for $\overline{V}_1$: $\{(\chi_{1,k}^{\ell})_{k,\ell}\}$ and $\{(\chi_{0,k}^{\ell})_{k,\ell}, (h_{0,k}^{\ell})_{k,\ell}\}$.

### *2.1. Multiwavelet Pyramid Algorithm for Alpert Bases*

There is a very simple pyramid algorithm for calculating the coefficients of Alpert expansions. For the vector functions $\chi_{j,k}$ and $h_{j,k}$, we adopt the convention that $\mu_{j,k} =$

$\langle f, \chi_{j,k}\rangle$ and $\overline{\alpha}_{j,k} = \langle f, h_{j,k}\rangle$ denote vectors with entries $\mu^{\ell}_{j,k} = \langle f, \chi^{\ell}_{j,k}\rangle$ and $\overline{\alpha}^{\ell}_{j,k} = \langle f, h^{\ell}_{j,k}\rangle$, respectively. Then from the two-scale relations (2.4) and (2.5) we have

$$\mu_{j,k} = H_0\mu_{j+1,2k} + H_1\mu_{j+1,2k+1} \tag{2.6}$$

$$\overline{\alpha}_{j,k} = G_0\mu_{j+1,2k} + G_1\mu_{j+1,2k+1}. \tag{2.7}$$

The *pyramid decomposition* is a way to calculate $(\mu_{j_0,k})_k, (\overline{\alpha}_{j_0,k})_k, \ldots, (\overline{\alpha}_{j_1-1,k})_k$ from $(\mu_{j_1,k})_k$ when $j_0 < j_1$. One simply applies (2.6)–(2.7) recursively, for $j = j_1 - 1$, $j_1 - 2, \ldots, j_0$.

The orthogonality of $U$ implies also

$$\mu_{j+1,2k} = H_0^T\mu_{j,k} + G_0^T\overline{\alpha}_{j,k} \tag{2.8}$$

$$\mu_{j+1,2k+1} = H_1^T\mu_{j,k} + G_1^T\overline{\alpha}_{j,k}. \tag{2.9}$$

The pyramid reconstruction is a way to calculate $(\mu_{j_1,k})_k$ from $(\mu_{j_0,k})_k, (\overline{\alpha}_{j_0,k})_k, \ldots, (\overline{\alpha}_{j_1-1,k})_k$. One simply applies (2.8)–(2.9) for $j = j_0, \ldots, j_1 - 1$.

### *2.2. Basis Properties*

The Alpert wavelets provide orthogonal bases for $L^2(\mathbb{R})$. We have both the homogeneous decomposition

$$f \sim \sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty} \langle f, h_{j,k}\rangle^T h_{j,k}$$

and the inhomogeneous decomposition

$$f \sim \sum_{k=-\infty}^{\infty} \langle f, \chi_{j_0,k}\rangle^T \chi_{j_0,k} + \sum_{j\ge j_0}\sum_{k=-\infty}^{\infty} \langle f, h_{j,k}\rangle^T h_{j,k},$$

where each equality is in the sense of *unconditional* $L^2$-convergence of the right side to the left side. These identities are easy to motivate from the pyramid algorithm. From $\overline{P}_{j_1} f = \overline{P}_{j_0} f + \sum_{j_0\le j<j_1} \overline{Q}_j f$ we get

$$P_{j_1} f = \sum_{k=-\infty}^{\infty} \langle f, \chi_{j_0,k}\rangle^T \chi_{j_0,k} + \sum_{j_0\le j<j_1}\sum_{k=-\infty}^{\infty} \langle f, h_{j,k}\rangle^T h_{j,k}.$$

Then from $P_{j_1} f \to f$ as $j \to \infty$ one motivates the inhomogeneous decomposition; and from $P_{j_0} f \to 0$ as $j \to -\infty$ one motivates the homogeneous decomposition.

## 3. MOMENT-INTERPOLATING REFINEMENT

Moment-interpolating refinement is an answer to the following problem:
Given the local moments of a function $f$,

$$\mu^{\ell}_{0,k} = \langle f, \chi^{\ell}_{0,k}\rangle, \qquad k \in \mathbb{Z},\ 0 \le \ell < m, \tag{3.1}$$

construct a smooth function $\tilde{f}$ matching those moments.

In this section we construct a linear operator $\mathcal{R}_0^*$ which solves this problem; i.e.,

$$\tilde{f} = \mathcal{R}_0^*((\mu_{0,k})_k)$$

gives a smooth function with prescribed moments

$$\langle \tilde{f}\chi_{0,k}\rangle = \mu_{0,k}, \qquad k \in \mathbb{Z}.$$

*3.1. Two-Scale Refinement*

We start by defining a two-scale refinement operator which, for a given sequence $(\mu_{j,k})_k$ of moment vectors $\mu_{j,k} = \langle f, \chi_{j,k}\rangle$, delivers a predicted sequence $(\tilde{\mu}_{j+1,k})_k$ of moment vectors at the next finer scale.

Fix $L > 0$ and set $D = (2L+1)m - 1$. The process works as follows, for each $k \in \mathbb{Z}$.

- (Moment interpolation). Find a polynomial $\pi_{j,k}$ of degree $D$ with the prescribed moments $\mu_{j,k'}$ for $k'$ in a neighborhood of $k$:

$$\langle \pi_{j,k}, \chi_{j,k+h}\rangle = \mu_{j,k+h}, \qquad -L \le h \le L. \tag{3.2}$$

- (Moment imputation). Calculate the moments of $\pi_{j,k}$ at the next finer scale:

$$\tilde{\mu}_{j+1,2k+h} = \langle \pi_{j,k}, \chi_{j,k+h}\rangle, \qquad h = 0, 1. \tag{3.3}$$

The moment-interpolation problem (3.2) has a unique solution. The vector equations impose exactly $D+1$ constraints on $\pi_{j,k}$, which is a polynomial of degree $D$. These constraints can be proven to be linearly independent over the space of polynomials of degree $D$. This will follow from results in Section 3.3 below.

This process implicitly defines an operator $R$ $(= R_{m,L})$ such that

$$(\tilde{\mu}_{j+1,k})_k = R((\mu_{j,k})_k).$$

Note that the operator does not depend on $j$; it can be given explicit forms as

$$\tilde{\mu}_{j+1,2k} = \sum_{h=-L}^{L} M_h^e \mu_{j,k-h}, \qquad \tilde{\mu}_{j+1,2k+1} = \sum_{h=-L}^{L} M_h^o \mu_{j,k-h} \tag{3.4}$$

for appropriate $m \times m$ matrices $(M_h^e, M_h^o)$, $h = -L, -L+1, \ldots, L$. (Software for calculating these matrices is available; see Section 9.)

*Remark 1.* The refinement operator $R$ is consistent with coarsening. Formally, if we define the coarsening operator $C((\mu_{j+1,k})_k) = (H_0\mu_{j+1,2k} + H_1\mu_{j+1,2k+1})_k$ then

$$C \circ R = I. \tag{3.5}$$

Indeed, let $(\tilde{\mu}_{j+1,k})_k = R((\mu_{j,k})_k)$. Let $(\tilde{\mu}_{j,k})_k = C((\tilde{\mu}_{j+1,k}))$. Then

$$\begin{aligned}\tilde{\mu}_{j,k} &= H_0\langle \pi_{j,k}, \chi_{j+1,2k}\rangle + H_1\langle \pi_{j,k}, \chi_{j+1,2k+1}\rangle \\ &= \langle \pi_{j,k}, H_0\chi_{j+1,2k} + H_1\chi_{j+1,2k+1}\rangle \\ &= \langle \pi_{j,k}, \chi_{j,k}\rangle = \mu_{j,k},\end{aligned}$$

where the last equality comes from the constraints defining $\pi_{j,k}$.

*Remark 2.* The refinement scheme is accurate on polynomials. Let $\pi$ be a polynomial of degree $D$, and suppose that $\mu_{j,k} = \langle \pi, \chi_{j,k}\rangle$ are the local moments of $\pi$. Apply two-scale refinement; then $\pi_{j,k} = \pi$ for every $k$. Hence $\tilde{\mu}_{j+1,2k+h} = \langle \pi_{j,k}, \chi_{j+1,2k+h}\rangle = \langle \pi, \chi_{j+1,2k+h}\rangle$ for all $k$ and for $h = 0, 1$. That is, the predicted local moments at scale $j+1$ are precisely correct.

### 3.2. Iterated Refinement

Start with prescribed local moment vectors $(\mu_{0,k})_k$ at scale $j = 0$, and iterate the two-scale refinement operator:

$$\begin{aligned}(\tilde{\mu}_{1,k})_k &= R((\mu_{0,k})_k) \\ &\vdots \\ (\tilde{\mu}_{h,k})_k &= R((\tilde{\mu}_{h-1,k})_k).\end{aligned}$$

Define

$$\tilde{f}_h = \sum_k \tilde{\mu}^T_{h,k}\chi_{h,k}, \qquad h = 1, 2, \ldots. \tag{3.6}$$

This is the function whose behavior at scale $2^{-h}$ has been imputed from unit scale by refinement. We note that the coarse scale moments of $\tilde{f}_h$ are as prescribed:

$$\langle \tilde{f}_h, \chi_{0,k}\rangle = \mu_{0,k} \qquad \forall k.$$

This follows from $h$ consecutive applications of (3.5).

DEFINITION 3.1. A MI scheme is $C^r$, $r \geq 0$, if, for any initial sequence $\{\mu_{0,k}\}_k \in l^\infty$, $\tilde{f}_h$ converges uniformly to a function $f \in C^r$.

*Main Result.* Let $m \in \{1, 2, 3\}$ and $L \in \{1, 2, 3, 4, 5\}$. Let $(\mu_{0,k})_k$ be an $l^\infty$ sequence. Then $\tilde{f}_h$ converges in $L^\infty$ as $h \to \infty$ to a continuous limit $\tilde{f}$. This limit is unique.

This result follows from Section 4 below.

From this, we obtain a linear operator $\mathcal{R}^*_0$ solving problem (3.1) via

$$\mathcal{R}^*_0((\mu_{0,k})_k) = \lim_{h\to\infty} \tilde{f}_h. \tag{3.7}$$

### 3.3. Hermite Interpolation

Our main result about properties of moment interpolation comes from its close relationship with another refinement scheme — Hermite interpolation.

Suppose we are given the values of a function and its first $m-1$ derivatives at the integers:

$$f^{(\ell)}(k) = \beta^\ell_{0,k}, \qquad k \in \mathbb{Z},\ 0 \leq \ell < m.$$

Hermite interpolation is the problem of finding a smooth function $f(t)$ obeying the given conditions.

A two-scale refinement scheme can be developed based on imputing the values at the half integers $f^{(\ell)}(k/2) = \beta^{\ell}_{1,2k+1}$ from the values at the integers.

Fix $L > 0$ and set $\tilde{D} = (2L+2)m - 1$. The process works as follows, for each $k \in \mathbb{Z}$.

- (Polynomial interpolation). Find a polynomial $\tilde{\pi}_{0,k}$ of degree $\tilde{D}$ with the prescribed point values $\beta_{0,k'}$ for $k'$ in a neighborhood of $k$:

$$\tilde{\pi}^{(\ell)}_{0,k}(k+h) = \beta^{\ell}_{0,k+h}, \qquad -L \le h \le L+1. \tag{3.8}$$

- (Polynomial imputation). Calculate the point values of $\pi_{0,k}$ at the next finer scale:

$$\tilde{\beta}^{(\ell)}_{1,2k+1} = \tilde{\pi}^{(\ell)}_{0,k}(k+1/2), \qquad k \in \mathbb{Z},\ 0 \le \ell < m. \tag{3.9}$$

This implicitly defines an operator $\tilde{R} = \tilde{R}_{m,L}$ which yields

$$(\tilde{\beta}_{j+1,k})_k = \tilde{R}((\tilde{\beta}_{j,k})_k).$$

One then iterates this process to obtain Hermite data $(\tilde{\beta}_{j,k})_k$ on all dyadic rationals $x_{j,k}$ for all $j \ge 0$, $k \in \mathbb{Z}$. By establishing a certain smoothness property of the Hermite data on the dense set $\{x_{j,k}\} \subset \mathbb{R}$ one can then uniquely extend $(\tilde{\beta}_{j,k})_k$ to a smooth function $f(t)$ defined on the whole real axis.

This scheme has been proposed by Merrien [20] and studied by Dyn and Levin [16]. In the case $m = 1$ this is just the scheme of Deslauriers–Dubuc interpolation [10, 13] about which a considerable amount is known. It is well known that the polynomial interpolation problem (3.8) is well posed; i.e., for any choice of $L = 1, 2, \ldots$ and $m = 1, 2 \ldots$, the problem (3.8) has a unique solution for any choice of data vectors $\beta_{0,k}$. Moreover, there is a very simple and numerically stable way to compute the polynomial via the classical divided difference approach [6] in numerical analysis.

DEFINITION 3.2. A Hermite scheme is $C^r$, $r \ge m-1$, if for any initial sequence $\{\beta_{0,k}\} \in l^{\infty}$ there exists a limit function $f \in C^r$ such that $f^{(\ell)}(x_{j,k}) = \beta^{\ell}_{j,k}$ for all $j \ge 0$, $k \in \mathbb{Z}$, and $\ell = 0, \ldots, (m-1)$. In this case, we write $f = \tilde{R}^*_0(\{\beta_{0,k}\}_k)$ (compare (3.7)).

As in (3.4), there exist $2L$ $m \times m$ matrices $A^{(j)}_h$ at each scale $j$ such that

$$\tilde{\beta}_{j+1,2k+1} = \sum_{h=-L}^{(L-1)} A^{(j)}_h \beta_{j,k-h} \qquad \text{and} \qquad \tilde{\beta}_{j+1,2k} = \beta_{j,k}. \tag{3.10}$$

As opposed to the MI case, the matrices $\{A^{(j)}_h\}$ are dependent on $j$. However, it can be shown that the $M(2^{-(j+1)})^{-1} A^{(j)}_h M(2^{-j})$ are independent of $j$, where $M(h) = \mathrm{diag}(1, h, h^2, \ldots, h^{(m-1)})$. Thus, the scaled Hermite vectors $\{\overline{\beta}_{j,k} = M(2^{-j})^{-1}\beta_{j,k}\}$ follow a *stationary* subdivision scheme with level-independent mask $\overline{A}_h = M(2^{-1})^{-1} A^{(0)}_h$. For more details, see [16].

### 3.4. *Intertwining with Hermite Interpolation*

The moment-interpolation problem looks at first glance to involve quite different notions than Hermite interpolation. For one thing, moment functionals exist for any locally integrable function, while derivatives at a point exist only for sufficiently smooth functions. Nevertheless there are close connections between the problems.

In this section, we establish a formal connection by showing that there is an intertwining relationship between the two schemes; under a certain correspondence of inputs, there is a one-to-one connection between outputs in one scheme and outputs in the other. The purpose of the derivation is three-fold. First, from the intertwining property and the well-posedness of Hermite interpolation, we can obtain the well-posedness of moment interpolation. Second, from the numerically stable algorithm for Hermit interpolation we can obtain a numerically stable algorithm for moment interpolation. This fact is used in our software to accurately generate the mask for moment-interpolating refinement. Third, and most importantly, the intertwining property gives us the smoothness of moment-interpolating refinement directly from information about the smoothness of Hermite-interpolating refinement.

*3.4.1.* $m = 1$. In the scalar case, the intertwining relationship is very easy to see. Suppose we take the prescribed local averages and form their cumulative sums: $\beta_{0,k} = \sum_{-\infty}^{k} \mu_{0,k}$. Then we can identify $\beta_{0,k}$ with $F(k+1)$ where $F(t) = \int_{-\infty}^{t} f(t)\,dt$. Now interpolating the point values of $F(k)$ gives us a function whose derivative $f = F'$ has prescribed averages: $\int_k^{k+1} f(t)\,dt = F(k+1) - F(k) = \mu_{0,k}$. Therefore a solution to the point-value interpolation problem for the cumulative sums leads, upon differentiation, to a solution to the average-interpolation problem.

Moreover, there is a very precise connection between Deslauriers–Dubuc and average interpolation (compare [11]). For $m = 1$ and a given value of $L$, apply the Deslauriers–Dubuc approach to the cumulative $(\beta_{0,k})_k$ of the prescribed averages $(\mu_{0,k})_k$. This gives, for each $k$, a polynomial $\tilde{\pi}_{0,k}$ of degree $\tilde{D}$. Apply the moment-interpolating approach, for $m = 1$ and the same value of $L$, to the prescribed averages $\mu_{0,k}$. This gives a polynomial $\pi_{0,k}$ of degree $D = \tilde{D} - 1$. Then we have

$$\pi_{0,k} = \frac{d}{dx}\tilde{\pi}_{0,k}.$$

The two schemes actually use in some sense the "same" polynomial to impute behavior to finer scales; it follows that

$$\tilde{\mu}_{1,2k} = \tilde{\beta}_{1,2k+1} - \tilde{\beta}_{1,2k}.$$

We can write this as saying that there are relations between the two operators,

$$R = \Delta \circ \tilde{R} \circ \Sigma,$$

where $(\Delta x)_k = x_{k+1} - x_k$ and $(\Sigma x)_k = \sum_{h \le k} x_h$. As $\Delta$ and $\Sigma$ are formal inverses of each other, this says that the operator $\Delta$ *intertwines* the two refinement schemes.

This pattern continues at higher orders, but becomes more complicated.

*3.4.2. General relation.* It is most natural to use the Bernstein basis to describe the general intertwining relation. Let

$$B^{m,\ell}(x) = \frac{(m-1)!}{\ell!(m-1-\ell)!} x^\ell (1-x)^{m-1-\ell}, \qquad 0 \le \ell < m,$$

denote the $\ell$th element of the Bernstein basis for polynomials of degree up to $m-1$. Denote $b^{m,\ell} = B^{m,\ell} 1_{[0,1]}$ and $b^{m,\ell}_{j,k} = 2^{j/2} b^{m,\ell}(2^j \cdot - k)$. Also, denote the multiple knots at scale $j$ by $t^j_i = \lfloor i/m \rfloor 2^{-j}$.

We first point out the close connection between Hermite sampling and moment sampling.

LEMMA 3.3. *Let* $(m-1)!F^{(m)} = f$.

$$[t^j_i, \ldots, t^j_{i+m}]F = \langle f, b^{m,\ell}_{j,k} \rangle,$$

*where* $k = \lfloor i/m \rfloor$ *and* $\ell = i - km$.

*Proof.* It is well known the lemma is true for $j = 0$ and $k = 0$ due to the fact that $\{b^{m,\ell} : 0 \le \ell < m\}$ are the B-splines of degree $m-1$ defined on $(0,1)$, with the knots $0, 1$ having multiplicity $m$ [9]. A simple rescaling gives the lemma. ■

This lemma gives a one–one correspondence between vectors of moments $\{\mu^\ell_{j,k} = \langle f, \chi^{m,\ell}_{j,k'} \rangle\}_{k \in \mathbb{Z}}$ and vectors of Hermite data $\{\beta_{j,k} = F^{(\ell)}(x_{j,k})\}_{k \in \mathbb{Z}}$. First of all, since the Bernstein polynomials of degree $m-1$ form a basis of the space of polynomials of degree up to $m-1$, and since the Legendre polynomials of degree up to $m-1$ form also such a basis, there is a matrix $\mathbf{B}$ of order $m \times m$, such that $\{b^{m,\ell}\}_{\ell=0}^{m-1} = \mathbf{B}\{\chi^\ell\}_{\ell=0}^{m-1}$ or more generally $\{b^{m,\ell}_{j,k}\}_{\ell=0}^{m-1} = \mathbf{B}\{\chi^\ell_{j,k}\}_{\ell=0}^{m-1}$. Hence the vector of $b$-moments,

$$\hat{\mu}_k = \{\langle f, b^{m,\ell}(\cdot - k) \rangle : 0 \le \ell < m\}, \tag{3.11}$$

is obtained by the matrix transformation $\hat{\mu}_k = \mathbf{B}\mu_k$.

Given a set of Hermite data $\{\beta_{j,k}\}_{k \in \mathbb{Z}}$, one can work down the divided difference table to obtain the $m$th order divided differences $[t^j_i, \ldots, t^j_{i+m}]F$. We denote this operation by $\Delta^m\{\beta_{j,k}\} = \{[t^j_i, \ldots, t^j_{i+m}]F\}$. Lemma 3.3 states that $\mathbf{B}^{-1} \circ \Delta^m$ maps the Hermite samples of $F$ to the moments (w.r.t. to the Legendre basis) of $f = (m-1)!F$.

For the converse direction of the correspondence, we assume that any given set of moment data $\{\mu_{j,k}\}$ is such that $\mu_{j,k} = 0$ for all $k < K$. Let $f = \sum_k \mu^T_{j,k} \chi_{j,k}$ and

$$F(x) = \frac{1}{(m-1)!} D^{-m} f(x) = \frac{1}{(m-1)!} \int_{-\infty}^{x} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_{m-1}} f(x_m)\, dx_m \cdots dx_1.$$

Then one obtains the $\{\beta_{j,k}\}_{k \in \mathbb{Z}}$ by sampling the Hermite data of $F$ at dyadic points $x_{j,k}$. Notice that a compactly supported $\{\mu_{j,k}\}$ does not give a compactly supported (nor even bounded) $\{\beta_{j,k}\}$ in this correspondence.

THEOREM 3.4. *Given a set of moment data* $\{\mu^\ell_{j,k'} : k - L \le k' \le k + L\}$ *defined on the dyadic intervals* $I_{j,k'}, k - L \le k' \le k + L$, *there exists a set of Hermite data* $\{\beta_{j,k'} : k - L \le k' \le k + L + 1\}$ *defined on the dyadic points* $x_{j,k'}, k - L \le k' \le k + L + 1$, *such that the (unique) Hermite interpolant of* $\{\beta_{j,k'}\}$, *denoted by* $p_{j,k}(x)$, *is such that* $\pi_{j,k}(x) = (m-\ell)!(d^m p_{j,k}(x))/dx^m$ *is the moment interpolant of* $\{\mu_{j,k'}\}$.

*Proof.* It suffices to prove the theorem for $j = 0$ and $k = 0$. Given $\{\mu^{\ell}_{0,k'} : -L \le k' \le L\}$, use the one–one correspondence to obtain $\{\beta_{0,k'} = F^{(\ell)}(k') : -L \le k' \le L+1\}$. Denote by $p$ the Hermite-interpolating polynomial of degree $(2L+2)m-1$ to the above Hermite data, which is known to exist and is unique (see, e.g., [6]). Since $p$ and $F$ share the same Hermite data of order up to $m-1$ at $k = -L, \ldots, L+1$, we have

$$[t_i, \ldots, t_{i+m}]F = [t_i, \ldots, t_{i+m}]p, \qquad -mL \le i < m(L+1).$$

Hence, by Lemma 3.3, the degree $(2L+1)m-1$ polynomial $\pi = (m-1)!p^{(m)}$ has the desired moments. ■

COROLLARY 3.5 (Well-posedness of moment interpolation). *The moment-interpolation Problem (3.2) has a unique solution.*

*Proof.* The determination of the moment-interpolating polynomial (3.2) is equivalent to the solution of a linear system of $(2L+1)m$ equations. The existence of a moment interpolant for any given set of moments proves that the system is nonsingular, and hence the solution must be unique. ■

Now we understand that the moment-interpolating polynomial $\pi_{j,k}$ of $\{\mu_{j,k-L}, \ldots, \mu_{j,k+L}\}$ is exactly the $m$th derivative (multiplied by $(m-1)!$) of the Hermite-interpolating polynomial of a corresponding Hermite data $\{\beta_{j,k-L}, \ldots, \beta_{j,k+L+1}\}$. In MI refinement, one obtains scale $j+1$ moment data by sampling the moments of $\pi_{j,k}$ at $I_{j+1,2k}$ and $I_{j+1,2k+1}$, whereas in Hermite-interpolating refinement, one obtains scale $j+1$ Hermite data by sampling the Hermite data of $p_{j,k}$ at $x_{j+1,2k+1}$. Using Lemma 3.3, we obtain a one–one correspondence between $\{b_{j+1,k}\}$ and $\{\mu_{j+1,k}\}$.

COROLLARY 3.6. *Using the notation in the proof of Theorem 3.4,*

$$\{\langle \pi_{j,k}, \chi^{\ell}_{j+1,2k+\epsilon}\rangle\}_{\ell=0}^{m-1} = \mathbf{B}^{-1}\{[t_i^{j+1}, \ldots, t_{i+m}^{j+1}]p_{j,k}\}_{i=km+\epsilon m}^{(k+1)m-1+\epsilon m}, \qquad \epsilon = 0, 1.$$

Notice that the above results can be stated compactly in the following commutative diagram.

$$\begin{array}{ccc}
\{\beta^{\ell}_{j,k}\} & \xrightarrow{\mathbf{B}^{-1}\circ\Delta^m} & \{\mu^{\ell}_{j,k}\} \\
\text{Hermite interpolation}\downarrow & & \downarrow\text{Moment interpolation} \\
\{p_{j,k}\} & \xrightarrow{(m-1)!\frac{d^m}{dx^m}} & \{\pi_{j,k}\} \\
\text{Hermite sampling at } x_{j+1,2k+1}\downarrow & & \downarrow\text{Moment sampling at } I_{j+1,2k}, I_{j+1,2k+1} \\
\{\beta^{\ell}_{j+1,k}\} & \xrightarrow{\mathbf{B}^{-1}\circ\Delta^m} & \{\mu^{\ell}_{j+1,k}\}
\end{array}$$

### 3.5. *Difference and Divided Difference Schemes*

In the analysis of scalar subdivision schemes, Dyn *et al.* [15] and Rioul [22] independently proposed the concept of divided difference schemes (related methodologies were proposed by Daubechies and Lagarias [7, 8] and by Micchelli and Prautzsch [21]). Roughly speaking, a refinement scheme of the form $f^{j+1} = Bf^j$ converges uniformly to a continuous limit if and only if the difference sequence $\partial f^j = f^j_{\cdot} - f^j_{\cdot-1}$ follows a

*difference scheme* of the form $\partial f^{j+1} = \partial B \partial f^j$ and if said scheme is contractive. Here by "$T$ contractive" we mean that the norm of the operator $T$ as a mapping from $\ell^\infty$ to $\ell^\infty$ is strictly less than 1. More generally, the scheme converges to a $C^r$ limit if and only if the difference sequence of the $r$th order divided difference sequence follows a derived scheme of the form $\partial \Delta^r f^{j+1} = \partial B^{[r]} \partial \Delta^r f^j$ and is contractive [15]. Here $B^{[r]}$ is the so-called *$r$th divided difference mask* and $\partial B^{[r]}$ the *difference mask* of $B^{[r]}$. Rioul [22] gave detailed results in determining exact (noninteger) Hölder exponents of refinement solutions using the same framework. He showed that one can also verify $C^r$ convergence and the precise Hölder regularity by studying "over-differencing" schemes $B^{[s]}$ for $s > r$.

Our analysis of the convergence and smoothness of Hermite-interpolatory schemes is based on generalizing results in [15, 22] to the Hermite case. Let $\{\beta_{j,k} : j > 0,\ k \in \mathbb{Z}\}$ be generated by the Hermite$(m, L)$ scheme with initial data $\{\beta_{0,k} : k \in \mathbb{Z}\}$. Recall each $\beta_{j,k} \in \mathbb{R}^m$, and, for $\ell = 0, \ldots, (m-1)$, $\beta^\ell_{j,k}$ is interpreted as $F^\ell(x_{j,k})$ for some $F \in C^{m-1}$. [1]

Let

$$t_{j,n} = 2^{-j} \left\lfloor \frac{n}{m} \right\rfloor.$$

For $m - 1 \le r \le 2mL$, define $u^{[r]}_{j,k} \in \mathbb{R}^m$ by

$$\left(u^{[r]}_{j,k}\right)_\ell = [t_{j,km+\ell}, \ldots, t_{j,km+\ell+r}]F, \qquad \ell = 0, \ldots, (m-1). \tag{3.12}$$

For example, when $r = m - 1 = 2$, one has

$$u^{[m-1]}_{j,k} = \left[ [x_{j,k}, x_{j,k}, x_{j,k}]F = \frac{\beta^2_{j,k}}{2!}, [x_{j,k}, x_{j,k}, x_{j,k+1}]F, [x_{j,k}, x_{j,k+1}, x_{j,k+1}]F \right]^T.$$

We will denote the operator that maps $\{\beta_{j,k}\}_k$ to $\{u^{[r]}_{j,k}\}_k$ by $\Delta^r$ and write $\{u^{[r]}_{j,k}\} = \Delta^r \beta_{j,k}$.

For $m - 1 \le r \le 2mL - 1$, define the forward differencing sequence of $u^{[r]}_{j,k}$ by

$$(\partial u^{[r]}_{j,k})_\ell = \begin{cases} (u^{[r]}_{j,k})_\ell - (u^{[r]}_{j,k-1})_{\ell-1} & 0 < \ell \le m-1, \\ (u^{[r]}_{j,k})_0 - (u^{[r]}_{j,k-1})_{m-1} & \ell = 0. \end{cases} \tag{3.13}$$

By the definition of divided difference, $\partial u^{[r-1]}_{j,k}$ and $u^{[r]}_{j,k}$ are related by

$$\left(u^{[r]}_{j,k}\right)_\ell = \left(\partial u^{[r-1]}_{j,k}\right)_\ell / \left(2^{-j} C(\ell; m, r)\right), \tag{3.14}$$

where $C(\ell; m, r) = \lfloor (\ell + r)/m \rfloor$.

In [16], Dyn and Levin showed that for any Hermite-type refinement scheme that reproduces polynomials of degree up to $D$, $\{u^{[r]}_{j,k}\}$, $r = (m-1), \ldots, (D+1)$, and $\{\partial u^{[r]}_{j,k}\}$, $r = (m-1), \ldots, D$, follow vector refinement schemes with compactly supported masks. Translating their results to our setting, we have

[1] Notice it is solely for notational convenience that we assume such a function $F$ exists. In fact the definitions of $u^{[r]}_{j,k}$ and $\partial u^{[r]}_{j,k}$ do not depend on $F$ but only on $\beta^\ell_{j,k}$.

• for $(m-1) \le r \le 2mL$, there exist finitely supported matrix sequences $\{M_\ell^{[r]}\}_\ell \in \ell^0(\mathbb{R}^{m\times m})$ such that

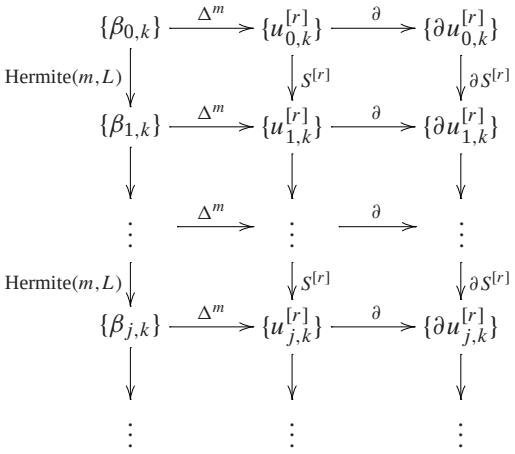
$$u_{j+1,k}^{[r]} = \sum_\ell M_{k-2\ell}^{[r]} u_{j,\ell}^{[r]}, \qquad \text{and} \tag{3.15}$$

• for $(m-1) \le r \le 2mL-1$, there exist finitely supported matrix sequences $\{C_\ell^{[r]}\}_\ell \in \ell^0(\mathbb{R}^{m\times m})$ such that

$$\partial u_{j+1,k}^{[r]} = \sum_\ell C_{k-2\ell}^{[r]} \partial u_{j,\ell}^{[r]}. \tag{3.16}$$

Dyn and Levin [16] also provide formulae for computing the matrices $\{M_\ell^{[r]}\}$ and $\{C_\ell^{[r]}\}$ from $\{A_\ell\}$ defined in Section 3.3. We will generally denote by $S^{[r]}$ the refinement scheme of $\{u_{j,k}^{[r]}\}$ and by $\partial S^{[r]}$ that of $\{\partial u_{j,k}^{[r]}\}$. We call $\{M_\ell^{[r]}\}$ and $\{C_\ell^{[r]}\}$ the refinement *masks* of $S^{[r]}$ and $\partial S^{[r]}$, respectively.

Their results can be described vividly by the following commutative diagram.




Although we will not repeat the proofs of the above facts, we stress that the existence of a refinement scheme with compactly supported mask for $\{u_{j,k}^{[r+1]}\}$ is closely connected to the fact that the scheme $S^{[r]}$ can *reproduce constants*. Moreover, the fact that Hermite$(m, L)$ reproduces polynomials of degree up to $2mL-1$ implies that $S^{[r]}$ reproduces polynomials of degree up to $2mL-r$. Hence, by induction, $S^{[r]}$ can be constructed for all $r \le 2mL$.

To end this section, we comment that the matrix sequence $\{\mathbf{B}^{-1}M_\ell^m\mathbf{B}\}$ is exactly the mask of the MI$(m, L-1)$ scheme. Precisely, with the notation of (3.4), we have $\mathbf{B}^{-1}M_{2h}^{[m]}\mathbf{B} = M_h^e$ and $\mathbf{B}^{-1}M_{2h-1}^{[m]}\mathbf{B} = M_h^o$. This follows directly from the one–one correspondence between Hermite and moment data described in Section 3.4. This observation is crucial for relating the regularity of MI schemes to that of Hermite schemes.

## 4. CONVERGENCE AND SMOOTHNESS

We now investigate the existence and smoothness of Hermite refinement schemes, using an approach due to Dyn and Levin [16]. We report the computational results obtained from this approach, which are sharp in some cases. A companion paper on the regularity

of Hermite schemes by Yu [25] develops a refined approach, together with sharper computational results.

For simplicity, we will use differentiability to measure smoothness. We say that the regularity order of Hermite$(m, L)$ is lower-bounded by $r$ if Hermite$(m, L)$ is $C^r$ in the sense of Definition 3.2. Similarly, we say that the regularity order of MI$(m, L)$ is lower-bounded by $r$ if MI$(m, L)$ is $C^r$ in the sense of Definition 3.1.

### *4.1. General Idea*

A Hermite scheme is $C^r$ (Definition 3.2) if and only if, for any given nontrivial bounded initial Hermite data $(\beta_{0,k})_k$, the difference sequence of the $r$th divided difference sequence $(u^{[r]}_{j,k})_k$ converges uniformly to zero or

$$\lim_{j\to\infty} \sup_{k,\ell} \left|\left(\partial u^{[r]}_{j,k}\right)_\ell\right| = 0. \tag{4.1}$$

By the discussion in Section 3.5, the difference sequences $\partial u^{[r]}_{j,k},\ j = 0, 1, 2, \ldots$, follow exactly a derived scheme with mask $\{C^{[r]}_\ell\}$. This mask represents implicitly the linear map $T\colon\ \ell^\infty(\mathbb{R}^m) \to \ell^\infty(\mathbb{R}^m)$ defined by $T(\partial u^{[r]}_{j,k}) = \partial u^{[r]}_{j+1,k}$. The *contractivity* condition (4.1) is equivalent to saying that

$$\lim_{j\to\infty} T^j(u) = 0 \qquad \forall u \in l^\infty(\mathbb{R}^m). \tag{4.2}$$

It is clear that a sufficient condition for (4.2) to happen is that $\|T\|_{(l^\infty, l^\infty)} < 1$. In fact it is enough to have

$$\|T^\ell\|_{(l^\infty, l^\infty)} = \mu < 1 \qquad \text{for some } \ell < \infty. \tag{4.3}$$

If (4.3) holds, then by letting $j = q^\ell + p,\ 0 \le p < \ell$, we obtain

$$\begin{aligned}\|T^j(u)\| &\le \|T^{q\ell+p}\|\|u\| \le \|T^\ell\|^q \|T\|^p \|u\| \\ &\le \|u\| \max(1, \|T\|, \|T\|^2, \ldots, \|T\|^{\ell-1})\mu^p \to 0 \qquad \text{as } j \to \infty,\end{aligned}$$

which implies (4.2).

Of course $T$ is armed with a subdivision structure, which makes the verification of condition (4.3) very easy. We have:

THEOREM 4.1 (Dyn [14]). *Let* $C(z) := \sum_k C^{[r]}_k z^k$ *and* $C(z)C(z^2)\cdots C(z^{2^{j-1}}) = \sum_k C^{[r]}_{j,k} z^k$. *It can be shown that*

$$\|T^j\|_{(l^\infty, l^\infty)} = \max_{0\le i<2^j} \left\| \sum_k |C^{[r]}_{j,i+k2^j}| \right\|_\infty, \tag{4.4}$$

*where* $(|C|)_{i,j} = |C_{i,j}|$.

The argument we have just presented suggests that the equivalent contractivity conditions (4.1)–(4.3) are equivalent to the $C^r$ convergence of Hermite schemes.

THEOREM 4.2 (Dyn and Levin [16]). *Define the contractivity factor order* $r$, $c^{[r]}_\ell$, *by (4.4); then a Hermite scheme is* $C^r$ *if and only if there exists* $\ell \ge 0$ *such that* $c^{[r]}_\ell < 1$.

This theorem justifies an iterative procedure for checking the convergence and smoothness of Hermite schemes. Call this the *Contractivity Bound Procedure*.

1. Input: $r$, a target regularity order to which Hermite$(m, L)$ is checked; $\ell^{\max}$, a maximum exponent to check in (4.3).
2. For $\ell = 0, 1, 2, \ldots, \ell^{\max}$, compute $c_\ell^{[r]}$ (recall (4.4)), stopping as soon as $c_\ell^{[r]} < 1$.
3. Output: If the algorithm finds $\ell \le \ell^{\max}$ such that $c_\ell^{[r]} < 1$,

   $C^r$ convergence of the Hermite scheme is verified, output $\ell^*(r; m, L) \equiv \ell$

   Else

   Nothing can be concluded, output "?".

Assuming exact arithmetic, the Contractivity Bound Procedure determines a lower bound on the regularity of the Hermite scheme whenever it terminates, i.e., whenever $\ell^* \le \ell^{\max}$. Unfortunately, the cost of implementing the algorithm increases exponentially in $\ell^{\max}$, while there can be no a priori limit on the size of $\ell^{\max}$ needed to obtain sharp results. Hence this procedure, implemented with current computing equipment, cannot always explore a sufficiently high range of $\ell$ to discover that contractivity holds at a given order $r$, even when it does hold at that order.

### 4.2. Computational Results

In Table 1, we show for each $m \in \{1, 2, 3\}$ and $L \in \{1, 2, 3, 4, 5\}$, a pair of indices $r/\ell^*(r; m, L)$ obtained from the Contractivity Bound Procedure. Table 1 shows that Hermite$(1, 1)$ is $C^0$, Hermite$(3, 5)$ is $C^5$, etc.

It is beyond the scope of this paper to discuss the *sharpness* of these lower bounds. In [25], we verify that the upper bounds are optimal in their integral parts, except for the case of $(m, L) = (2, 5)$. We verify in that paper that Hermite$(2, 5)$ is slightly smoother than $C^4$. It has turned out to be too computationally intensive (i.e., $\ell^*(4; 2, 5)$ is too large) to obtain this conclusion using the Contractivity Bound Procedure.

In general, the companion paper [25] gives a method which can be used when the Contractivity Bound Procedure is inconclusive. For instance, it can be shown that for any $\ell \ge 0$, $r - \log_2(c_\ell^{[r]})/\ell$ is a lower bound on the critical Hölder regularity of a Hermite scheme.

### 4.3. Moment-Interpolating Refinement Schemes

We now return to moment-interpolating refinement schemes. Section 3.4 showed that the mask for MI$(m, L)$ is exactly the $m$th order divided difference mask for Hermite$(m, L+1)$,

**TABLE 1**

$r/\ell^*(r; m, L)$ **for** $m \in \{1, 2, 3\}$ **and** $L \in \{1, 2, 3, 4, 5\}$

| | $L=1$ | $L=2$ | $L=3$ | $L=4$ | $L=5$ |
|---|---|---|---|---|---|
| $m=1$ | 0/1 | 1/2 | 2/2 | 3/5 | 4/8 |
| $m=2$ | 1/2 | 2/5 | 3/5 | 3/4 | 3/5 |
| $m=3$ | 2/3 | 3/7 | 4/7 | 4/6 | 5/14 |

**TABLE 2**
**Lower Bounds for Regularity Orders of MI$(m, L)$**

| | $L=1$ | $L=2$ | $L=3$ | $L=4$ |
|---|---|---|---|---|
| $m=1$ | 0 | 1 | 2 | 3 |
| $m=2$ | 0 | 1 | 1 | 1 |
| $m=3$ | 0 | 1 | 1 | 2 |

modulo a change of basis. As a consequence, a moment-interpolating refinement limit MI$(m, L)$ is exactly the $m$th derivative of a Hermite$(m, L+1)$ refinement limit.

From this rule, we have Table 2.

THEOREM 4.3. *If Hermite$(m, L+1)$ is $C^{m+r}$, then MI$(m, L)$ is $C^r$.*

*Proof.*

(I) Let $\beta_{0,k}$ be *any* bounded sequence of $m$-vectors. In particular, it need not be the $m$th order divided difference sequence of another bounded sequence. Apply the $m$th order divided difference mask of Hermite$(m, L)$, $\{M_\ell^{[m]}\}$, to refine $\beta_{0,k}$ into $\beta_{j,k}$, as in (3.15). Assume we have verified that Hermite$(m, L)$ is $C^{m+r}$, $r \geq 0$. Then it is easy to verify that

$$\sum_k \sum_\ell \beta_{j,k} 1_{I_{j,k}^\ell}, \qquad j = 0, 1, \ldots, \tag{4.5}$$

converges uniformly to a $C^r$ function, where $I_{0,0}^\ell$ is an arbitrary partition of $[0, 1)$ and $I_{j,k}^\ell = 2^{-j}(I_{0,0}^\ell + k)$. See [24] for details.

(II) Let $(\mu_{j,k})_k$, $j = 0, 1, \ldots$, be MI$(m, L)$ refinement sequences. We have

$$\sum_k \mu_{j,k}^T \chi_{j,k} = \sum_k \hat{\mu}_{j,k}^T \mathbf{B}^{-T} \mathbf{B}^{-1} b_{j,k}, \tag{4.6}$$

where $\hat{\mu}_{j,k}^T$ are Bernstein moments on dyadic cells $I_{j,k}$.

(III) Recall that $(\hat{\mu}_{j,k})_k$, $j = 0, 1, \ldots$, follows a vector refinement scheme with mask $\{M_\ell^{[m]}\}$. Therefore, by (I), $\sum_k \hat{\mu}_{j,k}^T 1_{I_{j,k}^\ell}$ converges uniformly to a $C^r$ function.

(IV) $\sum_k \hat{\mu}_{j,k}^T \mathbf{B}^{-T} \mathbf{B}^{-1} b_{j,k}$ and $\sum_k \hat{\mu}_{j,k}^T 1_{I_{j,k}^\ell}$ have precisely the same convergence and smoothness properties; thus MI$(m, L)$ is $C^r$ in the sense of Definition 3.1. This is due to the following two results:

1. We call $d = [d^0, \ldots, d^{m-1}]$ a *decomposition of unity* on $[0, 1)$ if $d^\ell \in L^\infty(\mathbb{R})$ and $\sum_\ell d^\ell = 1_{[0,1)}$. Let $d_{j,k}^\ell = d^\ell(2^j \cdot - k)$. If (4.5) converges uniformly to $f$, then it can be shown that, for any decomposition of unity $d$,

$$f = \sum_k \hat{\mu}_{j,k}^T d_{j,k}.$$

See also [24] for details.

2. $(1/m)\mathbf{B}^{-T}\mathbf{B}^{-1}b_{0,0}$ forms a decomposition of unity. Since $b_{0,0}$, the Bernstein basis, forms a decomposition of unity on $[0, 1)$, we prove the claim if we can show that

$$\sum_{j=0}^{m-1}\left(\frac{1}{m}\mathbf{B}^{-T}\mathbf{B}^{-1}\right)_{j,i} = 1, \qquad \forall i = 0, \ldots, (m-1).$$

In terms of eigenvectors and eigenvalues, the above may be written as

$$[\mathbf{1}, \ldots, 1] = [1, \ldots, \mathbf{1}]\left(\frac{1}{m}\mathbf{B}^{-T}\mathbf{B}^{-1}\right)$$

or, equivalently,

$$[1, \ldots, \mathbf{1}] = [\mathbf{1}, \ldots, 1](m\mathbf{B}\mathbf{B}^{T}).$$

Hence it suffices to show that the matrix $(\mathbf{B}\mathbf{B}^T)$ has columns all summing up to $1/m$. By orthogonality of $\chi_i$, $\mathbf{B}_{i,j} = \langle b^{m,i}, \chi_j\rangle$, $0 \le i, j \le (m-1)$, so

$$\begin{aligned}
\sum_j (\mathbf{B}\mathbf{B}^T)_{i,j} &= \sum_j \sum_k \langle b^{m,i}, \chi_k\rangle\langle b^{m,j}\chi_k\rangle \\
&= \sum_k \langle b^{m,i}, \chi_k\rangle \sum_j \langle b^{m,i}, \chi_k\rangle \\
&= \sum_k \langle b^{m,i}, \chi_k\rangle\langle 1, \chi_k\rangle \qquad (\text{since } \textstyle\sum_{i=0}^{m-1} b^{m,i} = 1) \\
&= \langle b^{m,i}, \chi_0\rangle \qquad (\text{since } \langle\chi_0, \chi_k\rangle = \delta_{k,0} \text{ by orthogonality}).
\end{aligned}$$

The lemma follows from the fact that $\int_0^1 b^{m,i}(x)\,dx = 1/m$ for $0 \le i \le (m-1)$, which can be shown easily by induction on $i$. ■

## 5. BIORTHOGONAL EXPANSIONS

Now we use the refinement operator to construct smooth multiwavelet scaling functions $\phi_{j,k}$ and smooth multiwavelets $\psi_{j,k}$ which (1) are dual (in an appropriate sense) to Alpert functions $(x_{j,k})$ and $(h_{j,k})$ and which (2) make Riesz bases.

### *5.1. Smooth Scaling Function and Wavelet*

Define now the sequence of local moment vectors $\mu^{k,\ell} = (\mu_{0,k'})_{k'}$ where $\mu_{0,k'}^{\ell'} = \delta_{k,k'} \cdot \delta_{\ell,\ell'}$. The refinement of the sequence $\mu^{0,\ell}$ produces a smooth limit $\phi^\ell(x)$:

$$\phi^\ell = \mathcal{R}_0^*(\beta^{0,\ell}).$$

This will be called the $\ell$th fundamental solution of the refinement equation, and $\phi(x)$ will denote the vector function with $\ell$th entry $\phi^\ell(x)$.

By linearity of refinement, if $(\mu_{0,k})_k$ is any bounded sequence of moment vectors, and if we set $\phi_{0,k}(x) = \phi(x-k)$, then

$$\mathcal{R}_0^*((\mu_{0,k})_k) = \sum_k \mu_{0,k}^T \phi_{0,k}.$$

Note that $\phi_{0,k}$ is compactly supported, so that the sum $\sum_k \mu_{0,k}^T \phi_{0,k}(x)$ involves only finitely many terms at each $x$. Now because $\phi^\ell$ has prescribed local moments, we have the relation $\langle \chi_{0,k'}^{\ell'}, \phi_{0,k}^{\ell} \rangle = \delta_{kk'}\delta_{\ell\ell'}$, from which follows

$$\left\langle \chi_{0,k}, \sum_{k'} \mu_{0,k'}^T \phi_{0,k'} \right\rangle = \mu_{0,k}.$$

We interpret this fact as follows. Let $V_0 = \text{Span}\{\phi^\ell(\cdot - k),\ k \in \mathbb{Z},\ 0 \le \ell < m\}$. Then

$$f = \sum_k \langle f, \chi_{0,k} \rangle^T \phi_{0,k}, \qquad f \in V_0. \tag{5.1}$$

This says that the *Alpert coefficients of an $f \in V_0$ are exactly the coefficients for an expansion of $f$ in terms of $\phi_{0,k}$*. More generally, we have the following.

THEOREM 5.1. *Set $\phi_{j,k}^{\ell} = 2^{j/2}\phi^\ell(2^j \cdot - k)$ for $j, k \in \mathbb{Z},\ 0 \le \ell < m$, and define*

$$V_j = \text{Span}\{\phi_{j,k}^{\ell} : k \in \mathbb{Z},\ 0 \le \ell < m\}.$$

*Then*

$$f = \sum \langle f, \chi_{j,k} \rangle^T \phi_{j,k}, \qquad f \in V_j, \tag{5.2}$$

*and*

$$f = \sum \langle f, \phi_{j,k} \rangle^T \chi_{j,k}, \qquad f \in \overline{V}_j. \tag{5.3}$$

Relations (5.2) and (5.3) express the biorthogonality of the systems $(\phi_{j,k}^{\ell})_{k,\ell}$ and $(\chi_{j,k}^{\ell})_{k,\ell}$ with $j$ fixed. Members of one system yield coefficients for expansion by the other system.

*Proof.* Of course (5.2) is just a dilation of (5.1) by a power $2^j$. As for (5.3), we have the special relation $\langle \chi_{j,k'}^{\ell'}, \phi_{j,k}^{\ell} \rangle = \delta_{k',k}\delta_{\ell',\ell}$, so that $\langle f, \phi_{j,k} \rangle = \langle f, \chi_{j,k} \rangle,\ f \in \overline{V}_j$, from which (5.3) follows. ■

For future use set

$$P_j f = \sum_k \langle f, \chi_{j,k} \rangle^T \phi_{j,k}. \tag{5.4}$$

Then (5.2) says that

$$f = P_j f, \qquad f \in V_j; \tag{5.5}$$

$P_j$ is an oblique projector onto $V_j$. Similarly, define $\tilde{P}_j f = \sum_k \langle f, \phi_{j,k} \rangle^T \chi_{j,k}$ then

$$f = \tilde{P}_j f, \qquad f \in \overline{V}_j; \tag{5.6}$$

$\tilde{P}_j$ is an oblique projector onto $\overline{V}_j$. It follows that

$$Q_j f = P_{j+1} f - P_j f$$

is an oblique projector; set $W_j = \text{Range}(Q_j)$.

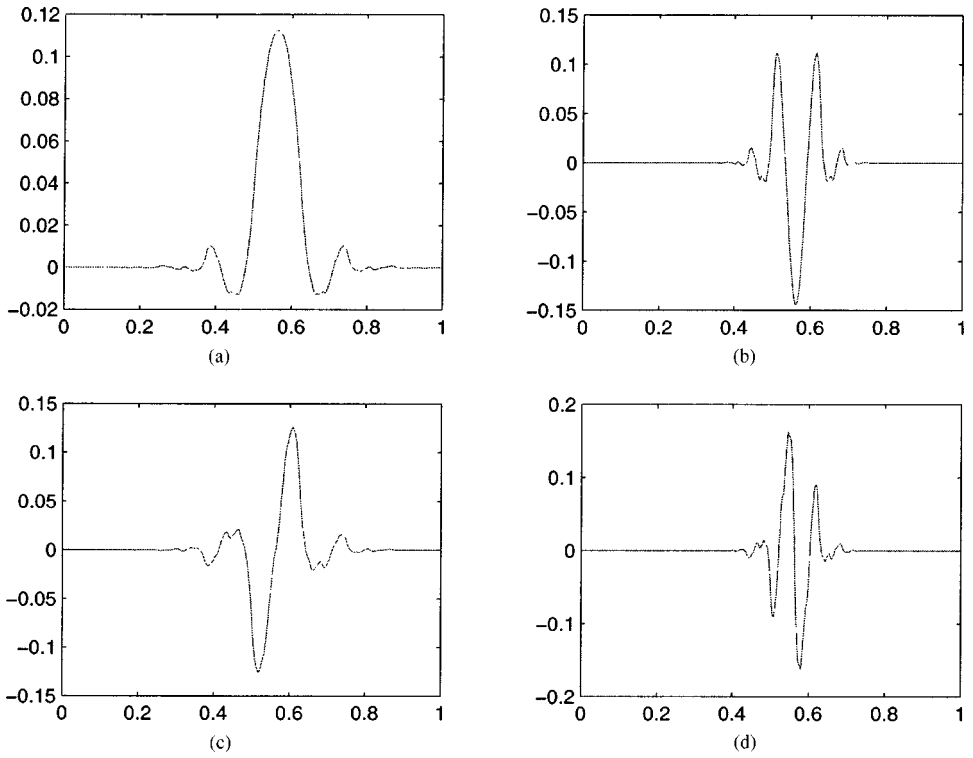



**FIG. 1.** Moment-interpolating multiwavelets $m = 2$, $L = 2$. (a) $\phi^0_{3,4}$, (b) $\psi^0_{3,4}$, (c) $\phi^1_{3,4}$, (d) $\psi^1_{3,4}$.

THEOREM 5.2. *Define the* multiwavelet *(vector-valued function)* $\psi_{j,k}$ *by*

$$\psi_{j,k} = G_0\phi_{j+1,2k} + G_1\phi_{j+1,2k+1}.$$

*Then*

$$f = \sum \langle f, h_{j,k}\rangle^T \psi_{j,k}, \qquad f \in W_j, \tag{5.7}$$

*and*

$$f = \sum \langle f, \psi_{j,k}\rangle^T h_{j,k}, \qquad f \in \overline{W}_j. \tag{5.8}$$

This says that the Alpert multiwavelets $h_{j,k}$ supply the coefficients for an expansion of $f \in W_j$ by the multiwavelets $\psi_{j,k}$, and vice versa.

*Proof.* An $f$ in $W_j$ is also in $V_{j+1}$; so by (5.2) it is represented by its local moments at scale $j+1$. But, as it is in $W_j$ its local moments at scale $j$ vanish: $\mu_{j,k}(f) = 0\ \forall k$. Hence from (2.8)–(2.9) we have that

$$\mu_{j+1,2k} = G_0^T\overline{\alpha}_{j,k}; \qquad \mu_{j+1,2k+1} = G_1^T\overline{\alpha}_{j,k},$$

for Alpert coefficient vectors $\overline{\alpha}_{j,k} = \overline{\alpha}_{j,k}(f)$.

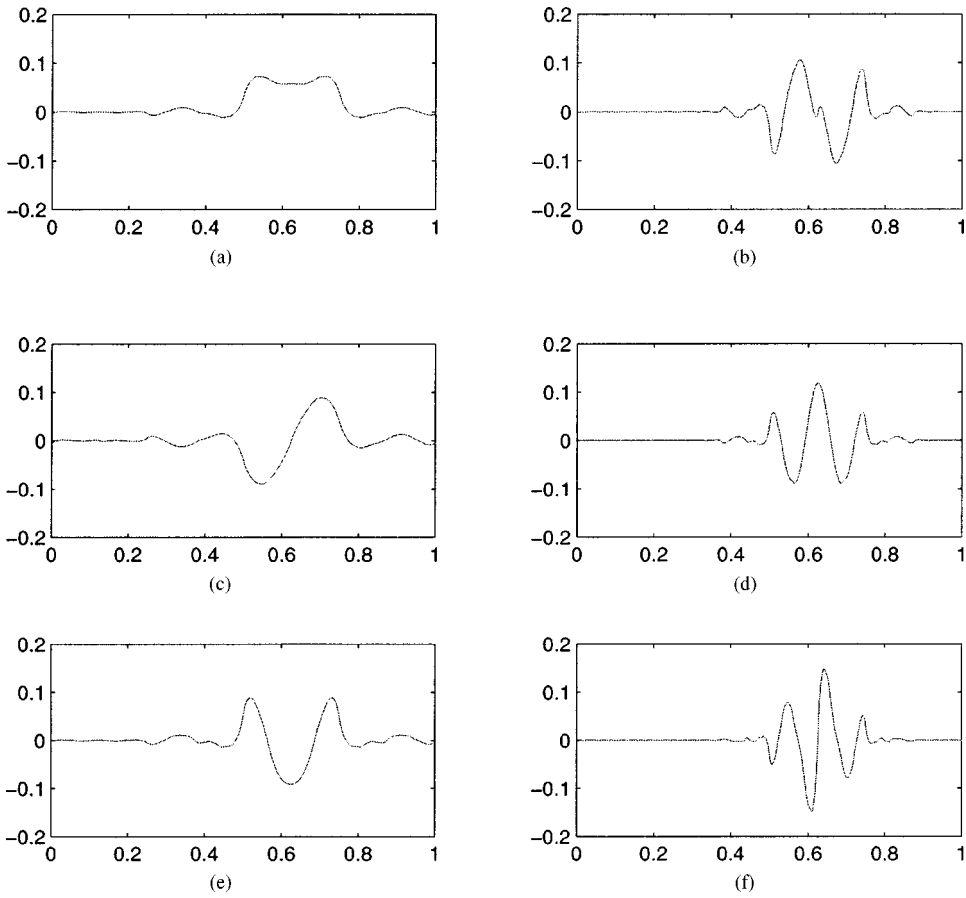



**FIG. 2.** Moment-interpolating multiwavelets $m = 3$, $L = 2$. (a) $\phi^0_{2,2}$, (b) $\psi^0_{2,2}$, (c) $\phi^1_{2,2}$, (d) $\psi^1_{2,2}$, (e) $\phi^2_{2,2}$, (f) $\psi^2_{2,2}$.

Combining these remarks: if $f \in W_j$, we have

$$\begin{aligned} f = \sum_k \mu^T_{j+1,k}\phi_{j+1,k} &= \sum_k \mu^T_{j+1,2k}\phi_{j+1,2k} + \mu^T_{j+1,2k+1}\phi_{j+1,2k+1} \\ &= \sum_k (G_0^T\overline{\alpha}_{j,k})^T\phi_{j+1,2k} + (G_1^T\overline{\alpha}_{j,k})^T\phi_{j+1,2k+1} \\ &= \sum_k \overline{\alpha}^T_{j,k}\phi_{j,k}, \end{aligned}$$

and (5.7) follows.

For (5.8), it is enough to show that if $f \in \overline{W}_j$ then $\langle f, \psi_{j,k}\rangle = \langle f, h_{j,k}\rangle$. We note that for an $f \in \overline{V}_{j+1}$, $\langle f, \phi_{j+1,k}\rangle = \langle f, \chi_{j+1,k}\rangle$, and so

$$\begin{aligned} \langle f, \psi_{j,k}\rangle &= G_0\langle f, \phi_{j+1,2k}\rangle + G_1\langle f, \phi_{j+1,2k+1}\rangle \\ &= G_0\langle f, \chi_{j+1,2k}\rangle + G_1\langle f, \chi_{j+1,2k+1}\rangle = \langle f, h_{j,k}\rangle. \ \blacksquare \end{aligned}$$

*5.2. Coefficient Functionals*

From (5.7) we have for the projector $Q_j$ that

$$Q_j f = \sum_k \alpha_{j,k}(f)^T \psi_{j,k},$$

where the coefficient functional $\alpha_{j,k}(f) = \langle Q_j f, h_{j,k} \rangle$. We now get more precise information.

THEOREM 5.3. *We have the formulas*

$$\alpha_{j,k}(f) = \langle f - P_j f, h_{j,k} \rangle, \tag{5.9}$$

*and*

$$\alpha_{j,k}(f) = \langle f - \pi_{j,k}(f), h_{j,k} \rangle, \tag{5.10}$$

*where $\pi_{j,k}$ denotes the moment-interpolating polynomial provided by the refinement scheme when applied to the data $(\mu_{j,k})_k$ at scale $j$.*

*Proof.* By construction of $P_{j+1}$, $\langle P_{j+1} f, \chi_{j+1,k} \rangle = \langle f, \chi_{j+1,k} \rangle$ for all $k$. Hence, as $h^\ell_{j,k} \in \overline{V}_{j+1}$,

$$\langle P_{j+1} f, h_{j,k} \rangle = \langle f, h_{j,k} \rangle \qquad \forall k.$$

Hence

$$\langle f - P_j f, h_{j,k} \rangle = \langle P_{j+1} f - P_j f, h_{j,k} \rangle = \langle Q_j f, h_{j,k} \rangle \qquad \forall k$$

which is (5.9); (5.10) is just the fact that $P_j f$ and $\pi_{j,k}$ have, by construction, the same local moments at scale $j+1$: $\mu_{j+1,2k+h}(P_j f) = \mu_{j+1,2k+h}(\pi_{j,k}),\ h = 0, 1.$ ■

From these results we see that $\alpha_{j,k}(f) = 0$ when $f$ is a polynomial of degree $D$ and that the $\alpha_{j,k}$ can be viewed as the Alpert coefficients of a function from which the part predictable at scale $j$ has been subtracted. It follows that if $f$ is smooth, the coefficients $\alpha_{j,k}$ will typically be much smaller at fine scales than the ordinary Alpert coefficients $\overline{\alpha}_{j,k}$.

THEOREM 5.4. *Suppose $f \in C^\alpha(\mathbb{R})$. Suppose $D = (2L+1)m - 1$. Then*

$$|\alpha^\ell_{j,k}| \le \text{Const}(2^{-j})^{(\alpha+1/2)}$$

*for $0 < \alpha < (D+1)$.*

This should be contrasted with the corresponding fact for Alpert wavelets, where the same conclusion holds for $0 < \alpha < m$; as $(D+1) > m$ this is a stronger result.

*5.3. Pyramid Algorithm*

Given the fine-scale moments $(\mu_{j+1,k})_k$, the above formulas tell us how to calculate the coefficients at the coarser scale:

$$\mu_{j,k} = H_0 \mu_{j+1,2k} + H_1 \mu_{j+1,2k+1}, \tag{5.11}$$

$$\alpha_{j,k} = G_0(\mu_{j+1,2k} - \tilde{\mu}_{j+1,2k}) + G_1(\mu_{j+1,2k+1} - \tilde{\mu}_{j+1,2k+1}), \tag{5.12}$$

where the $\tilde{\mu}_{j+1,k}$ are *predictions* of the local moments at the fine scale based on moment interpolation from the coarse scale data. These are obtained by using the refinement operator:

$$(\tilde{\mu}_{j+1,k})_k = R((\mu_{j,k})_k).$$

To go from coarse to fine, we first use the coarse scale moments to get the predictions $(\tilde{\mu}_{j+1,k})_k$ and then add them back in:

$$\overline{\alpha}_{j,k} = \alpha_{j,k} + G_0\tilde{\mu}_{j+1,2k} + G_1\tilde{\mu}_{j+1,2k};$$

the resulting $\overline{\alpha}_{j,k}$ are the Alpert wavelet coefficients of the object and thus can be used to reconstruct the finer-scale moments via the Alpert relations

$$\mu_{j+1,2k} = H_0^T\mu_{j,k} + G_0^T\overline{\alpha}_{j,k}, \tag{5.13}$$

$$\mu_{j+1,2k+1} = H_1^T\mu_{j,k} + G_1^T\overline{\alpha}_{j,k}. \tag{5.14}$$

The full pyramid algorithms, both for analysis and reconstruction, follow from these two-scale relations.

### 5.4. *Dual Wavelet*

The dual wavelet is defined as the representer of the coefficient functional:

$$\alpha_{j,k}(f) = \langle f, \tilde{\psi}_{j,k}\rangle, \qquad k \in \mathbb{Z}.$$

Define, for $-L \le h \le L$,

$$C_{2h} = G_0M_h^eH_0 + G_1M_h^oH_0,$$
$$C_{2h+1} = G_0M_h^eH_1 + G_1M_h^oH_1.$$

We define the mother dual wavelet by:

$$\tilde{\psi} = h_{0,0} - \sum_{h=-2L}^{2L-1} C_{-h}\chi_{1,h}.$$

THEOREM 5.5. $\tilde{\psi}_{j,k} = 2^{j/2}\tilde{\psi}(2^jx - k)$ *satisfies*

$$\langle Q_jf, h_{j,k}\rangle = \langle f, \tilde{\psi}_{j,k}\rangle, \qquad k \in \mathbb{Z}.$$

*As a result,* $Q_jf = \sum_k\langle f, \tilde{\psi}_{j,k}\rangle^T\psi_{j,k}$. *Moreover, if we define* $\tilde{Q}_jf = \sum_k\langle f, \psi_{j,k}\rangle^T\tilde{\psi}_{j,k}$ *and set* $\tilde{W}_j = \mathrm{Span}(\tilde{\psi}_{j,k}^\ell : k \in \mathbb{Z},\ 0 \le \ell < m)$, *then*

$$\tilde{Q}_jf = f \qquad \text{for } f \in \tilde{W}_j.$$

*Proof.* Formula (5.12) shows that we can write

$$\alpha_{j,k} = \overline{\alpha}_{j,k} - (G_0 \tilde{\mu}_{j+1,2k} + G_1 \tilde{\mu}_{j+1,2k+1}).$$

On the other hand, from (2.6) and (3.4)

$$\tilde{\mu}_{j+1,2k} = \sum_{h=-L}^{L} M_h^e (H_0 \mu_{j+1,2k-2h} + H_1 \mu_{j+1,2k+1-2h})$$

and

$$\tilde{\mu}_{j+1,2k+1} = \sum_{h=-L}^{L} M_h^o (H_0 \mu_{j+1,2k-2h} + H_1 \mu_{j+1,2k+1-2h}).$$

Collecting terms we get

$$\alpha_{j,k} = \overline{\alpha}_{j,k} - \sum_{h=-2L}^{2L-1} C_{-h} \mu_{j+1,2k-h}. \quad \blacksquare$$

*5.5. Decomposition Formulas*

We have the inhomogeneous multiwavelet expansion

$$f \sim \sum_{k=-\infty}^{\infty} \mu_{j0,k}^T \varphi_{j_0,k} + \sum_{j \geq j_0} \sum_{k=-\infty}^{\infty} \alpha_{j,k}^T \psi_{j,k}. \tag{5.15}$$

This can be motivated by the standard rewriting ideas. Let $f \in V_{j_1}$ and $j_0 < j_1$; then from $P_{j_1} f = P_{j_0} + Q_{j_0} f + \cdots + Q_{j_1 - 1} f$ we have immediately

$$f = \sum_k \mu_{j_0,k}^T \varphi_{j_0,k} + \sum_{j_0 \leq j < j_1} \sum_k \alpha_{j,k}^T \psi_{j,k}.$$

For general $f \in L^1(\mathbb{R})$, not in $V_{j_1}$, setting $f = P_{j_1} f + (f - P_{j_1} f)$ and letting $j_1$ tend to $\infty$, we get, for $f \in L^1(\mathbb{R})$ (since $P_j f \to f$ in $L^1$ as $j \to \infty$), the rule (5.15). A rigorous interpretation of the inhomogeneous expansion can be even broader. The MI coefficients make sense whenever $f$ is locally $L^1$, and so the inhomogeneous decomposition holds whenever $f$ is locally $L^1$, in the sense that a sequence of partial wavelet reconstructions

$$f_{J,K} = \sum_{|k| \leq 2^{j_0} K} \mu_{j0,k}^T \varphi_{j_0,k} + \sum_{j_0 \leq j \leq j_0 + J} \sum_{|k| \leq 2^j K} \alpha_{j,k}^T \psi_{j,k}$$

converges to $f$ as $J, K \to \infty$ (in $L^1(-R, R)$ norm, for any $R > 0$).

We also have the homogeneous expansion

$$f \sim \sum_{j=-\infty} \sum_{k=-\infty} \alpha_{j,k} \psi_{j,k}. \tag{5.16}$$

A rigorous interpretation of the homogeneous expansion is that, if $f \in L^1(\mathbb{R})$,

$$f_{J,K} = \sum_{-J \le j \le J} \sum_{|k| \le 2^j K} \alpha_{j,k} \psi_{j,k}$$

converges in $L^1(\mathbb{R})$ to $f$ as $J, K \to \infty$; this is a consequence of the earlier argument and of $P_{j_0} f \to 0$ in $L^1(\mathbb{R})$ as $j_0 \to -\infty$.

## 6. BASES FOR FUNCTIONAL SPACES

We may derive stronger results about the convergence of multiwavelet expansions by exploiting known properties of wavelet-like systems: systems in which the individual elements are scalar functions with wavelet-like characteristics but are not all dilations and translations of a single element. Such systems have been called *vaguelettes* by Y. Meyer and families of *molecules* by Frazier and Jawerth.

Such ideas give insights about multiwavelets by a process of vector-to-scalar interleaving. Starting from an MI multiwavelet system $(\psi^\ell_{j,k})_{j,k,\ell}$ constructed for a fixed $m$, $L$ combination, we may define a system of scalar functions $(v_{j,k})$ by the rule

$$v_{j,km+\ell}(t) = \sqrt{m} \cdot \psi^\ell_{j,k}(t/m), \qquad j,k \in \mathbb{Z},\ 0 \le \ell < m. \tag{6.1}$$

Similarly, starting from the dual multiwavelets $(\tilde{\psi}^\ell_{j,k})_{j,k,\ell}$, we can define

$$\tilde{v}_{j,km+\ell}(t) = \sqrt{m} \cdot \tilde{\psi}^\ell_{j,k}(t/m), \qquad j,k \in \mathbb{Z},\ 0 \le \ell < m. \tag{6.2}$$

These two systems are indexed like traditional wavelet systems, but they are not dilates and translates of a single fixed element. In the technical report on which this article is based, we use the concept of molecules [17, 18] to study these systems and obtain implications about MI multiwavelet expansions. We briefly state the conclusions, referring readers to [12] for more detail.

Frazier *et al.* [18] give a criterion for a family of scalar functions $(v_{j,k})$ to be a family of molecules for a Besov space $\dot{B}^\sigma_{p,q}$. In the case where $\sigma > 1/p - 1$, where $\psi$ is of compact support, and where the $(v_{j,k})$ derives from the vector-to-scalar interleaving (6.1), the criterion demands that the original multiwavelet mother $\psi^\ell \in C^\alpha$ for some $\alpha > \sigma$. The MI multiwavelet mother satisfies this hypothesis when the regularity $\mathrm{Reg}(m, L)$ is sufficiently large, and so *the MI multiwavelets furnish a family of molecules for* $\dot{B}^\sigma_{p,q}$, *provided* $\mathrm{Reg}(m, L) > \sigma > 1/p - 1$.

Frazier *et al.* [18] also give a criterion for the system of scalar functions $(\tilde{v}_{j,k})$ to be a norming family for a Besov space $\dot{B}^\sigma_{p,q}$. In the case where $\sigma > 1/p$, $\tilde{\psi}$ is of compact support, and $(\tilde{v}_{j,k})$ derives from vector-to-scalar interleaving (6.2), the criterion demands that each $\tilde{\psi}^\ell$ be bounded, with the vanishing moments property for polynomials of degree $D > \lfloor \sigma \rfloor$. Now $D = m(2L + 1) - 1$. The dual's mother MI multiwavelets satisfy this hypothesis for all $L$ sufficiently large, and so *duals of MI multiwavelets give a norming family for* $\dot{B}^\sigma_{p,q}$, *provided* $m(2L+1) - 1 > \sigma > 1/p - 1$.

Completely parallel results hold for homogeneous Triebel spaces $\dot{F}^{\sigma}_{p,q}$.

Results in [18] about molecules and norming families directly yield several implications about MI multiwavelet expansions; for details, see [12]. To begin with, the MI multiwavelet systems give Riesz bases for $L^2(\mathbb{R})$.

THEOREM 6.1. *Let* $\mathrm{Reg}(m, L) > 0$ *and* $D = m(2L+1) - 1 > 0$. *Then*

(1) $(\psi^{\ell}_{j,k})_{j,k,\ell}$ *and* $(\tilde{\psi}^{\ell}_{j,k})_{j,k,\ell}$ *are Riesz bases of* $L^2(\mathbb{R})$;
(2) *for* $f \in L^2(\mathbb{R})$ *we have*

$$f = \sum_{j,k} \langle f, \psi_{j,k} \rangle^T \tilde{\psi}_{j,k}$$
$$f = \sum_{j,k} \langle f, \tilde{\psi}_{j,k} \rangle^T \psi_{j,k},$$

*with both expansions unconditionally convergent in* $L^2$.

In addition, there are norm equivalence results in large ranges of the Besov–Triebel scales.

THEOREM 6.2. *Define a norm on the MI coefficients*

$$\theta = ((\mu_{j_0,\cdot}), (\alpha_{j_0,\cdot}), (\alpha_{j_0+1,\cdot}), \ldots)$$

*by*

$$\|\theta\|_{f^{\sigma}_{p,q}} \equiv \|(\mu_{j0,k})_k\|_{\ell_p} + \left\| \left( \sum_j \sum_k (2^{j\sigma} |\alpha_{j,k}| \chi_{j,k})^q \right)^{1/q} \right\|_{L^p}. \tag{6.3}$$

*If* $\mathrm{Reg}(m, L) > \sigma > 1/\min(1, p, q) - 1$, $p, q \in (1, \infty)$, *this is an equivalent norm for Triebel space* $F^{\sigma}_{p,q}(\mathbb{R})$. *Let* $\mathrm{Reg}(m, L) > \sigma > 1/p - 1$, $p, q \in (0, \infty]$. *Define a norm on the MI coefficients* $\theta$ *by*

$$\|\theta\|_{b^{\sigma}_{p,q}} = \|(\mu_{j_0,\cdot})\|_{\ell^p} + \left( \sum_{j \geq j_0} \left( 2^{js} \left( \sum_k |\alpha_{j,k}|^p \right)^{1/p} \right)^q \right)^{1/q}, \tag{6.4}$$

*with the calibration* $s \equiv \sigma + 1/2 - 1/p$. *This is an equivalent norm for the Besov space* $B^{\sigma}_{p,q}(\mathbb{R})$.

Such results, of course, imply results for the Sobolev and Hölder spaces, which are special cases of the Besov and Triebel scales.

COROLLARY 6.3. *If* $f \in B^{\sigma}_{p,q}$ *with* $\min(R, D) > \sigma > 1/p - 1$, *and* $0 < q < \infty$ *then the reconstruction of* $f$ *from its multiwavelet coefficients converges unconditionally to* $f$ *in the* $B^{\sigma}_{p,q}$ *norm — the order in which the individual terms are summed does not matter.*

As an example, consider the Sobolev space $W^1_2$, where $\sigma = 1, p = q = 2$. With MI wavelets of regularity $\mathrm{Reg}(m, L) > 2$, take an $f \in W^1_2$ and consider the partial reconstructions $f_\epsilon$ formed by setting to zero all MIMW coefficients of $f$ with norm smaller than the threshold $\epsilon$. Then $f_\epsilon \to f$ in $W^1_2$ norm as $\epsilon \to 0$.

## 7. EXPANSIONS ON THE INTERVAL

So far we have focused on the use of wavelet expansions on the line, but everything we have done works also for expansions on the interval $[0, 1]$. First, it is obvious that the Alpert system itself adapts very well to life on the interval; one simply restricts attention to Alpert functions and Alpert wavelets supported in the interval. Second, to construct the MI scaling functions and wavelets, everything is straightforward once one develops a refinement scheme adapted to life on the interval. This is relatively easy to do; conceptually, at the heart of the interval, the refinement scheme we have been using so far calls only for knowledge of local moments at intervals interior to the interval, and so the previous scheme can be used as is for such cases. At the edges of the interval, situations arise where the rule used on the line would refer to local moments of dyadic intervals not interior to $[0, 1]$; in such cases, one uses instead the $(2L + 1)$ closest interior intervals to the interval in question. The only restriction with such a rule is that one has to pick an expansion based on a coarsest scale $j_0$ obeying $2L + 1 < 2^{j_0}$.

The details of adapting to life on the interval have already been worked out carefully in the case $m = 1$ — see [11]. This shows that various functional space characterization properties for function spaces on the interval go through exactly as one might hope. We refer the reader to that article for details.

In the applications below we refer always to expansions adapted to life on the interval.

## 8. RECURSIVE DYADIC PARTITIONING

In this section we describe an interesting but apparently little known application of Alpert bases to fast recursive partitioning algorithms. In the next section we apply MI multiwavelets to obtain smooth recursive partitioning algorithms.

### *8.1. Recursive Partitioning*

A recursive dyadic partition (RDP) of $[0, 1)$ is any partition reachable by the following two rules.

1. The trivial partition $\mathcal{P} = \{[0, 1)\}$ is a recursive dyadic partition.
2. If $\mathcal{P} = \{[a_1, b_1), [a_2, b_2), \ldots, [a_k, b_k)\}$ is an RDP with $k$ pieces, then we get a new RDP $\mathcal{P}'$ obtained by splitting one piece, say the $i$th one, exactly in half:

$$\mathcal{P}' = \{[a_1, b_1), \ldots, [a_{i-1}, b_{i-1}), [a_i, (a_i + b_i)/2), [(a_i + b_i)/2, b_i), [a_{i+1}, b_{i+1}), \ldots\}.$$

Such partitions are naturally associated with binary trees. One can think of producing a given partition $\mathcal{P}$ in a stepwise fashion starting from a root which corresponds to $\mathcal{P}_0$; each enlargement is a sprouting of a new branch of the tree.

RDPs can conceivably subdivide the interval very finely in some parts of $[0, 1)$ and very coarsely in other parts; all that is required to reach such a partition is to apply the splitting rule more frequently to intervals covering certain locations and less frequently to intervals associated with other locations.

RDPs are naturally adapted to piecewise polynomial approximation. Given a function $f$, we denote by $E_m\{f|P\}$ the piecewise polynomial approximation $\tilde{f}$ to $f$ which is of the

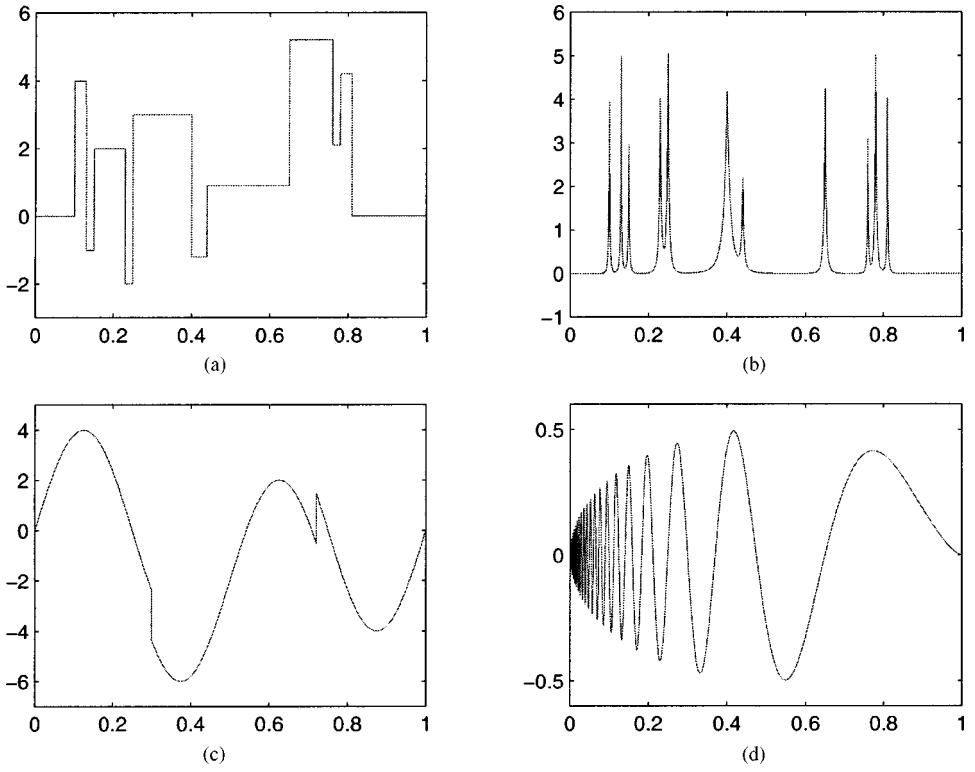



**FIG. 3.** Four objects: (a) Blocks; (b) Bumps; (c) HeaviSine; (d) Doppler.

form $\tilde{f} = \sum_{I \in P} p_I(x) \mathbb{1}_I(x)$, with each $p_I(x)$ a polynomial of degree $< m$, determined by least squares:

$$p_I = \underset{p \in \Pi_m}{\operatorname{argmin}} \|f - p\|_{L^2(I)}.$$

If the partition has small pieces where the function $f$ is very complicated, and large pieces where the function $f$ is very banal, this approach can maintain good accuracy while using relatively few coefficients.

Consider now the problem of *optimal recursive partitioning:* searching for a partition $\tilde{P}_N^*$ solving

$$\min \|f - E_m\{f|P\}\|_{L^2[0,1]} \text{ subject to } \#P \le N, \qquad P \text{ an RDP.} \tag{8.1}$$

This problem is attractive because it can be used to explain the idea of spatial adaptivity in lay terms. If the optimal partition is composed of pieces which are all roughly the same size, then the underlying object is spatially homogeneous: it requires approximation which is of the same scale everywhere. On the other hand, if the best partition contains blocks of very different sizes, this is an indication that the object contains some regions of much more rapid change than others.

To illustrate these points, we give computational examples. Figure 3 depicts four spatially inhomogeneous objects. Figure 4 shows optimal approximants with 64 pieces. The

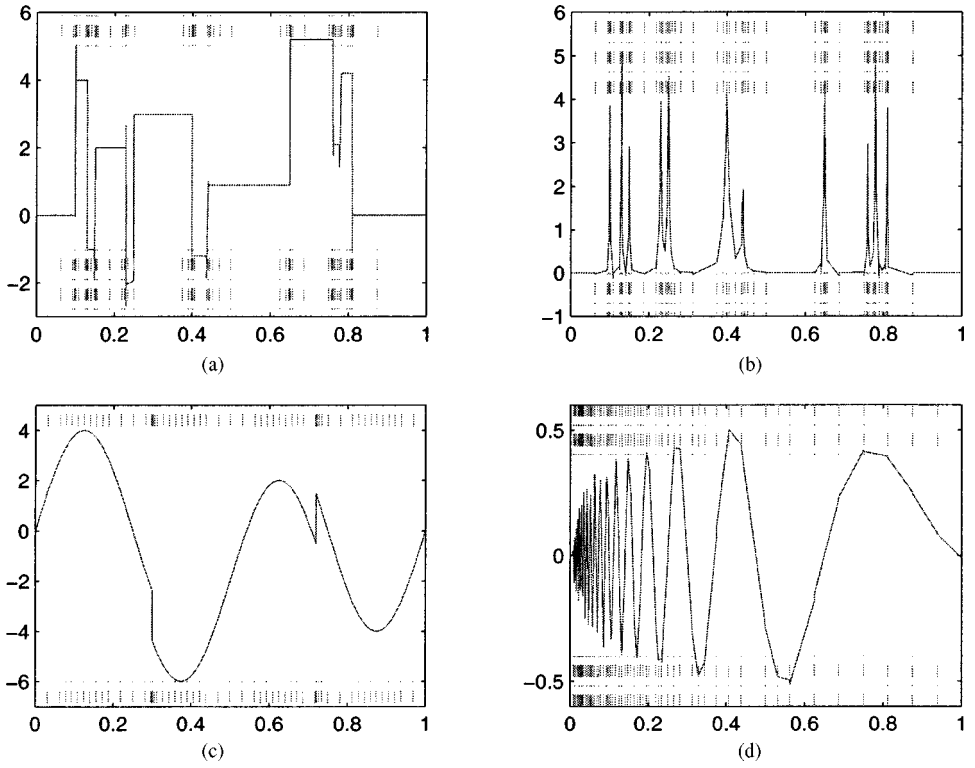



**FIG. 4.** Approximation by RDPs with 64 pieces. Tick marks denote boundaries of partition elements. (a) Blocks, MSE = 0.037835; (b) Bumps, MSE = 0.100508; (c) Bumps, MSE = 0.002377; (d) Doppler, MSE = 0.043558.

optimal partition responds visibly to the presence of the singularities and inhomogeneities in the underlying objects. Figure 5 shows spatially homogeneous approximation; note the far worse approximation in the vicinity of singularities, especially near $t = 0$ with the Doppler object.

### *8.2. Connection with Alpert Basis*

Optimal recursive partitioning is intimately connected with the Alpert basis.

THEOREM 8.1. *For each RDP $P$, the operator $E_m\{f|P\}$ is diagonal in the Alpert basis. In fact, let $\mathcal{I}(P)$ denote the collection of all dyadic intervals contained in* $[0, 1)$ *which contain* strictly *some interval $I \in P$, with the convention that $\mathcal{I}(P) \in \emptyset$ if $P$ is the trivial partition $P_0 = \{[0, 1)\}$. Then*

$$E_m\{f|P\} = \mu_{0,0}^T \chi_{0,0} + \sum_{I_{j,k} \in \mathcal{I}(P)} \overline{\alpha}_{j,k}^T h_{j,k}. \tag{8.2}$$

*Also,*

$$\|f - E_m\{f|P\}\|_{L^2[0,1)}^2 = \sum_{I_{j,k} \notin \mathcal{I}(P)} \|\overline{\alpha}_{j,k}\|^2. \tag{8.3}$$

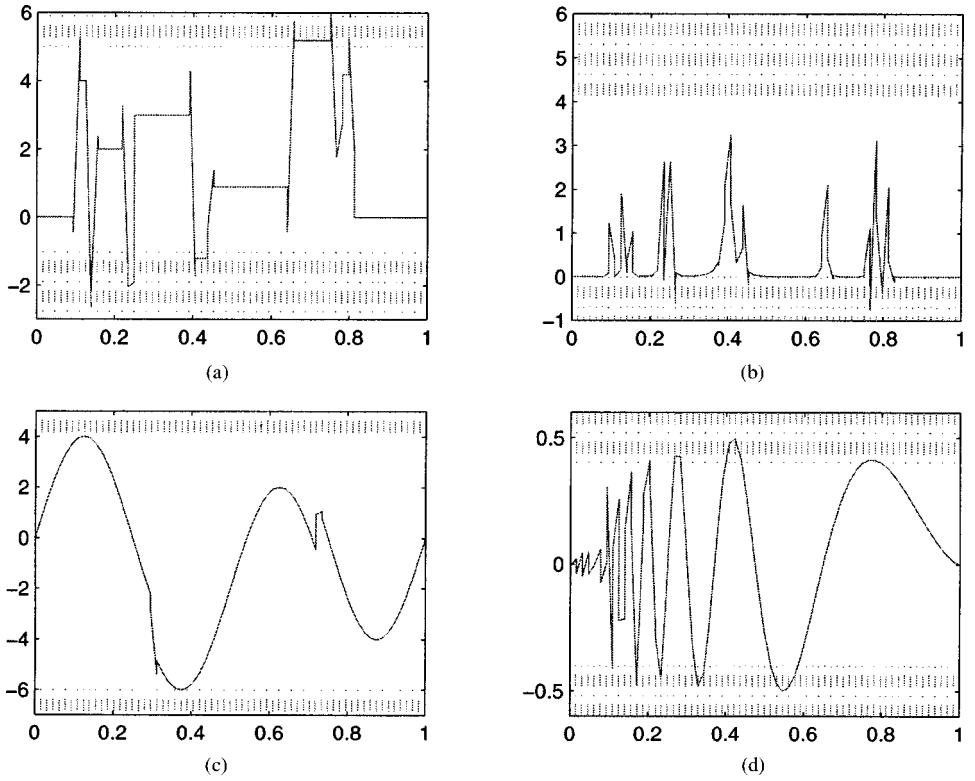



**FIG. 5.** Spatially homogeneous approximations with 64 pieces. (a) Blocks, MSE = 0.160728; (b) Bumps, MSE = 0.445318; (c) HeaviSine, MSE = 0.030392; (d) Doppler, MSE = 0.175505.

*Proof.* Consider a dyadic interval $I_{j,k} \subset I$ for an interval $I \in P$. Then for $\tilde{f} = E_m\{f|P\}$, we necessarily have $\overline{\alpha}_{j,k}(\tilde{f}) = 0$. Indeed, as the restriction of $\tilde{f}$ to $I$ is a polynomial $p_I$ of degree $< m$,

$$\overline{\alpha}_{j,k}(\tilde{f}) = \overline{\alpha}_{j,k}(p_I) = \langle p_I, h_{j,k} \rangle = 0.$$

The set $\mathcal{N}(P)$ of $\tilde{f}$ satisfying $\overline{\alpha}_{j,k} = 0$ whenever $I_{j,k} \subset I \in P$ has dimension $Nm$. The operator $E_m\{f|P\}$ has a range $R(P)$ with dimension $Nm$, where $N = \#P$. As $R(P) \subset N(P)$ and $\dim(R(P)) = \dim(N(P))$, we conclude that $R(P) = N(P)$. As the operator $E_m\{f|P\}$ is an orthoprojector, it acts as the identity on its range. The range contains (one can check) the Alpert functions $\chi^{\ell}_{0,0}$ and the Alpert wavelets $h^{\ell}_{j,k}$ where $I_{j,k} \subset \mathcal{I}(P)$; there are $Nm$ of these, they are orthonormal and so they span $R(P)$; hence (8.2) follows and the operator is diagonal in the Alpert basis. The identity (8.3) follows immediately by the orthogonality of the Alpert basis. ■

Say that a collection $\mathcal{I}$ of indices $I$ forms a *tree pattern* if, whenever $I \in \mathcal{I}$, every dyadic interval $I'$ containing $I$ also belongs to $\mathcal{I}$. The set $\mathcal{I}(P)$ defined in Theorem 8.1 forms a tree pattern. Hence the foregoing result has another formulation. *Every linear projection operator which is diagonal in the Alpert basis, taking the form*

$$P_{\mathcal{I}} f = \mu^T_{0,0} \chi_{0,0} + \sum_{I \in \mathcal{I}} \alpha^T_I h_I,$$

*where the indices in $\mathcal{I}$ form a tree pattern, is a projection operator $E_m\{f|P\}$ for some recursive dyadic $P$.*

### *8.3. Fast Algorithm for Partitioning*

A modified form of the optimal partitioning problem defining $P_N^*$ can be rapidly calculated using properties of the Alpert coefficients. The problem is modified in two ways.

- *Restricted search depth*. Let $\mathcal{P}_J$ denote the collection of RDPs containing only intervals wider than $2^{-J}$; we will restrict the search to partitions in $\mathcal{P}_J$ for a fixed $J$. (Note: the optimal $N$-element partition is always in $\mathcal{P}_{N-1}$, so the search depth restriction is not serious if $J$ is large enough.)
- *Lagrangian (penalized) form*. We put the problem in complexity penalized form, seeking to optimize $\|f - E_m\{f|P\}\|_2^2 + \lambda \# P$. For an appropriate $\lambda = \lambda(f, N)$ the solution partition will have $N$ elements, so we can obtain in this way an unconstrained problem with the same solution as the complexity-constrained problem.

In short our proposal is to solve the problem

$$\tilde{P}_{J,\lambda}^* = \operatorname*{argmin}_{P \in \mathcal{P}_J} \|f - E_m\{f|P\}\|_2^2 + \lambda \# P.$$

As will be seen, this form of the problem is computationally very tractable.

We first make some remarks about this problem.

- The complexity penalty parameter $\lambda$ is to some extent like a threshold. Indeed, an interval $I$ is in the optimal partition only if the Alpert coefficient $\|\alpha_I\|_2 < \sqrt{\lambda}$.
- This gives guidance for how large to choose $J$. If $J$ is so large that every wavelet coefficient at finer scales is smaller than $\lambda$, then $\overline{P}_{J,\lambda}^*$ is the same as $P_{J+h,\lambda}^*$ for every $h > 0$.
- In particular, suppose that $f \in B_{p,q}^{\sigma}(C)$. Then each Alpert multicoefficient obeys $\|\alpha_{j,k}(f)\|_2 \le A' \cdot C \cdot 2^{-j(\sigma+1/2-1/p)}$. Equating $\lambda = (A' \cdot C \cdot 2^{-j(\sigma+1/2-1/p)})^2$ we get that any $J > \log_2(\lambda^{-1/2}/A' \cdot C)/(\sigma + 1/2 - 1/p)$ is large enough so that $\overline{P}_{J,\lambda}^* = \overline{P}_{\infty,\lambda}^*$.
- One can find a value of $\lambda$ for which $\#\overline{P}_{J,\lambda}^* = N$ by a bisection search in $\lambda$.

ALGORITHM: PARTITION

*Description*.

Finds the partition which optimizes the complexity-penalized residual sum of squares.

*Inputs*.

$\overline{\alpha}_{j,k}$, $0 \le j < J$ — Alpert coefficients through level $J - 1$.

$\lambda$ — penalty factor.

*Results*.

$(c_{j,k})_{j,k}$ — value of subproblem associated with $j, k$.

$(\beta_{j,k})_{j,k}$ — indicator of membership of $I_{j,k}$ in optimal partition.

*Internal Data Structures*.

$(e_{j,k})_{j,k}$ — sum of squares of coefficients of descendants of $j, k$.

*Initialization.*

$$e_{j,k} = \|\overline{\alpha}_{j,k}\|_2^2, \qquad 0 \le j < J,\ 0 \le k < 2^j;$$
$$e_{J,k} = 0, \qquad 0 \le k < 2^J;$$
$$\beta_{j,k} = 1, \qquad 0 \le j < J;$$
$$c_{j,k} = 0 \qquad \forall j, k.$$

*Algorithm*

For $j = J-1, J-2, \ldots, 0$
  For $k = 0, \ldots, 2^j - 1$
    /* Compare Parent with Children */
    $e_{j,k} := e_{j,k} + e_{j+1,2k} + e_{j+1,2k+1}$
    if $e_{j,k} > \lambda + c_{j+1,2k} + c_{j+1,2k+1}$,
      $c_{j,k} := \lambda + c_{j+1,2k} + c_{j+1,2k+1}$
    else
      $\beta_{j,k} := 0$
      $c_{j,k} := e_{j,k}$
    end if
  end for
end for

The algorithm takes $O(N)$ operations to complete, and it gives the optimal partition because it uses the basic backwards induction process from dynamic programming.

### *8.4. Smooth Recursive Partitioning*

Given an RDP $P$, the approximant $E_m\{f|\mathcal{P}\}$ will, in general, be discontinuous. In this section we give two responses to this problem.

First, we develop a method for smoothing out a piecewise polynomial fit subordinate to an RDP; the method preserves the accuracy of the piecewise polynomial fit, while avoiding the discontinuities at boundaries of partition elements. We solve the problem of rapidly finding a smooth reconstruction matching the moments of the polynomial on each piece of the partition. MI multiwavelets are ideally suited to this.

Second, we develop a method for obtaining smooth approximants with a higher approximation order than piecewise polynomials. The method, which we call the method of Adapted Constraining Partitions, is in some sense dual to traditional recursive partitioning. Rather than demand that the reconstruction be made of atoms supported on pieces of the partition, we demand that the reconstruction match moments; the partition is adaptively constructed to get a "best" set of intervals for matching those moments.

In this section, if $I$ denotes one of the dyadic intervals $I_{j,k}$, then $\mu_I$ denotes the corresponding moment $\mu_{j,k}$, and similarly for $\alpha_I$.

*8.4.1. Moment interpolation on an RDP.* Given an RDP $P$ and Alpert moments

$$\mu_I = \langle \chi_I, f \rangle, \qquad I \in P,$$

we are interested in obtaining a smooth function $g$ obeying local moment conditions:

$$\langle \chi_I, g \rangle = \mu_I, \qquad I \in P.$$

If the partition is spatially homogeneous — all intervals the same length — this reduces to the problem defining MI refinement. If the partition is spatially inhomogeneous — involving finer scale discretization in certain regions than in others — we can apply MI refinement in a multiscale fashion to solve this problem. We now show how.

For simplicity, we assume that the partition $P$ is finite and therefore has a coarsest scale $2^{-j_0} = \max\{|I| : I \in P\}$ and a finest scale $2^{-j_1} = \min\{|I| : I \in P\}$. Let $\mathcal{I} = \mathcal{I}(P)$ be the collection of all intervals in $P$ and of all ancestors of intervals in $P$. Thus $\mathcal{I}$ is a tree-structured set — in principle, every local moment $\langle \chi_I, f \rangle$ where $I \in \mathcal{I}$ is known, i.e., derivable from the known information.

We will need to record which intervals at scale $j$ belong to $\mathcal{I}$ and which do not. At a given scale $j$, let $K_j^+$ be the collection of indices $k$, $0 \le k < 2^j$, such that $I_{j+1,2k}$ and $I_{j+2,2k+1}$ both belong to $\mathcal{I}$. Let $K_j^-$ be the complementary collection.

The algorithm has these steps:

S1. *Fine-To-Coarse:*

Procedure to derive moments $\mu_I$ for all intervals ancestors to $I$ for which local moments are given, i.e., for $I$ in $\mathcal{I}(P)$.

- Set $K_{j_1}^+ = \emptyset$.
- For each $j$ in the range $j_1, j_1 - 1, \ldots, j_0 - 1$,
  - Let $L_j = k$-indices of all intervals in $P$ of length $2^{-j}$.
  - Set $K_{j-1}^+ =$ parents of intervals in $L_j$ or of intervals in $K_j^+$.
  - Set $\mu_{j-1,k} = H_0 \mu_{j+1,2k} + H_1 \mu_{j+1,2k+1}$ for $k \in K_{j-1}^+$.

End.

S2. *Coarse-To-Fine:*

Procedure to imputed moments $b_I$ for all dyadic $I$ for which local moments were not given, at all scales $j_0 \le j \le J$.

- Set $\tilde{\mu}_{j_0,k} = \mu_{j_0,k}$, $0 \le k < 2^{j_0}$.
- For each $j$ in the range $j_0, \ldots, j_1 - 1$,
  - Set $b_{j,k} = \mu_{j,k}$ for $k \in K_j^+$.
  - Set $b_{j,k} = \tilde{\mu}_{j,k}$ for $k \in K_j^-$.
  - Predict the moments $(\tilde{\mu}_{j+1,k})_k = R((b_{j,k})_k)$.

End.

S3. Convert the pyramid of Alpert moments $b_{j,k}$ to an MI expansion.

S4. Reconstruct a functions from the MI expansion.

This algorithm allows one to quickly obtain a smooth function obeying the local moment constraints. The whole algorithm takes at most order $n$ time to compute for an ultimate resolution $n^{-1} = 2^{-J}$. The algorithm implicitly defines an operator taking $f$ into its reconstruction based only on local moments associated with $P$; below we will call the implicit operator $I_{m,L}\{f|P\}$.

If we apply $I_{m,L}\{f|P\}$ to the partition $P = \overline{P}_N^*$ solving (8.1), we get a smooth recursive partitioning.

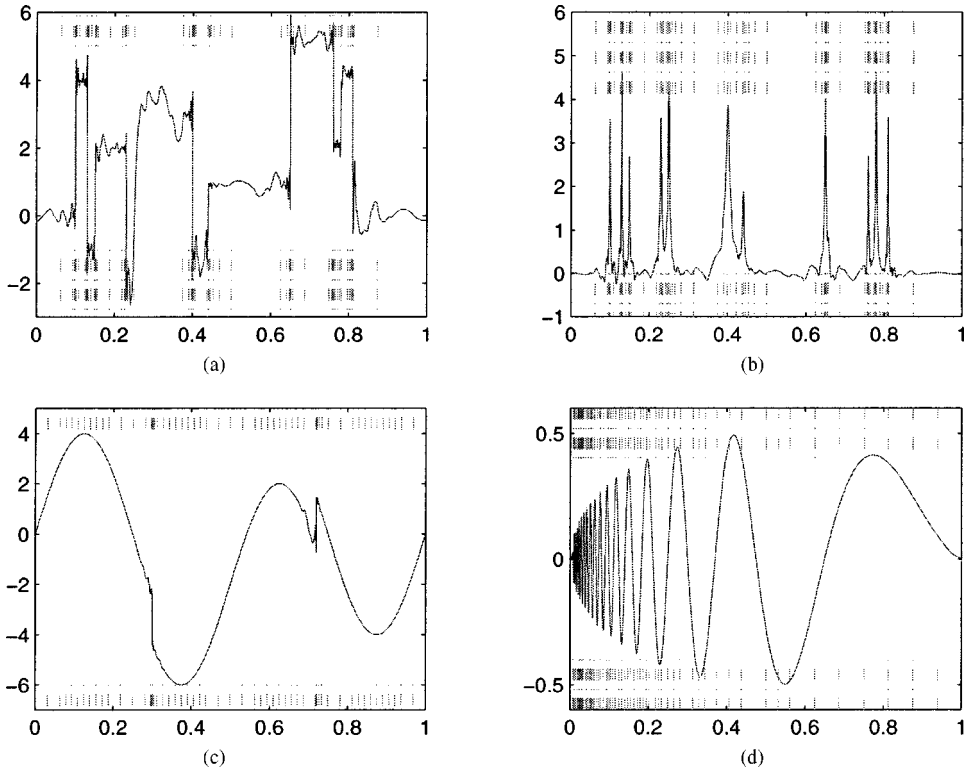



**FIG. 6.** Smooth reconstructions matching local moments to the piecewise polynomial RDP approximants of Fig. 2. (a) Blocks, MSE = 0.135963; (b) Bumps, MSE = 0.125741; (c) HeaviSine, MSE = 0.016023; (d) Doppler, MSE = 0.017537.

Figure 6 shows the results of this procedure on the same set of examples as in Fig. 4. In principle, the approximation error is equivalent to the approximation error of piecewise polynomial fit, while avoiding discontinuities. Unfortunately, for the Blocks and Heavisine signals the smoothing step generates significant overshoots and undershoots in the vicinity of discontinuities.

*8.4.2. Algorithm for adapted constraining partitions.* Combining ideas of the last few sections, we can obtain a fast partitioning algorithm combining high order accuracy and smoothness. The pieces of the partition constrain the local moments of the approximation; we adaptively choose the partition to get the most powerful local constraints. The problem of adapted constraining partition is in principle the problem of solving

$$\min \| f - I_{m,L}\{f|P\}\|_{L^2[0,1]} \text{ subject to } \#P \le N, \qquad P \text{ an RDP.} \tag{8.4}$$

The fast algorithm we propose below replaces $\| f - I_{m,L}\{f|P\}\|^2$ by the norm $\sum_{I \notin P} \|\alpha_I\|^2$. It has these steps:

- Apply Algorithm PARTITION initialized using the $MI$ coefficients $\alpha$ in place of the Alpert coefficients $\overline{\alpha}$.
- Use the partition $P^*_{J,\lambda}$ so obtained as the constraining partition.
- Obtain $I_{m,L}\{f|P^*_{J,\lambda}\}$.

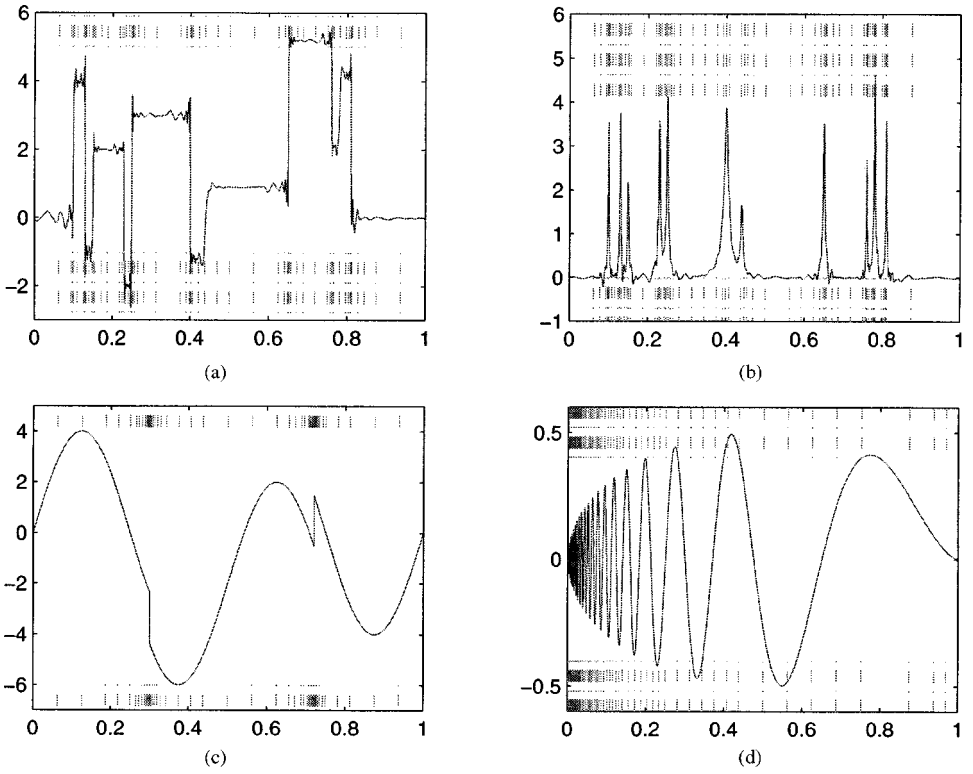



**FIG. 7.** Adapted constraining partitions with 64 pieces. (a) Blocks, MSE = 0.074696; (b) Bumps, MSE = 0.110900; (c) HeaviSine, MSE = 0.000487; (d) Doppler, MSE = 0.001740.

The algorithm is fast in the conventional sense that it requires order $O(n)$ operations to achieve an adapted reconstruction with ultimate resolution level $n$.

This approach is again attractive for explaining spatial adaptivity in lay terms. One expects that if the optimal constraining partition is composed of pieces which are all roughly the same size, then the underlying object is in some sense spatially homogeneous: it can be well approximated by obeying constraints of the same scale everywhere. On the other hand, if the best constraining partition contains blocks of very different sizes, this is an indication that the object contains some regions of much more rapid change than others.

Figure 7 shows that the adapted constraining partition method avoids the main problem suffered by the smoothed recursive partitioning algorithm of the last section, specifically the overshoot and undershoot that were evident in Figs. 6(a) and 6(c). Comparison of the MSE figures in the panel titles to Figs. 6 and 7 shows that the never method is substantially more accurate in these examples, for an equal degree of partition complexity; this follows from the very smooth nature of the objects in between singularities and the higher degree of approximation given by the MI system as compared to the comparable Alpert system.

Figure 8 gives a closer comparison of the performance of the following four approximation methods when applied to the Doppler signal: (i) piecewise polynomial approximation with a spatially homogeneous partition, (ii) piecewise polynomial approximation

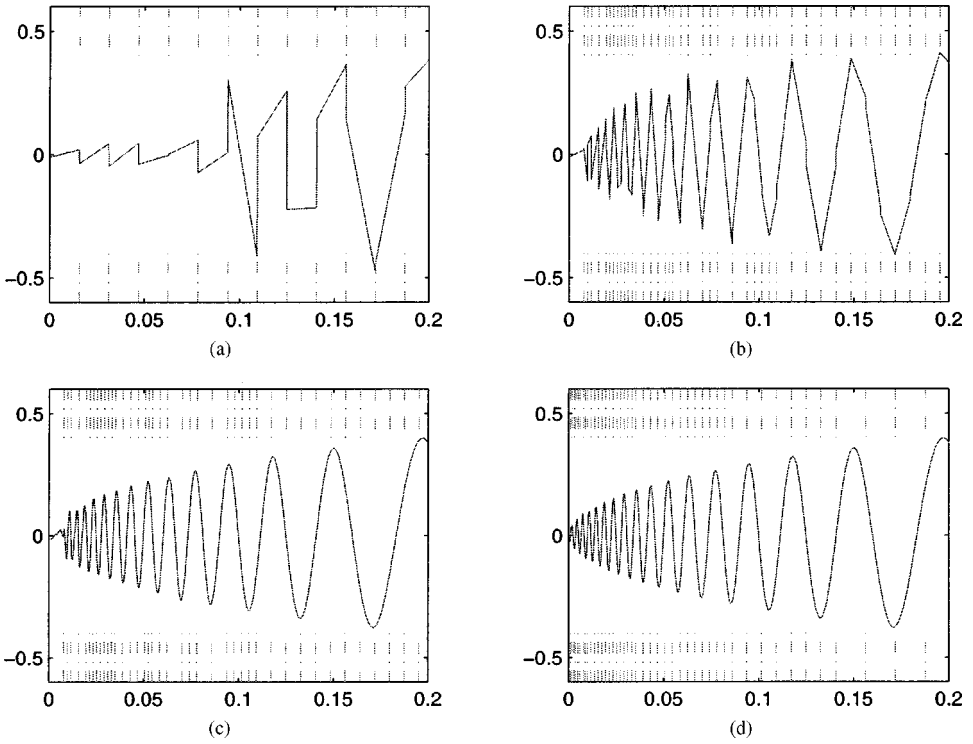



**FIG. 8.** (a) Spatially homogeneous piecewise polynomial fit. (MSE = 0.175505), (b) Piecewise polynomial fit with adapted partition (MSE = 0.043558), (c) Piecewise polynomial fit with adapted partition followed by multiwavelet smoothing (MSE = 0.017537), (d) Smooth adapted constraining partitions (MSE = 0.001740). 64 pieces are used in each case. The MSEs refer to the reconstruction errors for the whole Doppler signal, whereas the plots show only the reconstructed signals in the interval $[0, 0.2]$.

with an adapted partition (using algorithm of Section 8.3), (iii) a smoothing of (ii), produced by $I_{m,L}\{f|\overline{P}_N^*\}$ as described in Section 8.4.1, (iv) smooth adapted constraining partitions.

As expected,

- method (i) performs the worst;
- methods (ii) and (iii) perform comparably in terms of MSE and both methods outperform method (i);
- method (iii) gives smooth reconstruction, whereas method (ii) gives discontinuous reconstruction;
- method (iv) outperforms the first three methods — both in terms of MSE and smoothness.

## 9. SOFTWARE IMPLEMENTATION

All computational results reported in this paper are reproducible, meaning that the code which generated the figures and tables is available over the Internet, following the discipline indicated in [4], at `http://www-stat.stanford.edu/~wavelab`.

For each figure given here, it is possible to inspect the code which generated the figure and even to modify the examples and re-run them.

The system itself is a component of the `WaveLab` software library, release .800. The directory `Papers/Alpert` contains all the scripts necessary to reproduce the figures in this paper, and the directory `MultiWavelets` contains the basic tools for the multiwavelet transform. The directory `Refinement` contains the basic tools to estimate the smoothness of general scalar refinement schemes, as well as Hermite and MI refinement schemes.

## ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation NSF-DMS-95-05151 and the Air Force Office of Scientific Research MURI95-F49620-96-0028 and other sponsors. Both David Donoho and Thomas Yu would like to thank Nira Dyn and David Levin of the Department of Mathematics at Tel Aviv University for hospitality during their visits there in 1996 and 1997, respectively.

## REFERENCES

1. B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin, Wavelet-like bases for the fast solution of second-kind integral equations, *SIAM J. Math. Anal.* **14** (1993), 159–184.
2. B. K. Alpert, A class of bases in $L^2$ for sparse representation of integral operators, *SIAM J. Math. Anal.* **24** (1993), 246–262.
3. L. Breiman, J. Friedman, R. Olshen, and C. J. Stone, "Classification and Regression Trees," Wadsworth, Belmont, CA, 1983.
4. J. Buckheit and D. L. Donoho, Wavelab and reproducible research, *in* "Wavelets and Statistics," (A. Antoniadis and G. Oppenheim, Eds.), pp. 55–82, Springer-Verlag, New York, 1995.
5. A. Cohen, I. Daubechies, and J. Feauveau, Bi-orthogonal bases of compactly supported wavelets, *Comm. Pure Appl. Math.* **45** (1992), 485–560.
6. S. Conte and C. de Boor, "Elementary Numerical Analysis," 3rd ed., McGraw-Hill, New York, 1980.
7. I. Daubechies and J. Lagarias, Two-scale difference equations I. Existence and global regularity of solutions, *SIAM J. Math. Anal.* **22** (1991), 1388–1410.
8. I. Daubechies and J. Lagarias, Two-scale difference equations II. Local regularity, infinite products of matrices and fractals, *SIAM J. Math. Anal.* **23** (1992), 1031–1079.
9. C. de Boor, "A Practical Guide to Splines," Applied Mathematical Sciences, Vol. 27, Springer-Verlag, New York, 1978.
10. G. Deslauriers and S. Dubuc, Symmetric iterative interpolation processes, *Constr. Approx.* **5** (1989), 49–68.
11. D. L. Donoho, Smooth wavelet decompositions with blocky coefficient kernels, *in* "Recent Advances in Wavelet Analysis," (L. L. Schumaker and G. Webb, Eds.), pp. 259–308, Academic Press, Boston, 1993.
12. D. L. Donoho, N. Dyn, D. Levin, and T. P.-Y. Yu, Smooth multiwavelet duals of Alpert bases by moment-interpolating refinement, Technical report, Department of Statistics, Stanford University, 1998. Available at ftp://stat.stanford.edu/reports/donoho/moment.ps.Z.
13. S. Dubuc, Interpolation through an iterative scheme, *J. Math. Anal. Appl.* **114** (1986), 185–204.
14. N. Dyn, "Subdivision Schemes in Computer-Aided Geometric Design," pp. 36–104. Advances in Numerical Analysis II, Wavelets Subdivision Algorithms and Radial Basis Functions, Clarendon Press, Oxford, 1992.
15. N. Dyn, J. Gregory, and D. Levin, Analysis of uniform binary subdivision schemes for curve design, *Constr. Approx.* **7** (1991), 127–147.
16. N. Dyn and D. Levin, Analysis of hermite-interpolatory subdivision schemes, *in* "Spline Functions and the Theory of Wavelets," (S. Dubuc, Ed.), AMS Series CRM Proceedings and Lecture Notes, pp. 105–113, 1998.

17. M. Frazier and B. Jawerth, The $\varphi$-transform and applications to distribution spaces, *in* "Function Spaces and Applications," (M. Cwikel *et al*., Eds.), Lecture Notes in Mathematics, Vol. 1302, pp. 223–246, Springer-Verlag, New York/Berlin, 1988.
18. M. Frazier, B. Jawerth, and G. Weiss, "Littlewood–Paley Theory and the Study of Function Spaces," American Math. Soc., Providence, 1991.
19. A. Harten, Multiresolution representation of cell-averaged data, Technical report, UCLA, 1994.
20. J. L. Merrien, A family of Hermite interpolants by bisection algorithms, *Numer. Algorithms* **2** (1992), 187–200.
21. C. Micchelli and H. Prautzsch, Uniform refinement of curves, *Linear Algebra Appl.* **114**/**115** (1989), 841–870.
22. O. Rioul, Simple regularity criteria for subdivision schemes, *SIAM J. Math. Anal.* **23** (1992), 1544–1576.
23. J. M. Shapiro, Embedded image coding using zerotrees of wavelet coefficients, *IEEE Trans. Signal Process.* **41** (1993), 3445–3462.
24. T. P.-Y. Yu, "New Developments in Interpolating Wavelet Transforms," PhD thesis, Program of Scientific Computing and Computational Mathematics, Stanford University, August 1997.
25. T. P.-Y. Yu, Parametric families of Hermite subdivision schemes in dimension 1, Available at http://www.rpi.edu/yut/Papers/hermitel.pdf, 1999.