SIAM J. MATH. ANAL.
Vol. 31, No. 5, pp. 1030–1061

NONLINEAR PYRAMID TRANSFORMS BASED ON MEDIAN-INTERPOLATION*

DAVID L. DONOHO† AND THOMAS P.-Y. YU‡

Abstract. We introduce a nonlinear refinement subdivision scheme based on *median-interpolation*. The scheme constructs a polynomial interpolating adjacent block medians of an underlying object. The interpolating polynomial is then used to impute block medians at the next finer triadic scale. Perhaps surprisingly, expressions for the refinement operator can be obtained in closed-form for the scheme interpolating by polynomials of degree $D = 2$. Despite the nonlinearity of this scheme, convergence and regularity can be established using techniques reminiscent of those developed in analysis of linear refinement schemes.

The refinement scheme can be deployed in multiresolution fashion to construct a nonlinear pyramid and an associated forward and inverse transform. In this paper we discuss the basic properties of these transforms and their possible use in removing badly non-Gaussian noise. Analytic and computational results are presented to show that in the presence of highly non-Gaussian noise, the coefficients of the nonlinear transform have much better properties than traditional wavelet coefficients.

Key words. pyramid transform, subdivision scheme, wavelet, median, robust statistics, nonlinear analysis, interpolation

AMS subject classifications. 26A15, 26A16, 26A18, 26A27, 39B05, 41A05, 41A46, 42C40, 94A12, 62G05, 62G35

PII. S0036141097330294

1. Introduction.

Recent theoretical studies [14, 13] have found that the orthogonal wavelet transform offers a promising approach to noise removal. They assume that one has noisy samples of an underlying function f

$$y_i = f(t_i) + \sigma z_i,\ i = 1, \dots, n, \tag{1.1}$$

where $(z_i)_{i=1}^n$ is a standard Gaussian white noise and σ is the noise level. In this setting, they show that one removes noise successfully by applying a wavelet transform, thresholding the wavelet coefficients, and inverting the transform. Here "success" means near-asymptotic minimaxity over a broad range of classes of smooth f. Other efforts [20, 17] have shown that the Gaussian noise assumption can be relaxed slightly; in the presence of non-Gaussian noise that is not too heavy-tailed (e.g., the density has sufficiently rapid decay at $\pm\infty$), one can use level-dependent thresholds which are somewhat higher than in the Gaussian case and continue to obtain near-minimax results.

1.1. Strongly non-Gaussian noise.

In certain settings, data exhibit strongly non-Gaussian noise distributions; examples include analogue telephony [30], radar signal processing [1], and laser radar imaging [19]. By strongly non-Gaussian we mean subject to very substantial deviations much more frequently than under Gaussian assumptions.

*Received by the editors November 19, 1997; accepted for publication (in revised form) April 2, 1999; published electronically April 20, 2000. This research was partially supported by NSF DMS-95-05151, by AFOSR MURI95-F49620-96-1-0028, and by other sponsors.

http://www.siam.org/journals/sima/31-5/33029.html

†Department of Statistics, Stanford University, Stanford, CA 94305 (donoho@stat.stanford.edu).

‡Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180-3590 (yut@rpi.edu).

Thresholding of linear wavelet transforms *does not* work well with strongly non-Gaussian noise. Consider model (1.1) in a specific case: let (z_i) be independently and identically Cauchy distributed. The Cauchy distribution has no moments $\int x^\ell f(x)dx$ for $\ell = 1, 2, \ldots$, in particular neither mean nor variance.

Under this model, typical noise realizations $(z_i)_{i=1}^n$ contain a few astonishingly large observations: the largest observation is of size $O(n)$. (In comparison, for Gaussian noise, the largest observation is of size $O(\sqrt{\log(n)})$.) Moreover, a linear wavelet transform of independently and identically distributed (i.i.d.) Cauchy noise does not result in independent, nor identically distributed wavelet coefficients. In fact, coefficients at coarser scales are more likely to be affected by the perturbing influence of the few large noise values, and so one sees a systematically larger stochastic dispersion of coefficients at coarse scales. Invariance of distribution across scale and $O(\sqrt{\log(n)})$ behavior of maxima are fundamental to the results on wavelet denoising in [13, 14]. The Cauchy situation therefore lacks key quantitative properties which were used in denoising in the Gaussian case.

It is not just that this situation lacks properties which would make the proofs "go through." If we try to apply ideas which were successful under Gaussian theory we meet with abject failure, as simple computational examples given later will illustrate.

1.2. Median-interpolating pyramid transform. Motivated by this situation, this paper develops a kind of *nonlinear "wavelet transform."* The need to abandon linearity is clear a priori. It is well known that linearity of approach is essentially tantamount to a Gaussian assumption and that non-Gaussian assumptions typically lead to highly nonlinear approaches. For example, maximum likelihood estimators in classical statistical models are often linear in the Gaussian case, but highly nonlinear under Cauchy and similar assumptions.

Central to our approach is the notion of *median-interpolating* (MI) refinement scheme. Given data about the medians of an object on triadic blocks at a coarse scale, we predict the medians of triadic blocks at the next finer scale. We do this by finding a median-interpolating polynomial—a polynomial with the same coarse-scale block medians—and then calculating the block medians of this polynomial for blocks at the next finer scale. The procedure is nonlinear: the interpolating polynomial is a nonlinear functional of the coarse-scale medians; and the imputed finer-scale medians are nonlinear functionals of the interpolating polynomial. Perhaps surprisingly, in the case of interpolation by quadratic polynomials, the interpolating polynomial and its finer-scale medians can both be found in closed-form.

Using MI refinement, we can build forward and inverse transforms which can be computed rapidly and which exhibit favorable robustness and regularity properties.

The forward transform deploys the median in a multiresolution pyramid; it computes "block medians" over all triadic cells. MI refinement is used to predict medians of finer-scale blocks from coarser-scale blocks; the resulting prediction errors are recorded as transform coefficients.

The inverse transform undoes this process; using coarser-scale coefficients it builds MI predictions of finer-scale coefficients; adding in the prediction errors recorded in the transform array leads to exact reconstruction.

This way of building transforms from refinement schemes is similar to the way interpolating wavelet transforms are built from Deslauriers–Dubuc interpolating schemes in [9] and in the way biorthogonal wavelet transforms are built from average-interpolating refinement in [10]. The basic idea is to use data at coarser scales to predict data at finer scales and to record the prediction errors as coefficients asso-

ciated with the finer scales. Despite structural similarities our MI-based transforms exhibit important differences:

- Both the forward and inverse transforms can be nonlinear;
- The transforms are based on a triadic pyramid and a 3-to-1 decimation scheme;
- The transforms are expansive (they map n data into $\sim 3/2n$ coefficients).

Terminologically, because the forward transform is expansive, it should be called a *pyramid transform* rather than a wavelet transform. We call the transform itself the *median-interpolating pyramid transform* (MIPT).

The bulk of our paper is devoted to analysis establishing two key properties of these transforms.

- *Regularity.* Take block medians at a single level and refine to successively finer and finer levels using the quadratic polynomial MI scheme. Detailed analysis shows that the successive refinements converge uniformly to a continuous limit with Hölder-α regularity for some $\alpha > 0$. We prove that $\alpha > .0997$ and we give computational and analytical evidence pointing to $\alpha > 1 - \epsilon$ for all $\epsilon > 0$.

  This result shows that MIPT has important similarities to linear wavelet and pyramid transforms. For example, it provides a notion of nonlinear multiresolution analysis: just as in the linear MRA case, one can decompose an object into "resolution levels" and examine the contributions of different levels separately; each level contributes a regular curve oscillating with wavelength comparable to the given resolution, with large oscillations in the spatial vicinity of significant features.
- *Robustness.* It is well known that the median is robust against heavy-tailed noise distributions [21, 22]. In the present setting this phenomenon registers as follows. We are able to derive thresholds for noise removal in the MIPT which work well for *all* distributions in rather large classes, irrespective of the heaviness of the tails. In particular, we show that at all but the finest scales, the same thresholds work for both Gaussian and Cauchy data. Hence a noise-removal scheme based on thresholding of MIPT coefficients depends only very weakly on assumptions about noise distribution.

There is considerable applied interest in developing median-based multiscale transforms, as one can see from [2, 23, 28, 29, 26]. The analysis we give here suggests that our framework will turn out to have strong theoretical justification and may provide applied workers with helpful new tools.

1.3. Contents. Section 2 introduces the notion of median-interpolating refinement, shows how one of the simplest instances may be computed efficiently, gives computational examples, and proves some basic properties. Section 3 establishes convergence and smoothness results for the quadratic median-interpolating refinement scheme. Section 4 develops a nonlinear pyramid transform and describes properties of transform coefficients. Proofs of these properties are recorded in section 6. Section 5 applies the pyramid transform to the problem of removing highly non-Gaussian noise.

2. Median-interpolating refinement schemes. In this section we describe a notion of two-scale refinement which is nonlinear in general, and which yields an interesting analogue of the refinement schemes occurring in the theory of biorthogonal wavelets.

2.1. Median-interpolation. Given a function f on an interval I, let $\mathrm{med}(f|I)$ denote a median of f for the interval I, defined by

$$\mathrm{med}(f|I) = \inf\{\mu : m(t \in I : f(t) \geq \mu) \geq m(t \in I : f(t) \leq \mu)\}, \tag{2.1}$$

where $m()$ denotes Lebesgue measure on $\mathbb{R}$.

Now suppose we are given a triadic array $\{m_{j,k}\}_{k=0}^{3^j-1}$ of numbers representing the medians of f on the triadic intervals $I_{j,k} = [k3^{-j}, (k+1)3^{-j})$:

$$m_{j,k} = \mathrm{med}(f|I_{j,k}), \quad 0 \leq k < 3^j, \ \ j \geq 0.$$

The goal of median-interpolating refinement is to use the data at scale j to infer behavior at the finer scale $j+1$, obtaining imputed medians of f on intervals $I_{j+1,k}$. Obviously we are missing the information to impute perfectly; nevertheless we can try to do a reasonable job.

We employ *polynomial-imputation*. Starting from a fixed even integer D, it involves two steps.

[M1] **(Interpolation).** For each interval $I_{j,k}$, find a polynomial $\pi_{j,k}$ of degree $D = 2A$ satisfying the *median-interpolation condition*:

$$\mathrm{med}(\pi_{j,k}|I_{j,k+l}) = m_{j,k+l} \ \text{ for } \ -A \leq l \leq A. \tag{2.2}$$

[M2] **(Imputation).** Obtain (pseudo-) medians at the finer scale by setting

$$\tilde{m}_{j+1,3k+l} = \mathrm{med}(\pi_{j,k}|I_{j,3k+l}) \ \text{ for } \ l = 0, 1, 2. \tag{2.3}$$

An example is given in Figure 2.1 for degree $D = 2$. Some questions come up naturally:

[Q1] Is there a unique polynomial $\pi_{j,k}$ satisfying the nonlinear equations (2.2)?

[Q2] If so, is there an effective algorithm to find it?

[Q3] If so, what are the properties of such a procedure?

2.1.1. Average-interpolation. A scheme similar to the above, with "med" replaced by "ave," is relatively easy to study and provides useful background. Given a function f on an interval I, write $\mathrm{ave}(f|I) = |I|^{-1} \int_I f(t)dt$ for the average value of f over the interval I. Now suppose we are given a triadic array $\{a_{j,k}\}_{k=0}^{3^j-1}$ of numbers representing the averages of f on the triadic intervals $I_{j,k}$. Average-interpolating refinement uses the data at scale j to impute behavior at the finer scale $j+1$, obtaining the (pseudo-) averages of f on intervals $I_{j+1,k}$. Fix an even integer D, it runs as follows:

[A1] **(Interpolation).** For each interval $I_{j,k}$, find a polynomial $\pi_{j,k}$ of degree $D = 2A$ satisfying the *average-interpolation condition*:

$$\mathrm{ave}(\pi_{j,k}|I_{j,k+l}) = a_{j,k+l} \ \text{ for } \ -A \leq l \leq A. \tag{2.4}$$

[A2] **(Imputation).** Obtain (pseudo-) cell averages at the finer scale by setting

$$\overline{a}_{j+1,3k+l} = \mathrm{ave}(\pi_{j,k}|I_{j,3k+l}) \ \text{ for } l = 0, 1, 2. \tag{2.5}$$

This type of procedure has been implemented and studied in (the dyadic case) [10, 11]. The analogues of questions [Q1]–[Q2] have straightforward "Yes" answers. For any degree D one can find coefficients $c_{h,l}^{(D)}$ for which

$$\overline{a}_{j+1,3k+l} = \sum_{h=-A}^{A} c_{h,l}^{(D)} a_{j,k+h}, \quad l = 0, 1, 2, \tag{2.6}$$

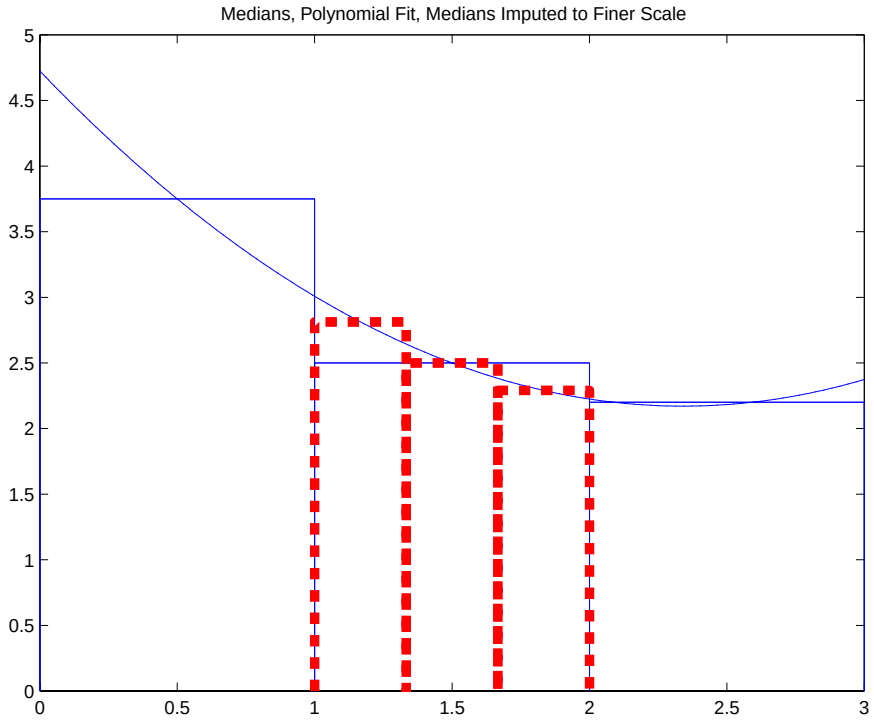

FIG. 2.1. *Median-interpolation, $D = 2$. The rectangular blocks surrounded by solid lines correspond to $m_{0,0}, m_{0,1}, m_{0,2}$, the blocks surrounded by thickened dashed lines correspond to imputed medians $\tilde{m}_{1,2}, \tilde{m}_{1,3}, \tilde{m}_{1,4}$, the parabola corresponds to the median-interpolant $\pi_{0,1}$.*

exhibiting the fine-scale imputed averages $\overline{a}_{j+1,k}$'s as linear functionals of the coarse-scale averages $a_{j,k}$. Moreover, using analytic tools developed in wavelet theory [4] and in refinement subdivision schemes [8, 16] one can establish various nice properties of refinement by average-interpolation—see below.

2.1.2. $D = 0$. We return to median-interpolation. The case $D = 0$ is the simplest by far; in that case one is fitting a constant function $\pi_{j,k}(t) = \text{Const}$. Hence $A = 0$, and (2.2) becomes $\pi_{j,k}(t) = m_{j,k}$. The imputation step (2.3) then yields $m_{j+1,3k+l} = m_{j,k}$ for $l = 0, 1, 2$. Hence refinement proceeds by imputing a constant behavior at finer scales.

2.1.3. $D = 2$. The next simplest case is $D = 2$ and will be the focus of attention in this article. To apply (2.2) with $A = 1$, we must find a quadratic polynomial solving

$$\text{med}(\pi_{j,k}|I_{j,k+l}) = m_{j,k+l} \quad \text{for } l = -1, 0, 1. \tag{2.7}$$

In general this is a system of nonlinear equations. One can ask [Q1]–[Q3] above for this system. The answers come by studying the operator $\Pi_{(2)} : \mathbb{R}^3 \to \mathbb{R}^3$ defined as the solution to the problem: given $[m_1, m_2, m_3]$, find $[a, b, c]$ such that the quadratic polynomial $\pi(x) = a + bx + cx^2$ satisfies

$$\text{med}(\pi|[0, 1]) = m_1, \tag{2.8}$$

$$\operatorname{med}(\pi|[1,2]) = m_2, \tag{2.9}$$

$$\operatorname{med}(\pi|[2,3]) = m_3. \tag{2.10}$$

In this section, we work out explicit algebraic formulae for $\Pi_{(2)}$. It will follow from these that (2.7) *has an unique solution, for every* m_1, m_2, m_3, and that this solution is a Lipschitz function of the m_i.

$\Pi_{(2)}$ possesses two purely formal invariance properties which are useful below.

- *Reversal equivariance.* If $\Pi_{(2)}(m_1, m_2, m_3) = a + bx + cx^2$, then $\Pi_{(2)}(m_3, m_2, m_1) = a + b(3-x) + c(3-x)^2$.
- *Affine equivariance.* If $\Pi_{(2)}(m_1, m_2, m_3) = \pi$, then $\Pi_{(2)}(a + bm_1, a + bm_2, a + bm_3) = a + b\pi$.

Reversal equivariance is, of course, tied to the fact that median-interpolation is a spatially symmetric operation. From affine equivariance, it follows that when $m_2 - m_1 \neq 0$ we have

$$\begin{aligned}\Pi_{(2)}(m_1, m_2, m_3) &= m_1 + \Pi_{(2)}(0, m_2 - m_1, m_3 - m_1) \\ &= m_1 + (m_2 - m_1)\Pi_{(2)}\left(0, 1, 1 + \frac{m_3 - m_2}{m_2 - m_1}\right).\end{aligned} \tag{2.11}$$

Thus $\Pi_{(2)}$ is characterized by its action on very special triples; it is enough to study the univariate function $\Pi_{(2)}(0, 1, 1+d)$, $d \in \mathbb{R}$. (The exceptional case when $m_2 - m_1 = 0$ can be handled easily; see the discussion after the proof of Proposition 2.2.)

To translate (2.8)–(2.10) into manageable algebraic equations, we begin with the following proposition.

PROPOSITION 2.1 (median-imputation, $D = 2$). *Suppose the quadratic polynomial* $\pi(x)$ *has its extremum at* x^*. *Let* $s = q - p$.

[L] *If* $x^* \notin [p + s/4, p + 3s/4]$, *then*

$$\operatorname{med}(\pi(x)|[p,q]) = \pi((p+q)/2). \tag{2.12}$$

[N] *If* $x^* \in [p + s/4, p + 3s/4]$, *then*

$$\operatorname{med}(\pi(x)|[p,q]) = \pi(x^* \pm s/4). \tag{2.13}$$

Proof. We assume x^* is a minimizer (the case of a maximizer being similar). The key fact is that $\pi(x)$, being a quadratic polynomial, is symmetric about x^* and monotone increasing in $|x - x^*|$.

If $x^* \in [p + s/4, p + 3s/4]$, then $[x^* - s/4, x^* + s/4] \subseteq [p, q]$, $\{x \in [p,q] \mid \pi(x) \leq \pi(x^* \pm s/4)\} = [x^* - s/4, x^* + s/4]$. Thus $m\{x \in [p,q] \mid \pi(x) \leq \pi(x^* \pm s/4)\} = s/2 = m\{x \in [p,q] \mid \pi(x) \geq \pi(x^* \pm s/4)\}$, which implies $\operatorname{med}(\pi(x)|[p,q]) = \pi(x^* \pm s/4)$.

If $x^* < p + s/4$, then $\{x \in [p,q] \mid \pi(x) \leq \pi((p+q)/2)\} = [p, p + s/2]$ and $\{x \in [p,q] \mid \pi(x) \geq \pi((p+q)/2)\} = [p + s/2, q]$ have equal measure. Thus $\operatorname{med}(\pi(x)|[p,q]) = \pi((p+q)/2)$. The same conclusion holds when $x^* > p + 3s/4$. Thus we have (2.12). □

The two cases identified above will be called the "Linear" and "Nonlinear" cases. Equations (2.8)–(2.10) always give rise to a system of three algebraic equations in three variables a, b, c. Linearity refers to dependence on these variables. When (2.13) is invoked, $x^* = -b/2c$, and so the evaluation of π—at a location depending on x^*—is a nonlinear functional.

A similar division into cases occurs when we consider median-interpolation.

PROPOSITION 2.2 (median-interpolation, $D = 2$). $\Pi_{(2)}(0, 1, 1+d) = a + bx + cx^2$ *can be computed by the following formulae:*

[N1] *If* $\frac{7}{3} \le d \le 5$, *then* $x^* \in [\frac{1}{4}, \frac{3}{4}]$, *and*

$$a = 11 + \frac{7}{2}d - \frac{5}{2}r,\ b = -\frac{32}{3} - \frac{13}{3}d + \frac{8}{3}r,\ c = \frac{8}{3} + \frac{4}{3}d - \frac{2}{3}r, \tag{2.14}$$

where $r = \sqrt{16 + 16d + d^2}$.

[N2] *If* $\frac{1}{5} \le d \le \frac{3}{7}$, *then* $x^* \in [\frac{9}{4}, \frac{11}{4}]$, *and*

$$a = -\frac{3}{2} - 2d + \frac{1}{2}d - \frac{5}{2}r,\ b = -\frac{11}{3} + \frac{16}{3}d - \frac{4}{3}r,\ c = -\frac{4}{3} - \frac{8}{3}d + \frac{2}{3}r, \tag{2.15}$$

where $r = \sqrt{1 + 16d + 16d^2}$.

[N3] *If* $-3 \le d \le -\frac{1}{3}$, *then* $x^* \in [\frac{5}{4}, \frac{7}{4}]$, *and*

$$a = -\frac{7}{12} + \frac{1}{12}d + \frac{r}{12},\ b = \frac{13}{10} - \frac{3}{10}d - \frac{r}{5},\ c = -\frac{4}{15} + \frac{4}{15}d + \frac{r}{15}, \tag{2.16}$$

where $r = -\sqrt{1 - 62d + d^2}$.

[L] *In all other cases,*

$$a = -\frac{7}{8} + \frac{3}{8}d,\ b = 2 - d,\ c = -\frac{1}{2} + \frac{d}{2}. \tag{2.17}$$

Proof. Fix a polynomial π. To calculate its block medians on blocks $[0,1]$, $[1,2]$, $[2,3]$, we can apply Proposition 2.1 successively to the choices $[p,q] = [0,1]$, $[1,2]$, $[2,3]$. We see that either the extremum of π lies in the middle half of one of the three intervals $[0,1]$, $[1,2]$, $[2,3]$ or it does not. If it does not lie in the middle half of any interval, the relation of the block medians to the coefficients is linear. If it does lie in the middle half of some interval, the relation of the block medians to the coefficients will be linear in two of the blocks—those where the extremum does not lie—and nonlinear in the block where the extremum does lie. Hence there are four basic cases to consider: (i) $x^* \in [1/4, 3/4]$, (ii) $x^* \in [9/4, 11/4]$, (iii) $x^* \in [5/4, 7/4]$, and (iv) $x^* \notin [1/4, 3/4] \cup [9/4, 11/4] \cup [5/4, 7/4]$. The first three involve some form of nonlinearity; the remaining case is linear.

Now to *solve for* a polynomial π with prescribed block medians, we can see at this point that *if we knew in advance the value of* $x^*(\pi)$, we could identify one of cases (i)–(iv) as being operative. It is easy to set up for any one of these cases a system of algebraic equations defining the desired quadratic polynomial. By writing down the system explicitly and solving it, either by hand or with the assistance of an algebraic software tool, we can obtain explicit formulae for the coefficients π. This has been done for cases (i)–(iv) and results are recorded above in (2.14)–(2.17). We omit the detailed calculation.

At this point, we have identified four different cases relating polynomials to their block medians. Within a given case, the relationship between a polynomial and its block medians is one-one. However, it remains for the moment at least conceivable that for a given collection of block medians, there would be two different cases which gave the same block medians, and hence nonunique interpolation.

We are rescued by a small miracle: *with six exceptions, a given set of block medians is consistent with exactly one of the four cases.*

To understand this, note that each of the four cases, involving a hypothesis on x^*, is consistent with block medians $[0, 1, 1 + d]$ only for a special set of values of d. We now proceed to identify the set of values of d which may arise in each given case,

case by case (but out of numerical order). Starting with case (iv), we can show that *if* $x^* \notin [1/4, 3/4] \cup [9/4, 11/4] \cup [5/4, 7/4]$*, then the associated block medians* $[0, 1, 1+d]$ *must obey* $d \notin [7/3, 5] \cup [1/5, 3/7] \cup [-3, -1/3]$. As we are in case (iv), formula (2.17) gives $\Pi_{(2)}(0, 1, 1+d) = (-7/8 + 3/8d) + (2-d)x + ((d-1)/2)x^2$, and hence

$$x^* = (d-2)/(d-1). \tag{2.18}$$

By a routine calculation, case (iv) and (2.18) combine to conclude that $d \notin [-3, -1/3] \cup [1/5, 3/7] \cup [7/3, 5]$. For future reference, set $\mathcal{L} = ([-3, -1/3] \cup [1/5, 3/7] \cup [7/3, 5])^c$.

Now turn to case (i); we can show that *if* $x^* \in [1/4, 3/4]$*, then the associated block medians* $[0, 1, 1+d]$ *must obey* $d \in [7/3, 5]$. As we are in case (i), formula (2.14) applies, and

$$x^* = \frac{32 + 13d - 8\sqrt{16 + 16d + d^2}}{4(4 + 2d - \sqrt{16 + 16d + d^2})}. \tag{2.19}$$

This and $x^* \in [1/4, 3/4]$ combine to conclude that $d \in [7/3, 5]$. For future reference, set $\mathcal{N}_1 = [7/3, 5]$.

Similar calculations show that in case (ii) we have *if* $x^* \in [9/4, 11/4]$*, then the associated block medians* $[0, 1, 1+d]$ *must obey* $d \in \mathcal{N}_2 \equiv [1/5, 3/7]$. Also in case (iii) we have *if* $x^* \in [5/4, 7/4]$*, then the associated block medians* $[0, 1, 1+d]$ *must obey* $d \in \mathcal{N}_3 \equiv [-3, -1/3]$.

We now have a collection of 4 sets: $\mathcal{L}$ and $\mathcal{N}_i$, $i = 1, 2, 3$. The sets have disjoint interiors and together cover the whole range of possible values for d. For d in the interior of one of these sets, exactly one of the four cases is able to generate the block medians $[0, 1, 1+d]$. The exceptional values of d, not in the interior of one of the sets, lie in the intersection of one of the nonlinear branches $\mathcal{N}_i$ and the linear branch $\mathcal{L}$. They are $-3, -1/3, 1/5, 3/7, 7/3, 5$. Analysis "by hand" shows that at each exceptional value of d, the formula for cases (iv) and the formula for the appropriate case (i)–(iii) give identical polynomials. Formally, one sees that if $a_L(d)$, $b_L(d)$, $c_L(d)$ denote formulas from one of the expressions (2.14)–(2.17) associated with an interval immediately to the left of an exceptional value (d_E, say), and $a_R(d)$, $b_R(d)$, $c_R(d)$ denote corresponding formulas associated with the interval immediately to the right of that same exceptional value, then

$$\lim_{d \uparrow d_E} a_L(d) = \lim_{d \downarrow d_E} a_R(d)$$

and similarly for b_L, b_R, c_L, c_R.

Thus the formulas listed above *cohere globally*. Each individual formula gives $a(d)$, $b(d)$, and $c(d)$ valid on the hypothesis that x^* lies in a certain range; but because of the continuous joining at the exceptional values, the different formulas combine to produce globally monotone functions of d. See Figure 2.2. □

The degenerate case $m_2 - m_1 = 0$ can be handled as follows: (i) if $m_1 = m_2 = m_3$, then $\Pi_{(2)}(m_1, m_2, m_3) = m_1$; (ii) otherwise $m_3 - m_2 \neq 0$, then use reversal equivariance followed by the formulae in Proposition 2.2. Notice that [N1]–[N3] are nonlinear rules, whereas [L] is linear. Figure 2.2 illustrates the nonlinearity of $\Pi_{(2)}$. Panels (a)–(c) show $a(d), b(d)$, and $c(d)$, where $a(d)+b(d)x+c(d)x^2 = \Pi_{(2)}(0, 1, d+1)$. Panel (d) shows $\partial a/\partial d$. Proposition 2.2 implies that $\Pi_{(2)}$ is basically *linear* outside

$$\mathcal{N} = \left\{ [m_1, m_2, m_3] \;\mid\; \frac{m_3 - m_2}{m_2 - m_1} \in \mathcal{N}_0 \right\}, \tag{2.20}$$

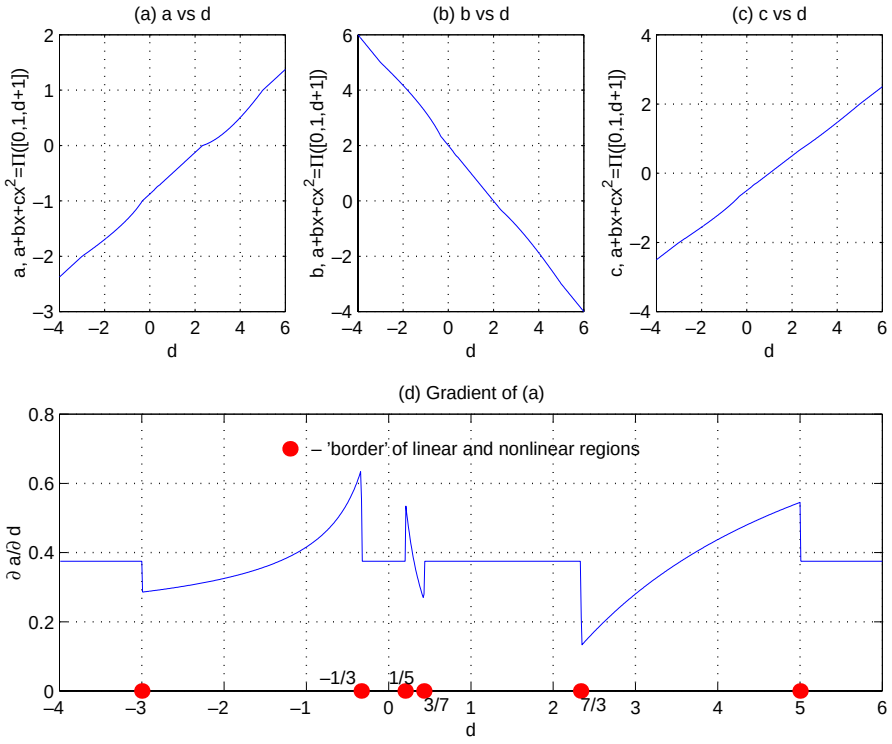

FIG. 2.2. *Nonlinearity structure of* $\Pi_{(2)}$.

where

$$\mathcal{N}_0 = [-3, -1/3] \cup [1/5, 3/7] \cup [7/3, 5]. \tag{2.21}$$

Precisely, if $\mu, \beta, a\mu + b\beta \in \mathcal{N}^c$, then

$$\Pi_{(2)}(a\mu + b\beta) = a\,\Pi_{(2)}(\mu) + b\,\Pi_{(2)}(\beta).$$

Figure 2.2(d) illustrates this point.

We now combine Propositions 2.1 and 2.2 to obtain closed-form expressions for the two-scale median-interpolating refinement operator in the quadratic case. First of all, Proposition 2.1 implies the well-posedness of median-interpolation in the case $D = 2$. Hence there exists a median-interpolating refinement operator $Q : \mathbb{R}^3 \to \mathbb{R}^3$ such that if $\pi = \pi_{j,k}$ is the fitted polynomial satisfying

$$\mathrm{med}(\pi | I_{j,k+l}) = m_l \quad \text{for} \quad -1 \le l \le 1. \tag{2.22}$$

Then

$$Q([m_{-1}, m_0, m_1]) = [\mathrm{med}(\pi | I_{j,3k}), \mathrm{med}(\pi | I_{j,3k+1}), \mathrm{med}(\pi | I_{j,3k+2})]. \tag{2.23}$$

Note that the refinement calculation is *independent* of the scale and spatial indices j and k, so Q is indeed a map from $\mathbb{R}^3$ to $\mathbb{R}^3$.

The operator Q shares two equivariance properties with $\Pi_{(2)}$:

- *Reversal equivariance.*

  $Q(m_1, m_2, m_3) = -\text{reverse}(Q(-m_3, -m_2, -m_1))$, where $\text{reverse}(p, q, r) = (r, q, p)$.

- *Affine equivariance.*

$$Q(a + bm_1, a + bm_2, a + bm_3) = a + b\,Q(m_1, m_2, m_3).$$

Q is characterized by its action on triplets $(0, 1, 1 + d)$, since if $m_2 - m_1 \neq 0$,

$$Q(m_1, m_2, m_3) = m_1 + (m_2 - m_1)Q\left(0, 1, 1 + \frac{m_3 - m_2}{m_2 - m_1}\right), \tag{2.24}$$

while if $m_2 - m_1 = 0$ and $m_3 - m_2 \neq 0$, then $Q(m_1, m_2, m_3) = \text{reverse}(Q(m_3, m_2, m_1))$ and $Q(m_3, m_2, m_1)$ can then be calculated from (2.24). Of course, when $m_2 - m_1 = m_3 - m_2 = 0$, $Q(m_1, m_2, m_3) = m_1$.

We now derive a closed-form expression for $Q(0, 1, 1 + d) = (q_1(d), q_2(d), q_3(d))$, say. By reversal equivariance, $q_3(d) = 1 + d - dq_1(\frac{1}{d})$ if $d \neq 0$. If $d = 0$, the median-interpolant of $(0, 1, 1 + d)$, π, is given by [L] of Proposition 2.2, with its maximum $x^* = 2$. A simple calculation gives $q_3(0) = \pi(\frac{11}{6}) = \frac{10}{9}$. We now work on obtaining expressions for q_1 and q_2.

PROPOSITION 2.3 (median-refinement, $D = 2$).

$$q_1(d) = \begin{cases} \frac{59}{27} + \frac{7}{27}d - \frac{8}{27}\sqrt{16 + 16d + d^2} & \text{if } d \in [\frac{7}{3}, 5], \\ \frac{26}{27} + \frac{16}{27}d - \frac{4}{27}\sqrt{1 + 16d + 16d^2} & \text{if } d \in [\frac{1}{5}, \frac{3}{7}], \\ \frac{77}{135} + \frac{13}{135}d + \frac{8}{135}\sqrt{1 - 62d + d^2} & \text{if } d \in [-3, -\frac{1}{3}], \\ -\frac{1}{288}\frac{323 - 214d + 35d^2}{-1 + d} & \text{if } d \in [-11, -3], \\ \frac{7}{9} - \frac{d}{9} & \text{otherwise}; \end{cases} \tag{2.25}$$

$$q_2(d) = \begin{cases} -\frac{1}{270}\frac{1097 - 1174d + 17d^2 + (278 - 8d)\sqrt{1 - 62d + d^2}}{-4 + 4d - \sqrt{1 - 62d + d^2}} & \text{if } d \in [-\frac{10}{7}, -\frac{7}{10}], \\ \frac{23}{30} + \frac{7}{30}d + \frac{1}{15}\sqrt{1 - 62d + d^2} & \text{if } d \in [-3, -\frac{1}{3}] \backslash [-\frac{10}{7}, -\frac{7}{10}], \\ 1 & \text{otherwise.} \end{cases} \tag{2.26}$$

Proof. Let π denote the median-interpolant of $(0, 1, 1 + d)$ associated with intervals $[0, 1], [1, 2], [2, 3]$. Recall that median-interpolation follows four branches [N1]–[N3] and [L] (cf. Proposition 2.2), whereas median-imputation follows two branches [N] and [L] (cf. Proposition 2.1.) The main task is to identify ranges of d for which median-interpolation and median-imputation use specific combinations of branches. The refinement result can then be described by obtaining algebraic expressions for each of those ranges. The calculations are similar to those in the proof of Proposition 2.2.

For q_1, there are five distinct cases:

1. $d \in [\frac{7}{3}, 5] \Rightarrow x^* \in [\frac{1}{4}, \frac{3}{4}]$: interpolation by branch [N1] and imputation by branch [L];
2. $d \in [\frac{1}{5}, \frac{3}{7}] \Rightarrow x^* \in [\frac{9}{4}, \frac{11}{4}]$: interpolation by branch [N2] and imputation by branch [L];
3. $d \in [-3, -\frac{1}{3}] \Rightarrow x^* \in [\frac{5}{4}, \frac{7}{4}]$: interpolation by branch [N3] and imputation by branch [L];

4. $d \in [-11,-3] \Rightarrow x^* \in [\frac{13}{12}, \frac{5}{4}]$: interpolation by branch [L] and imputation by branch [N];
5. $d \notin [\frac{7}{3}, 5] \cup [\frac{1}{5}, \frac{3}{7}] \cup [-3, -\frac{1}{3}] \cup [-11,-3] \Rightarrow x^* \in [\frac{13}{12}, \frac{5}{4}]$: interpolation by branch [L] and imputation by branch [L].

In each case, use the corresponding formulae in Proposition 2.2 to calculate π and then evaluate $q_1(d) = \mathrm{med}(\pi|[1, \frac{4}{3}])$ by Proposition 2.1.

For q_2, there are three distinct cases:

1. $d \in [-\frac{10}{7}, -\frac{7}{10}] \Rightarrow x^* \in [\frac{17}{12}, \frac{19}{12}]$: interpolation by branch [N3] and imputation by branch [N];[1]
2. $d \in [-3, -\frac{10}{7}] \cup [-\frac{7}{10}, -\frac{1}{3}] \Rightarrow x^* \in [\frac{5}{4}, \frac{17}{12}] \cup [\frac{19}{12}, \frac{7}{4}]$: interpolation by branch [N3] and imputation by branch [L];
3. $d \notin [-3, -\frac{1}{3}] \Rightarrow x^* \notin [\frac{5}{4}, \frac{7}{4}]$: $\mathrm{med}(\pi|[1,2]) = \mathrm{med}(\pi|[4/3, 5/3]) = \pi(3/2) \equiv 1$.

In the first two cases, again use the corresponding formulae in Proposition 2.2 to calculate π followed by evaluating $q_2(d) = \mathrm{med}(\pi|[\frac{4}{3}, \frac{5}{3}])$ using Proposition 2.1. ∎

2.1.4. $D > 2$. Higher degree median-interpolation is also well-posed: [Q1] in section 2.1 has an affirmative answer for *all* integers A. A nonconstructive proof was found independently by the second author and Goodman [18]. However, it seems difficult to obtain a closed-form expression for the nonlinear refinement operator in case $D > 2$. It is possible to develop an iterative algorithm for MI that seems to converge exponentially fast to the median-interpolating polynomial for orders $D > 2$; see [15]. Experience with this algorithm suggests that MI is stable even for orders $D > 2$.

2.2. Multiscale refinement. The two-scale refinement scheme described in section 2 applied to an initial median sequence $(\tilde{m}_{j_0,k})_k \equiv (m_{j_0,k})_k$ implicitly defines a (generally nonlinear) refinement operator $R_{MI} = R$

$$R((\tilde{m}_{j,k})_k) = (\tilde{m}_{j+1,k})_k, \qquad j \geq j_0. \tag{2.27}$$

We can associate resulting sequences $(\tilde{m}_{j,k})_k$ with piecewise constant functions on the line via

$$\tilde{f}_j(\cdot) = \sum_{k=-\infty}^{\infty} \tilde{m}_{j,k}\, 1_{I_{j,k}}(\cdot) \ \text{ for } j \geq j_0. \tag{2.28}$$

This defines a sequence of piecewise constant functions defined on successively finer and finer meshes.

In case $D = 0$, we have

$$\tilde{f}_{j_0+h} = f_{j_0} \ \text{ for all } \ h \geq 0,$$

so the result is just a piecewise constant object taking value $m_{j_0,k}$ on $I_{j_0,k}$.

In case $D = 2$, we have no closed-form expression for the result. The operator R is nonlinear, and proving the existence of a limit $\tilde{f}_{j+h}$ as $h \to \infty$ requires work.

We mention briefly what can be inferred about multiscale average-interpolation from experience in subdivision schemes and in wavelet analysis. Fix $D \in \{2, 4, \dots\}$, and let $\overline{R} = \overline{R}^{(D)}$ denote the average-interpolation operator implicitly defined by (2.6).

[1] It is worth mentioning that this is the only case where *both* the interpolation and imputation are done using *nonlinear* rules.

Set $a_{0,k} = 1_{\{k=0\}}$. Iteratively refine this sequence by the rule $(a_{j+1,k})_k = \overline{R}((a_{j,k})_k)$. Define

$$\overline{f}_j(\cdot) = \sum_{k=-\infty}^{\infty} a_{j,k}\, 1_{I_{j,k}}(\cdot) \qquad \text{for } j \geq j_0. \tag{2.29}$$

The resulting sequence of $\overline{f}_j$ converges as $j \to \infty$ to a continuous limit $\phi = \phi^{(D)}$. This is called the fundamental solution of the multiscale refinement process. Due to the linearity of average-interpolation, if we refine an arbitrary bounded sequence $(a_{j_0,k})_k$ we get a continuous limit which is a superposition of shifted and dilated fundamental solutions:

$$\overline{f}(t) = \sum a_{j_0,k}\phi(2^{j_0}t - k). \tag{2.30}$$

For median-interpolation, such a superposition result cannot hold because of the nonlinearity of the refinement scheme for $D = 2, 4, \ldots$.

Figure 2.3 illustrates the application of multiscale refinement. Panel (a) shows the $D = 2$ refinement of three Kronecker sequences $m_{0,k}^{k'} = 1_{\{k=k'\}}$, $k' = 0, 1, 2$, as well as refinement of a Heaviside sequence $1_{\{k\geq 3\}}$. Panel (b) shows the $D = 2$ refinement of a Heaviside sequence $1_{\{k\geq 0\}}$. The sequence refined in (b) is the superposition of sequences refined in (a). Panel (c) gives a superposition of shifts of (a) for $k \geq 0$; if an analogue of (2.30) held for median refinement, this should be equal to panel (b). Panel (d) gives the discrepancy, (b)–(c). Note the vertical scales. While the discrepancy from "superposability" is not large, it is definitely nonzero and not simply an artifact of rounding or other numerical processes.

3. Convergence of median-interpolation, $D = 2$. We now study some convergence properties of iterative median-interpolation. It turns out that for any bounded sequence $(m_{0,k})_k$, the sequence of nonlinear refinements $\tilde{f}_j$ converges to a bounded uniformly continuous limit $f(t)$. Moreover the limit has global Hölder exponent $\alpha > 0$. In this section, we will simplify notation and "drop tildes"; we denote a typical member of a refinement sequence by $m_{j,k}$ rather than $\tilde{m}_{j,k}$.

3.1. Weak convergence and stability. Let Q be the refinement operator as defined in (2.23), and denote $Q^j = Q \circ \cdots \circ Q$ (Q composed with itself j times). We first show that, with any initial sequence $\{m_{j_0,k}\}$, $\{f_j\}$ converges at a dense set of points.

LEMMA 3.1. *For any $[m_1, m_2, m_3] \in \mathbb{R}^3$, the limit $\lim_{j\to\infty} Q^j([m_1, m_2, m_3])$ exists.*

Proof. Let $T_{j,k}$ denote the triple of intervals $[I_{j,k-1}, I_{j,k}, I_{j,k+1}]$. If π is the median-interpolant of $[m_1, m_2, m_3]$ on $T_{j_0,k}$, then it is also the interpolant for $Q([m_1, m_2, m_3])$ on the triple $T_{j_0+1,3k+1}$ arising from triadic subdivision of the central interval of $T_{j_0,k}$. If we refine the central subinterval of $T_{j_0+1,3k+1}$, we see that π continues to be the interpolant of the resulting medians. In fact, π is the interpolant for $Q^j([m_1, m_2, m_3])$ on the triple arising from the jth such generation of subdivision of central intervals, i.e., for $T_{j_0+j,\, 3^j k+(3^j-1)/2}$ for every $j > 0$. As $j \to \infty$, the sequence of sets $(T_{j_0+j,\, 3^j k+(3^j-1)/2})_j$ collapses to the midpoint of $I_{j_0,k}$. Therefore, by continuity of π and continuity of the imputation operator,

$$\lim_{j\to\infty} Q^j([m_1, m_2, m_3]) = [m, m, m], \text{ where } m = \pi\left(3^{-j_0}\left(k + \frac{1}{2}\right)\right),$$

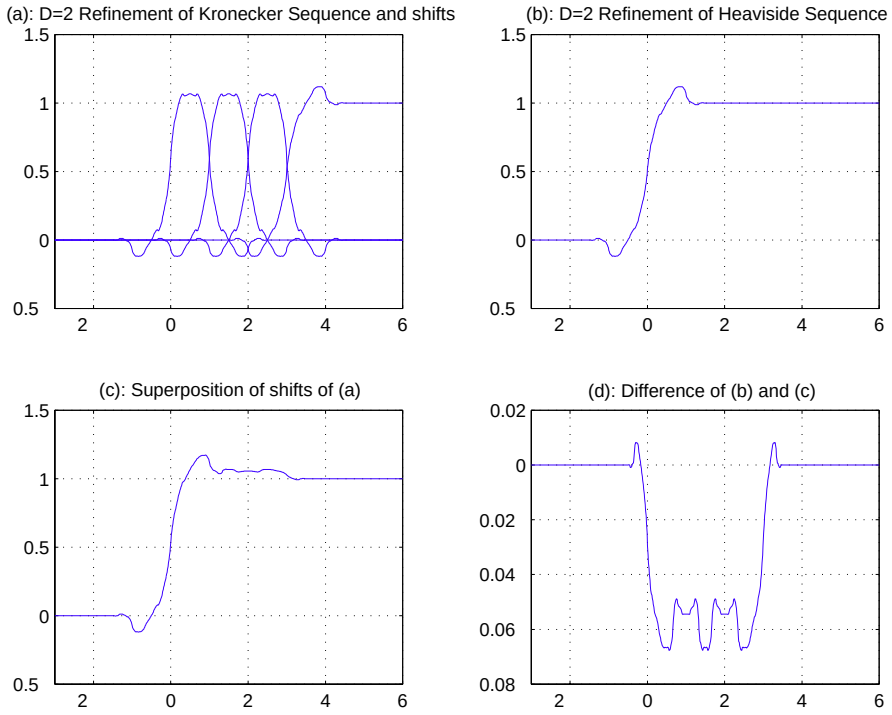

FIG. 2.3. *Discrepancy from "superposability" of multiscale median-interpolating refinement.*

the value of π at the midpoint of $I_{j_0,k}$. □

LEMMA 3.2 (convergence at triadic rationals). *For any initial median sequence* $\{m_{j_0,k}\}$*, the (nonlinear) iterative refinement scheme based on quadratic median-interpolation converges on a countable dense subset of the real line, i.e., there exists a countable dense set* $S \subset \mathbb{R}$ *and a function* $f : S \to \mathbb{R}$ *such that* $\lim_{j\to\infty} f_j(x) = f(x)$ *for every* $x \in S$.

Proof. Let $t_{j,k}$ be the midpoint of the triadic interval $I_{j,k}$. Assume we have applied the refinement scheme j_1 times to the input sequence $\{m_{j_0,k}\}_k$ (so that the values $m_{j_0+j_1,k-1}$, $m_{j_0+j_1,k}$, and $m_{j_0+j_1,k+1}$ have been calculated). We then have, for every $j > 0$, $f_{j_0+j_1+j}(t_{j_0+j_1,k}) = m^{(j)}$, where $m^{(j)}$ is the middle entry of $Q^j([m_{j_0+j_1,k-1}, m_{j_0+j_1,k}, m_{j_0+j_1,k+1}])$. By Lemma 3.1, $(m^{(j)})_j$ converges to a definite value as $j \to \infty$. We may take $S = \{t_{j,k} \mid j \geq j_0, k \in \mathbb{Z}\}$, the set of midpoints of all arbitrarily small triadic intervals, which is dense in $\mathbb{R}$. □

LEMMA 3.3. *For any* $j > 0$ *and* $k \in \mathbb{Z}$*, let* $d_{j,k} = m_{j,k} - m_{j,k-1}$*. Then*

$$|f(t_{j,k}) - m_{j,k}| \leq \frac{1}{15} \max\{|d_{j,k-1}|, |d_{j,k}|\}, \tag{3.1}$$

where the upper bound is attained if and only if $d_{j,k-1} = -d_{j,k}$.

Proof. From the proof of Lemma 3.2, $f(t_{j,k}) = \pi(t_{j,k})$, where π median-interpolates $m_{j,k-1}$, $m_{j,k}$, and $m_{j,k+1}$ for the triple $[I_{j,k-1}, I_{j,k}, I_{j,k+1}]$. Thus $|f(t_{j,k}) - m_{j,k}| = |\pi(t_{j,k}) - m_{j,k}|$. Unless $d = d_{j,k+1}/d_{j,k} \in [-3, -1/3]$, $\pi(t_{j,k}) = m_{j,k}$ and (3.1) holds trivially. When $d = d_{j,k+1}/d_{j,k} \in [-3, -1/3]$, $\pi(x)$ is given by [N3] of Proposition 2.2.

Without loss of generality, we can work with $j = 0$, $k = 0$ and denote (for simplicity) $m_i = m_{0,i}$, $i = -1, 0, 1$. Without loss of generality, we assume $\max\{|m_0 - m_{-1}|, |m_1 - m_0|\} = 1$, $m_{-1} = 0$, $m_0 = 1$ and $d = \frac{m_1 - m_0}{m_0 - m_{-1}} = m_1 - 1 \in [-1, -\frac{1}{3}]$. Then $|\pi(t_{j,k}) - m_{j,k}|/\max(|d_1|, |d_2|) = |\pi(3/2) - 1|$, where $\pi(x)$ is given by [N3] in Proposition 2.2, and we have

$$\begin{aligned}\max \frac{|\pi(t_{j,k}) - m_{j,k}|}{\max(|d_1|, |d_2|)} &= \max_{d \in [-1, -\frac{1}{3}]} |a + b(3/2) + c(3/2)^2 - 1| \\ &= \max_{d \in [-1, -\frac{1}{3}]} \left| \frac{7}{30}(d-1) + \frac{\sqrt{1 - 62d + d^2}}{15} \right| = \frac{1}{15}.\end{aligned}$$

The maximum is attained at $d = -1$. □

3.2. Hölder continuity. We now develop a basic tool for establishing Hölder continuity of refinement schemes.

THEOREM 3.4. *Let $(m_{0,k})_k$ be a bounded sequence, and $m_{j,k}$ be the refinement sequences generated by the quadratic median-interpolating refinement scheme constructed in section* 2.2. *Let $d_{j,k} := m_{j,k} - m_{j,k-1}$ and $f_j := \sum_k m_{j,k} 1_{I_{j,k}}$. Suppose that for some $\alpha > 0$ and all $j \geq 0$, $\sup_k |d_{j,k}| \leq C \cdot 3^{-j\alpha}$. Then f_j converges uniformly to a bounded limit $f \in \dot{C}^\alpha$. The converse is also true if $\alpha \leq 1$.*

This is analogous to results found in the literature of linear refinement schemes (c.f. Theorem 8.1 of Rioul [25]). The proof of the forward direction uses basically the same arguments as in the linear case, except that one must deal with nonlinearity using a general affine-invariance property of medians. Similar arguments could be applied in the study of cases $D > 2$. The proof of the converse direction, on the other hand, relies on Lemma 3.3 and is therefore specific to the $D = 2$ triadic case.

Proof. ($\Rightarrow$) We show that $\{f_j\}$ is a Cauchy sequence. Consider

$$\sup_x |f_{j+1}(x) - f_j(x)| = \sup_k \max_{\epsilon=0,1,2} |m_{j+1,3k+\epsilon} - m_{j,k}|. \tag{3.2}$$

The functions $m_{j+1,3k+\epsilon} = q_\epsilon(m_{j,k-1}, m_{j,k}, m_{j,k+1})$ obey

$$q_\epsilon(m_{j,k-1}, m_{j,k}, m_{j,k+1}) = m_{j,k} + q_\epsilon(-d_{j,k}, 0, d_{j,k+1}).$$

Moreover, by Lemma 6.2, these functions are Lipschitz: $q_\epsilon(m_1, m_2, m_3) \leq c \max_{i=1,2,3}\{|m_i|\}$. Therefore,

$$\sup_x |f_{j+1}(x) - f_j(x)| = \sup_k \max_{\epsilon=0,1,2} |q_\epsilon(-d_{j,k}, 0, d_{j,k+1})| \leq c \sup_k |d_{j,k}| \leq c \cdot C \cdot 3^{-j\alpha}$$

and $||f_{j+p} - f_j||_\infty \leq cC(3^{-j\alpha} + \cdots + 3^{-(j+p-1)\alpha}) \leq C' 3^{-j\alpha}$. ($C'$ is independent of p.) Hence $\{f_j\}$ is a Cauchy sequence that converges uniformly to a function f. Furthermore, $f \in \dot{C}^\alpha$ because $\sup_{3^{-(j+1)} \leq |h| \leq 3^{-j}} |f(x+h) - f(x)| \leq |f(x+h) - f_j(x+h)| + |f(x) - f_j(x)| + |f_j(x+h) - f_j(x)| \leq C' 3^{-j\alpha} + C' 3^{-j\alpha} + C 3^{-j\alpha} \leq 3^\alpha (2C' + C)|h|^\alpha$.

($\Leftarrow$) If $f \in \dot{C}^\alpha$, $\alpha \leq 1$, then, by definition, $\sup_k |f(t_{j,k}) - f(t_{j,k-1})| \leq c \cdot 3^{-j\alpha}$. But

$$\begin{aligned}|m_{j,k} - m_{j,k-1}| &\leq |m_{j,k} - f(x_{j,k})| + |f(t_{j,k}) - f(t_{j,k-1})| + |m_{j,k-1} - f(t_{j,k-1})| \\ &\leq \frac{1}{15} \max\{|d_{j,k}|, |d_{j,k-1}|\} + c \cdot 3^{-j\alpha} + \frac{1}{15} \max\{|d_{j,k-1}|, |d_{j,k-2}|\}.\end{aligned}$$

The last inequality is due to Lemma 3.3. Maximizing over k on both sides of the above inequality, followed by collecting terms, gives $\sup_k |d_{j,k}| \leq (15/13\ c) \cdot 3^{-j\alpha}$. □

3.3. Nonlinear difference scheme. As in Theorem 3.4, let $d_{j,k} = m_{j,k} - m_{j,k-1}$ denote the sequence of interblock differences. It is a typical property of any *constant-reproducing* linear refinement scheme that the difference sequences can themselves be obtained from a linear refinement scheme, called the *difference scheme*. The coefficient mask of that scheme is easily derivable from that of the original scheme; see [16, 25]. More generally, a linear refinement scheme that can reproduce all lth degree polynomials would possess $l+1$ difference (and divided difference) schemes [16]. A partial analogy to this property holds in the nonlinear case: the $D = 2$ median-interpolation scheme, being a nonlinear refinement scheme with quadratic polynomial reproducibility, happens to possess a (nonlinear) first difference scheme but no higher order ones.

Let $d_{j,k} \neq 0, d_{j,k+1} \neq 0, d_{j,k+2}$ be given. Then, by (2.24),

$$\begin{aligned}
&(m_{j+1,3k}, m_{j+1,3k+1}, m_{j+1,3k+2}) \\
&\qquad = m_{j,k-1} + d_{j,k}\, Q\left(0, 1, 1 + \frac{d_{j,k+1}}{d_{j,k}}\right) \\
&\qquad = m_{j,k-1} + d_{j,k}\, \left(q_1\left(\frac{d_{j,k+1}}{d_{j,k}}\right), q_2\left(\frac{d_{j,k+1}}{d_{j,k}}\right), q_3\left(\frac{d_{j,k+1}}{d_{j,k}}\right)\right), \\
&(m_{j+1,3k+3}, m_{j+1,3k+4}, m_{j+1,3k+5}) \\
&\qquad = m_{j,k} + d_{j,k+1} Q\left(0, 1, 1 + \frac{d_{j,k+2}}{d_{j,k+1}}\right) \\
&\qquad = m_{j,k-1} + d_{j,k} + d_{j,k+1}\left(q_1\left(\frac{d_{j,k+2}}{d_{j,k+1}}\right), q_2\left(\frac{d_{j,k+2}}{d_{j,k+1}}\right), q_3\left(\frac{d_{j,k+2}}{d_{j,k+1}}\right)\right).
\end{aligned}$$

Hence $d_{j+1,3k+1} d_{j+1,3k+2}, d_{j+1,3k+3}$ are only dependent on $d_{j,k}, d_{j,k+1}, d_{j,k+2}$ and there exist three functionals $\partial q_0 : \mathbb{R}^2 \to \mathbb{R}$, $\partial q_1 : \mathbb{R}^2 \to \mathbb{R}$, $\partial q_2 : \mathbb{R}^3 \to \mathbb{R}$ such that $d_{j+1,3k+1} = \partial q_0(d_{j,k}, d_{j,k+1})$, $d_{j+1,3k+2} = \partial q_1(d_{j,k}, d_{j,k+1})$, and $d_{j+1,3k+3} = \partial q_2(d_{j,k}, d_{j,k+1}, d_{j,k+2})$, where

$$\begin{aligned}
\partial q_0(d_0, d_1) &= d_0\left(q_2\left(\frac{d_1}{d_0}\right) - q_1\left(\frac{d_1}{d_0}\right)\right) \quad (\text{when } d_0 \neq 0), \\
\partial q_1(d_0, d_1) &= d_0\left(q_3\left(\frac{d_1}{d_0}\right) - q_2\left(\frac{d_1}{d_0}\right)\right) \quad (\text{when } d_0 \neq 0), \\
\partial q_2(d_0, d_1, d_2) &= d_0 + d_1 q_1\left(\frac{d_2}{d_1}\right) - d_0 q_3\left(\frac{d_1}{d_0}\right) \quad (\text{when } d_0 \neq 0 \text{ and } d_1 \neq 0), \\
&= d_0 + d_1 q_1\left(\frac{d_2}{d_1}\right) - d_0\left(1 + \frac{d_1}{d_0} - \frac{d_1}{d_0} q_1\left(\frac{d_0}{d_1}\right)\right). \qquad (3.3)
\end{aligned}$$

The degenerate cases can be handled easily. One of those will be of use later, namely,

$$\partial q_0(0, d_1) = \frac{d_1}{9} = \lim_{d_0 \to 0} d_0\left(q_2\left(\frac{d_1}{d_0}\right) - q_1\left(\frac{d_1}{d_0}\right)\right). \qquad (3.4)$$

Similar limits hold for ∂q_1 and ∂q_2.

The difference scheme inherits two nice equivariance properties from median-interpolation:

- *Reversal equivariance.*

$$\begin{aligned}
\partial q_0(d_1, d_0) &= \partial q_1(d_0, d_1), \partial q_1(d_1, d_0) \\
&= \partial q_0(d_0, d_1), \text{ and } \partial q_2(d_2, d_1, d_0) = \partial q_2(d_0, d_1, d_2).
\end{aligned}$$

- *Affine equivariance.*

$$\partial q_\epsilon(b(d_0,d_1)) = b\,\partial q_\epsilon(d_0,d_1), \epsilon = 0,1, \text{ and}$$
$$\partial q_2(b(d_0,d_1,d_2)) = b\,\partial q_2(d_0,d_1,d_2).$$

The above discussion implies the existence of three (nonlinear) operators $\partial Q_\epsilon : \mathbb{R}^3 \to \mathbb{R}^3$, $\epsilon = 0,1,2$ that govern the difference scheme:

$$\partial Q_\epsilon([d_{j,k-1}, d_{j,k}, d_{j,k+1}]^T) = [d_{j+1,3k+\epsilon}, d_{j,3k+\epsilon}, d_{j,3k+\epsilon}]^T \text{ for all } j \geq 0, k \in \mathbb{Z}. \tag{3.5}$$

Uniform convergence will follow from the fact that these operators are *shrinking* in the sense that

$$S_\infty(\partial Q_\epsilon) := \max_{d\in\mathbb{R}^3} \frac{||\partial Q_\epsilon(d)||_\infty}{||d||_\infty} = \beta < 1, \quad \epsilon = 0,1,2. \tag{3.6}$$

As the ∂Q_ϵ are nonlinear, this is slightly weaker than being *contractive*. We will prove an inequality like this in the next section.

It is easy to check that $\partial Q_\epsilon(d) = 0$ if and only if $d = 0$ and that $S_\infty(\partial Q_{\epsilon_1} \circ \partial Q_{\epsilon_2}) \leq S_\infty(\partial Q_{\epsilon_1}) S_\infty(\partial Q_{\epsilon_2})$. In order to bound the decay rate of $\max_k |d_{j,k}|$ (and hence the critical Hölder exponent for median-interpolating refinements), we can use the estimate

$$\sup_k |d_{j,k}| \leq \sup_k |d_{0,k}| \max_{\epsilon_i=0,1,2} S_\infty(\partial Q_{\epsilon_j} \circ \cdots \circ \partial Q_{\epsilon_1}). \tag{3.7}$$

Assuming (3.6), we can bound the right-hand side of (3.7) crudely by

$$\max_{\epsilon_i=0,1,2} S_\infty(\partial Q_{\epsilon_j} \circ \cdots \circ \partial Q_{\epsilon_1}) \leq \max_{\epsilon_i=0,1,2} S_\infty(\partial Q_{\epsilon_j}) \times \cdots \times S_\infty(\partial Q_{\epsilon_1}) \tag{3.8}$$
$$= \beta^j = 3^{-j\,\alpha}, \tag{3.9}$$

where $\alpha = \log_3(1/\beta) > 0$. Hence, uniform convergence follows from Theorem 3.4.

Actually, the inequality (3.8) contains slack. It is possible to improve on it by adapting to the nonlinear case approaches developed by Rioul [25] and Dyn, Gregory, and Levin [16] in the study of linear refinement schemes. We state without proof the following: Define α_j by

$$3^{-j\alpha_j} = \max_{\epsilon_i=0,1,2} S_\infty(\partial Q_{\epsilon_j} \circ \cdots \circ \partial Q_{\epsilon_1}). \tag{3.10}$$

Let $\alpha := \sup_j \alpha_j$. Then $\lim_j \alpha_j = \alpha$ and median-interpolating refinements are $\dot{C}^{\alpha-\epsilon}$ for $\epsilon > 0$. This observation is potentially useful because it provides a way to compute *lower bounds* for the Hölder regularity of median-interpolating refinement limits. In the next section, we apply this idea with the choice of $j = 1$, which results in the crude lower bound $\alpha_1 = \log_3(135/121)$. A better bound might be obtained if one could manage to compute the right-hand side of (3.10) for a larger j.

3.4. The difference scheme is shrinking. Armed with the closed-form expression for the quadratic median-interpolating refinement scheme, we can explicitly calculate $S_\infty(\partial Q_\epsilon)$ despite the nonlinearity of the operator.

THEOREM 3.5. *$S_\infty(\partial Q_\epsilon) = 121/135 < 1$ for $\epsilon = 0,1,2$. Consequently, by Theorem 3.4 and (3.7)–(3.9), for any bounded initial sequence $m_{0,k}$, the sequence of*

nonlinear refinements $f_j = \sum_k m_{j,k} 1_{I_{j,k}}$ *converges uniformly to a bounded uniformly continuous function* $f \in \dot{C}^\alpha$, *where* $\alpha = \log_3(135/121) \approx 0.0997$.

Proof. See [31] for the computational details. The main idea of the proof is to verify that $\partial q_0(d_0, d_1)$ and $\partial q_1(d_0, d_1)$ are monotone increasing in d_0 for fixed d_1 and monotone increasing in d_1 for fixed d_0; and that $\partial q_2(d_0, d_1, d_2)$ is monotone decreasing in d_0 and d_2 for fixed d_1 and monotone increasing in d_1 for fixed d_0 and d_2. Thus $S_\infty(\partial q_0) := \max_{|d_0|,|d_1|\le 1} \partial q_0(d_0, d_1) = \partial q_0(1, 1) = 1/3$, and, by symmetry, $S_\infty(\partial q_1) = S_\infty(\partial q_0) = 1/3$; $S_\infty(\partial q_2) = \partial q_2(-1, 1, -1) = 121/135$. The theorem follows from the fact that $S_\infty(\partial Q_\epsilon) = \max_{i=0,1,2} S_\infty(\partial q_i)$ for $\epsilon = 0, 1, 2$. □

3.5. Discussion. The regularity bound $\alpha \ge \log_3(135/121) \approx 0.0997$ is probably very far from sharp. We now discuss evidence suggesting that the sharp Hölder exponent is nearly 1.

3.5.1. Linearized median-interpolation. We recall from Figure 2.2, and Propositions 2.1, 2.2, and 2.3 that there is an underlying *linear branch* associated with the median scheme. A sufficient but not necessary condition for applicability of this branch is that the block medians be consistent with a polynomial π that is monotone throughout $[a, b]$.

In the linear branch, the median functional amounts to *midpoint evaluation*: $\mathrm{med}(\pi|[a, b]) = \pi((a + b)/2)$. The resulting refinement rule is a linear scheme that we call the *LMI* scheme, with coefficient mask $[-1/9, 0, 2/9, 7/9, 1, 7/9, 2/9, 0, -1/9]$. It is a symmetric interpolatory scheme and can be viewed as a triadic variant of Deslauriers–Dubuc schemes. The mask has a positive Fourier transform, and the convergence and critical Hölder regularity of the scheme can be determined quite easily by applying the theory of linear refinement schemes [25, 5, 8].

The LMI scheme has refinement limits which are "almost Lipschitz" [31]. For any given bounded initial sequence of block values at scale 0, the LMI scheme converges to a bounded uniformly continuous limit f obeying the regularity estimate $\sup_x |f(x + h) - f(x)| \le C|h| \log(1/|h|)$. Moreover, the above global regularity bound cannot be improved. (If we study *local* rather than global regularity, it can be shown, using techniques in [5], that the bound can be improved for *most* x.)

See Figure 3.1 for pictures of median-interpolating and linearized median-interpolating refinement limits of the Kronecker sequence $\{m_{0,k} = \delta_{0,k}\}$.

3.5.2. Critical Hölder exponent conjectures. We conjecture that MI and LMI share the same global Hölder regularity. This is a rather natural conjecture to make, since the difference between MI and LMI is actually very small—as one sees from the near-linearity of the functions displayed in Figure 2.2. In [31], computational evidence was provided to support the conjecture. In particular, the experiments there suggest the following:

1. The actual decay behavior in (3.10) is $O(j3^{-j})$, which is much faster than the rate bound calculated in Theorem 3.5. This rate would imply that median-interpolating refinement limits are almost Lipschitz. See panel (d) of Figure 3.1.
2. Both the MI refinement sequences $m^{MI}_{j,k}$ and the LMI refinement sequences $m^{LMI}_{j,k}$ appear to possess a *stationarity property*: let k^*_j be the value of k maximizing $|d^{MI}_{j,k}|$; here there exists an integer k^* such that $3^{-j} k^*_j = k^*$ for all large enough j. See panel (b) of Figure 3.1. The same phenomenon is observed for $d^{LMI}_{j,k}$; see panel (c) of Figure 3.1. Stationarity is a *provable*

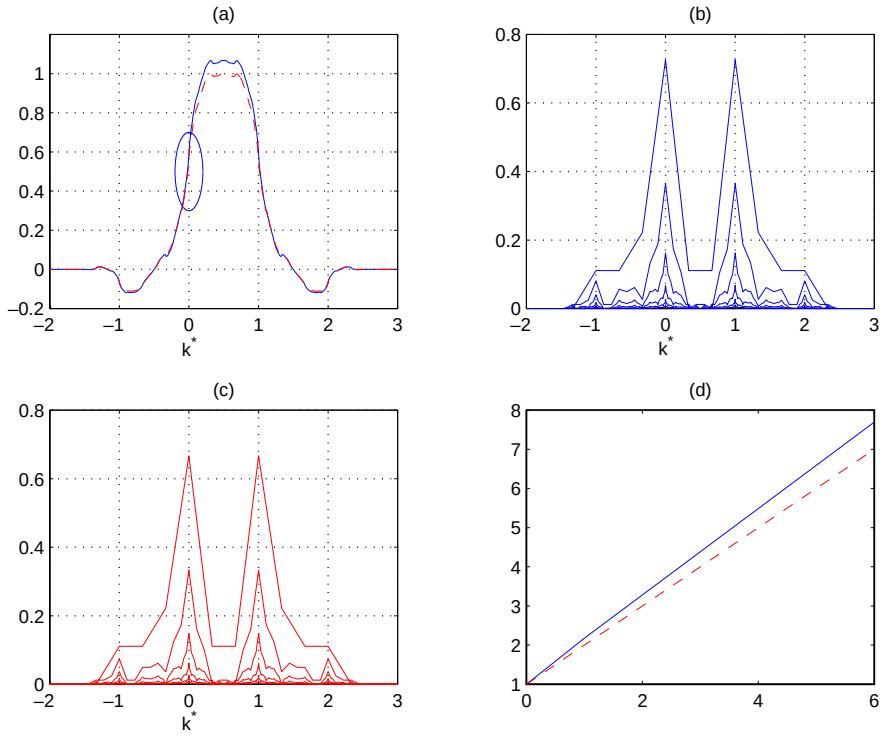

FIG. 3.1. *MI versus LMI:* (a) *MI- and LMI- refinements (solid and dashed lines, respectively) of* $m_{0,k} = \delta_{0,k}$, (b) $|d^{MI}_{j,k}|$ *versus* $k3^{-j}$, $j = 1, \ldots, 6$, (c) $|d^{LMI}_{j,k}|$ *versus* $k3^{-j}$, $j = 1, \ldots, 6$, (d) $3^j \max_k |d^{MI}_{j,k}|$ *and* $3^j \max_k |d^{LMI}_{j,k}|$ *versus* j *(solid and dashed lines, respectively).*

property of the LMI scheme. It is *empirically* a property of the MI scheme as well.

3. It appears that in the vicinity of the spatial location $x = k^*$, the limit function is monotone and (consequently) median-interpolating refinement is repeatedly using its *linear branch*. Therefore, it seems that $\sup_k |d^{MI}_{j,k}|$ and $\sup_k |d^{LMI}_{j,k}|$ share the same asymptotics. See again Figure 3.1.

A more ambitious open question is the following: Let $x \in \mathbb{R}$, and let $k_j(x)$ be defined by $x \in I_{j,k_j(x)}$. We call x an *asymptotically linear point* if, for large enough j, median-interpolation is only using its linear branch to determine $m_{j+1,k_{j+1}(x)}$ from $m_{j,k_j(x)+\epsilon}$, $\epsilon = -1, 0, 1$. In order to understand deeply the relation between median-interpolation and linearized median-interpolation, it would be useful to determine the structure of the set of asymptotically linear points.

4. Median-interpolating pyramid transform. We now apply the refinement scheme to construct a nonlinear pyramid and associated nonlinear multiresolution analysis.

4.1. Pyramid algorithms. While it is equally possible to construct pyramids for decomposition of functions $f(t)$ or of sequence data y_i, we keep an eye on applications and concentrate attention on the sequence case. So we assume we are given a discrete dataset $y_i, i = 0, \ldots, n-1$, where $n = 3^J$ is a triadic number. We aim to use

the nonlinear refinement scheme to decompose and reconstruct such sequences.

ALGORITHM FMIPT: PYRAMID DECOMPOSITION.

1. *Initialization.* Fix $D \in 0, 2, 4, \ldots$ and $j_0 \geq 0$. Set $j = J$.
2. *Formation of block medians.* Calculate

$$m_{j,k} = \mathrm{med}(y_i : i/n \in I_{j,k}). \tag{4.1}$$

   (Here med() refers to the discrete median rather than the continuous median.)
3. *Formation of refinements.* Calculate

$$\tilde{m}_{j,k} = R((m_{j-1,k})) \tag{4.2}$$

   using refinement operators of the previous section.
4. *Formation of detail corrections.* Calculate

$$\alpha_{j,k} = m_{j,k} - \tilde{m}_{j,k}.$$

5. *Iteration.* If $j = j_0 + 1$, set $m_{j_0,k} = \mathrm{med}(y_i : i/n \in I_{j_0,k})$ and terminate the algorithm, else set $j = j - 1$ and goto 2.

ALGORITHM IMIPT: PYRAMID RECONSTRUCTION.

1. *Initialization.* Set $j = j_0 + 1$. Fix $D \in 0, 2, 4, \ldots$ and $j_0 \geq 0$, as in the decomposition algorithm.
2. *Reconstruction by refinement.*

$$(m_{j,k}) = R((m_{j-1,k})) + (\alpha_{j,k})_k.$$

3. *Iteration.* If $j = J$ goto 4, else set $j = j + 1$ and goto 2.
4. *Termination.* Set

$$y_i = m_{J,i}, \quad i = 0, \ldots, n-1.$$

An implementation is described in [31]. Important details described there include the treatment of boundary effects and efficient calculation of block medians.

DEFINITION 4.1. *Gather the outputs of the pyramidal decomposition algorithm into the sequence*

$$\theta = ((m_{j_0,k})_k, (\alpha_{j_0+1,k})_k, (\alpha_{j_0+2,k})_k, \ldots, (\alpha_{J,k})_k).$$

We call θ the MIPT of y and we write $\theta = MIPT(y)$. Applying the pyramidal reconstruction algorithm to θ gives an array which we call the inverse transform, and we write $y = MIPT^{-1}(\theta)$.

The reader may wish to check that $MIPT^{-1}(MIPT(y)) = y$ for every sequence y.

We will also use below the average-interpolating pyramid transform (AIPT), defined in a completely parallel way, using only the average-interpolation refinement operator $\overline{R}$. We write $\theta = AIPT(y)$ and $y = AIPT^{-1}(\theta)$.

Complexity. Both transforms have good computational complexity. The refinement operator for $AIPT$, in common with wavelet transforms and other multiscale algorithms, has order $O(n)$ computational complexity. The coarsening operator can be implemented with the same complexity because of a *causality* relationship:

$$\mathrm{ave}(y_i|I_{j,k}) = \mathrm{ave}(\mathrm{ave}(y_i|I_{j+1,3k}),\ \mathrm{ave}(y_i|I_{j+1,3k+1}),\ \mathrm{ave}(y_i|I_{j+1,3k+2})). \tag{4.3}$$

Similarly, the refinement operator of $MIPT$ of order $D = 2$ has complexity $O(n)$ due to the propositions of section 2.1.3. However, for the coarsening operator there is no direct causality relationship. The analogue of (4.3) obtained by replacing "ave" by "med" does not hold.

To rapidly calculate all medians over triadic blocks, one can maintain sorted lists of the data in each triadic block; the key coarsening step requires merging three sorted lists to obtain a single sorted list. This process imposes only a $\log_3(n)$ factor in additional cost. For a more detailed description of the implementation, we refer to [31]. As a result, $MIPT$ can be implemented by an $O(n \log_3 n)$ algorithm, whereas $MIPT^{-1}$ can be implemented with $O(n)$ time-complexity.

4.2. Properties. P1. *Coefficient localization.* The coefficient $\alpha_{j,k}$ in the pyramid only depends on block medians of blocks at scale $j-1$ and j which cover or abut the interval $I_{j,k}$.

P2. *Expansionism.* There are 3^{j_0} résumé coefficients $(m_{j_0,k})$ in θ and 3^j coefficients $(\alpha_{j,k})_k$ at each level j. Hence

$$\mathrm{Dim}(\theta) = 3^{j_0} + 3^{j_0+1} + \cdots + 3^J.$$

It follows that $\mathrm{Dim}(\theta) = 3^J(1 + 1/3 + 1/9 + \cdots) \sim 3/2 \cdot n$. The transform is about 50% expansionist.

P3. *Coefficient decay.* Suppose that the data $y_i = f(i/n)$ are noiseless samples of a continuous function $f \in \dot{C}^\alpha$, $0 \le \alpha \le 1$, i.e., $|f(s) - f(t)| \le C|s-t|^\alpha$ for a fixed C. Then for $MIPT$ $D = 0$ or 2, we have

$$|\alpha_{j,k}| \le C' C 3^{-j\alpha}. \tag{4.4}$$

Suppose f is $\dot{C}^{r+\alpha}$ for $r = 1$ or 2, i.e., $|f^{(r)}(s) - f^{(r)}(t)| \le C|s-t|^\alpha$, for some fixed α and C, $0 < \alpha \le 1$. Then, for $MIPT$ $D = 2$,

$$|\alpha_{j,k}| \le C' C 3^{-j(r+\alpha)}. \tag{4.5}$$

P4. *Gaussian noise.* Suppose that $y_i = \sigma z_i$, $i = 0, \ldots, n-1$, and that z_i is i.i.d $N(0,1)$, a standard Gaussian white noise. Then

$$P(\sqrt{3^{J-j}}\,|\alpha_{j,k}| \ge \xi) \le C_1 \cdot \exp\left(-C_2 \frac{\xi^2}{\sigma^2}\right),$$

where the $C_i > 0$ are absolute constants.

These properties are things we naturally expect of linear pyramid transforms, such as those of Adelson and Burt, and P1, P3, and P4 we expect also of wavelet transforms. In fact these properties hold not just for MIPT but also for AIPT.

A key property of MIPT but *not* AIPT is the following.

P5. *Cauchy noise.* Suppose that $y_i = \sigma z_i$, $i = 0, \ldots, n-1$, and that z_i is i.i.d standard Cauchy white noise. Then

$$P(\sqrt{3^{J-j}}\,|\alpha_{j,k}| \ge \xi) \le C_1' \cdot \exp\left(-C_2' \frac{\xi^2}{\sigma^2}\right),$$

where $0 \le \xi \le \sqrt{3^{J-j}}$ and the $C_i' > 0$ are absolute constants.

For a linear transform, such as AIPT, the coefficients of Cauchy noise have Cauchy distributions, and such exponential bounds cannot hold. Moreover, the spread of the

resulting Cauchy distributions does not decrease with increasing j. In contrast, P5 shows that the spread of the MIPT coefficients gets smaller with larger j, and that deviations more than a few multiples of the spread are very rare.

Properties P1 and P2 need no further proof; P3–P5 are proved in the appendix.

4.3. MRA. MI refinement allows us to mimic the multiresolution analysis of wavelet theory. Given the sequence $(m_{j,k})$ of block medians of y at scale j, we may apply $J-j$ iterations of two-scale refinement to these medians, getting a sequence of length n which we can call $P_j y$. An equivalent definition is as follows:

- Decomposition. $\theta = MIPT(y)$.
- Suppression of details. Let $\tilde{\theta}$ be a partial copy of θ, where we set $\alpha_{j',k} = 0$ for $j' > j$.
- Reconstruction. $P_j y = MIPT^{-1}(\tilde{\theta})$.

P_j is a nonlinear approximation of y at the scale j, because it uses only the block medians at scale j in its construction.

We can also form $Q_j y = P_j y - P_{j-1} y$, listing the details present in the approximation at scale j but not present at scale $j-1$.

4.4. Examples. We collect here a few examples of the MIPT for $D = 2$.

Figure 4.1 shows three different noiseless signals: (a) Sinusoid; (b) Heaviside; (c) Doppler. It also shows the pyramid coefficients of noiseless data for $D = 2$.

Figure 4.2 shows an MRA decomposition of the same three signals. This display shows $P_{j_0} y$, $Q_{j_0+1} y, \ldots, Q_J y$.

5. Denoising by MIPT thresholding. We now consider applications of pyramid transforms to multiscale denoising. In general, we act as we would in the wavelet denoising case.

- *Pyramid decomposition.* Calculate $\theta = MIPT(y)$.
- *Hard thresholding.* Let $\eta_t(y) = y \cdot 1_{\{|y|>t\}}$ be the hard thresholding function and let
$$\hat{\theta} = ((m_{j_0,k})_k, (\eta_{t_{j_0+1}}(\alpha_{j_0+1,k}))_k, \ldots).$$
Here the (t_j) is a sequence of threshold levels.
- *Pyramid reconstruction.* Calculate $\hat{f} = MIPT^{-1}(\hat{\theta})$.

In this approach, coefficient amplitudes smaller than t_j are judged negligible, as noise rather than signal. Hence the thresholds t_j control the degree of noise rejection but also of valid signal rejection. One hopes, in analogy with the orthogonal transform case studied in [12], to set thresholds which are small but which are very likely to exceed every coefficient in case of a pure noise signal. If the MIPT performs as we hope, the MIPT thresholds can be set "as if" the noise were Gaussian and the transform were AIPT, even when the noise is very non-Gaussian. This would mean that the median pyramid is immune to bad effects of impulsive noise.

5.1. Choice of thresholds. Motivated by P4 and P5, we work with the "L^2-normalized" coefficients $\bar{\alpha}_{j,k} = \sqrt{3^{J-j}}\alpha_{j,k}$ in this section.

In order to choose thresholds $\{t_j\}$ which are very likely to exceed every coefficient in case of a pure noise signal, we find t_j satisfying $P(|\bar{\alpha}_{j,k}| > t_j) \le c \cdot 3^{-J}/J$ where the MIPT coefficients arise from a pure noise signal $(X_i)_{i=0}^{3^J-1}$, $X_i \sim_{i.i.d.} F$. Then we have

$$P(\exists (j,k) \text{ s.t. } |\bar{\alpha}_{j,k}| > t_j) \le c \cdot \sum_{j=j_0}^{J} \sum_{k=0}^{3^j-1} \frac{1}{J} 3^{-J} \to 0 \text{ as } J \to \infty. \tag{5.1}$$

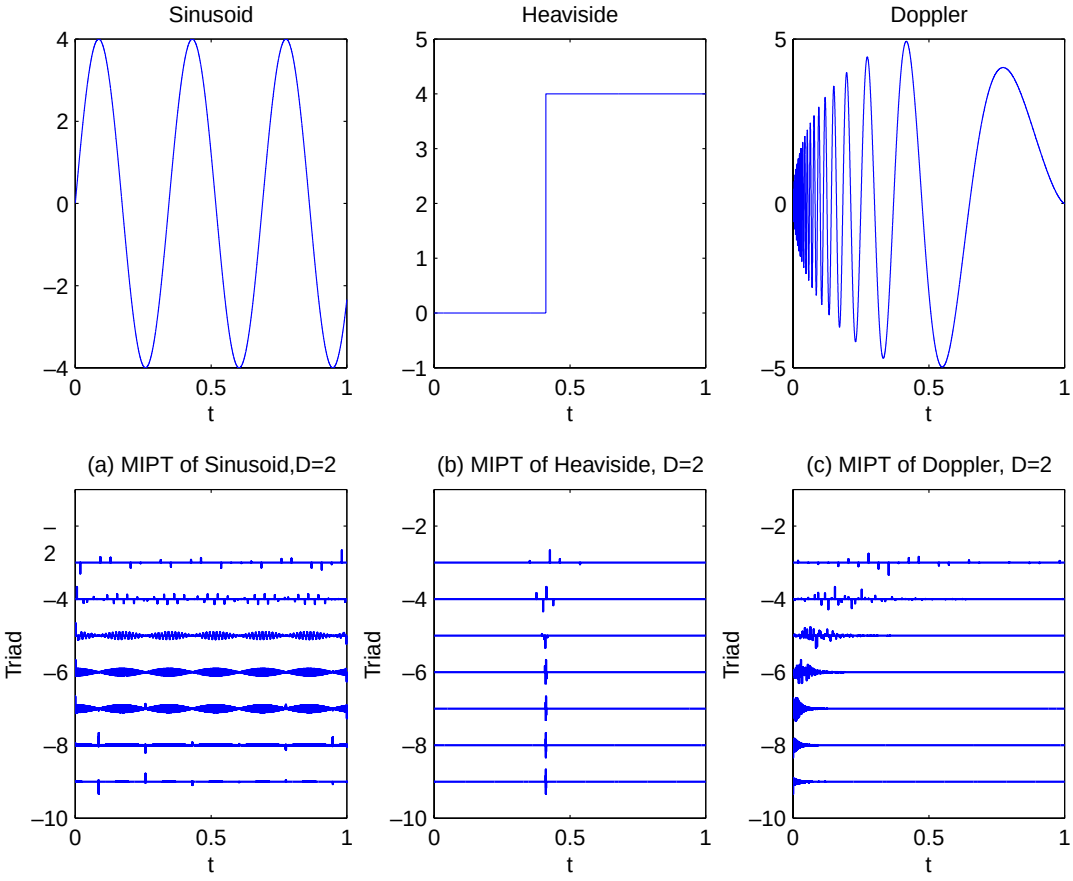

FIG. 4.1. *MIPT coefficients of three synthetic signals: Sinusoid, Heaviside, and Doppler. In each case, the plot of* $(\alpha_{j,k})_k$ *is scaled by* $\max_k |\alpha_{j,k}|$.

By (6.5), we can simply choose t_j satisfying $P(\sqrt{3^{J-j}}|\operatorname{med}(X_1,\ldots,X_{3^{J-j}})| > t_j) \le 3^{-J}/J$. Corollary 6.5 gives, when F is a symmetric law,

$$t_j := t_j(F) = \sqrt{3^{J-j}}F^{-1}\left(\frac{1}{2}+\frac{1}{2}\sqrt{1-\left(\frac{1}{2J3^J}\right)^{\frac{2}{3^{J-j}}}}\right). \tag{5.2}$$

Careful study of (5.2) suggests to us that away from the finest scales, the magnitude of t_j is governed by the behavior of F^{-1} near $1/2$. Hence after standardizing the level and slope of F at $p = 1/2$ we expect that the threshold depends very little on F.

The discussion of the last few paragraphs has been informal, but the "weak dependence of thresholds on F" can be formalized. Consider classes of smooth distributions $\mathcal{F}(M,\eta)$ defined as follows. First, the distributions have densities f symmetric about 0, so that $F^{-1}(1/2) = 0$. Second, scale is standardized so that each density obeys $f(0) = 1/\sqrt{2\pi}$, the same as the standard Gaussian $N(0,1)$. This is of course equivalent to setting $(F^{-1})'(1/2) = \sqrt{2\pi}$. Third, we impose on $F^{-1}(p)$ some regularity near

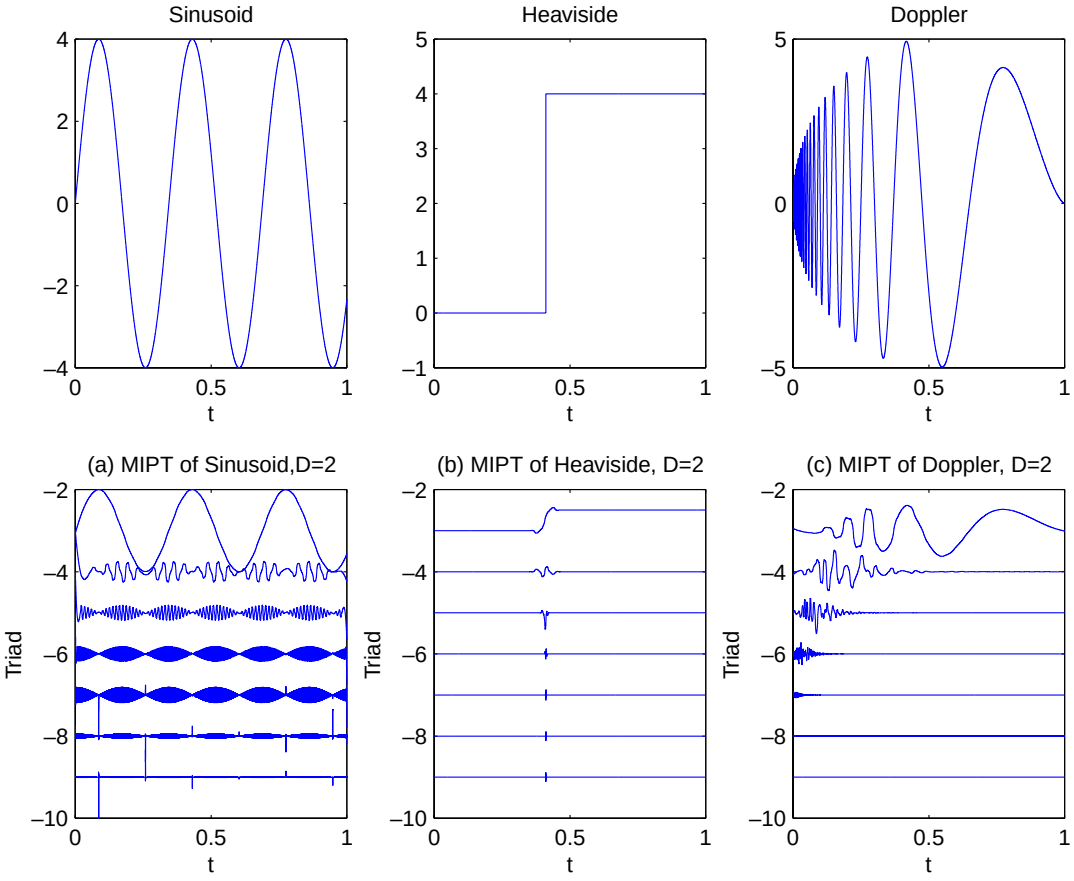

FIG. 4.2. *Nonlinear multiresolution analysis of three synthetic signals: Sinusoid, Heaviside, and Doppler. For Heaviside and Doppler, the plots of Q_j are scaled the same for all j; whereas for the Sinusoid, each Q_j is scaled by* $\max_k |(Q_j)_k|$.

$p = 1/2$: the existence of two continuous derivatives throughout $[1/2 - \eta, 1/2 + \eta]$. Our classes of symmetric distributions $\mathcal{F}(M, \eta)$ are then

$$\mathcal{F}(M,\eta) := \{F : \ f \text{ symmetric}, \ (F^{-1})'(1/2) = \sqrt{2\pi}, \ |(F^{-1})''(p)| \le M, \ |p - 1/2| \le \eta\},$$

where $M > 0$ and $0 < \eta < 1/2$ are absolute constants. The appendix proves the following theorem.

THEOREM 5.1. *For any $\epsilon > 0$ and $\theta \in (0,1)$, there exists $J^* = J^*(\epsilon, \theta, M, \eta)$ such that if $J \ge J^*$, then*

$$\max_{j \le \lfloor \theta J \rfloor} |t_j(F_1) - t_j(F_2)| \le \epsilon \qquad \textit{for all } F_1, F_2 \in \mathcal{F}(M,\eta).$$

5.2. Alpha-stable laws. Theorem 5.1 shows that a single set of MIPT thresholds can work not only for Gaussian data but also for a wide family of distributions—provided that we avoid the use of coefficients at the finest scales. To illustrate the theorem, we consider symmetric α-stable laws ($S\alpha S$) [27]. Alpha-stable laws are good models for many applications because of their high variability [24, 27].

Each symmetric α-stable law $S\alpha S$ is specified by its characteristic function $\exp(-\sigma^\alpha|\theta|^\alpha)$, with two parameters, (α,σ), $\alpha \in (0,2]$, $\sigma > 0$. The case $\alpha = 2$ is the Gaussian distribution with standard deviation $\sqrt{2}\sigma$ and density function $1/(\sqrt{2\pi}\sqrt{2}\sigma)\exp(-t^2/(4\sigma^2))$. The case $\alpha = 1$ is the Cauchy distribution with density $\sigma/(\pi(\sigma^2+t^2))$.

For our purposes, we consider $S\alpha S$ densities with σ calibrated so that the density at zero has the same value $1/\sqrt{2\pi}$ as the standard Gaussian. We denote the density and distribution of a $S\alpha S$ standardized in this way by f_α and F_α, respectively. Notice that

$$f_\alpha(t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-i\omega t}e^{-\sigma^\alpha|\omega|^\alpha}d\omega = \frac{1}{\pi}\int_0^\infty e^{-\sigma^\alpha\omega^\alpha}\cos(\omega t)d\omega \tag{5.3}$$

and therefore $f_\alpha(0) = \frac{1}{\sigma\pi}I(\alpha)$, where $I(\alpha) = \int_0^\infty e^{-\omega^\alpha}d\omega$. So f_α is properly calibrated by choosing

$$\sigma = \sigma_\alpha = \sqrt{2/\pi}\cdot I(\alpha). \tag{5.4}$$

It is clear from (5.3) that $f_\alpha(t)$ is smooth; the appendix proves the following lemma.

LEMMA 5.2. *Let $0 < \alpha_0 < 2$. $\{F_\alpha : \alpha \in [\alpha_0, 2]\}$ is a subset of a $\mathcal{F}(M,\eta)$ for appropriate M and η.*

Combining this with Theorem 5.1 gives the following corollary.

COROLLARY 5.3. *For any $\epsilon > 0$, $\theta \in (0,1)$ and $\alpha_0 \in (0,2)$, there exists $J^* = J^*(\epsilon,\theta,\alpha_0)$ such that if $J \geq J^*$, then*

$$\max_{j\leq\lfloor\theta J\rfloor}|t_j(F_{\alpha_1}) - t_j(F_{\alpha_2})| \leq \epsilon \qquad \text{for all } \alpha_0 \leq \alpha_1,\alpha_2 \leq 2.$$

To illustrate Corollary 5.3, we compare $t_j(F_2)$ with $t_j(F_1)$ in Table 5.1.

While the Gaussian and Cauchy are widely different distributions, their MIPT thresholds are very close at coarse scales.

5.3. Denoising in Gaussian noise. In order to test the above ideas, we first report on the behavior of MIPT with Gaussian noise. Figure 5.1 shows two objects—Heaviside and Doppler—contaminated by Gaussian noise.

The figure also shows the results for (a) Doppler signal, thresholding in MIPT domain; (b) Doppler signal, thresholding in AIPT domain; (c) Heaviside signal, thresholding in MIPT domain; (d) Heaviside signal, thresholding in AIPT domain.

The thresholds t_j in both cases were set by (5.2). The performance of MIPT is comparable to the performance of AIPT, as we expect.

5.4. Denoising in heavy-tailed noise. Next we report on the behavior of MIPT with Cauchy noise. Figure 5.2 shows two objects—Heaviside and Doppler—contaminated by Cauchy noise.

The figure also shows the results for (a) Doppler signal, thresholding in MIPT domain; (b) Doppler signal, thresholding in AIPT domain; (c) Heaviside signal, thresholding in MIPT domain; (d) Heaviside signal, thresholding in AIPT domain.

The thresholds t_j for MIPT thresholding were again set by (5.2). As a control experiment, the same set of thresholds were used for AIPT thresholding. The performance of MIPT is much better than the performance of AIPT, as we expect.

TABLE 5.1
MIPT thresholds for Gaussian and Cauchy white noise.

N	j	n_j	$t_j(N(0,1))$	t_j(Cauchy)	N	j	n_j	$t_j(N(0,1))$	t_j(Cauchy)
3^{10}					3^{11}				
	3	2187	6.6286	6.6764		3	6561	6.9052	6.9231
	4	729	6.6306	6.7756		4	2187	6.9059	6.9600
	5	243	6.6366	7.0866		5	729	6.9082	7.0724
	6	81	6.6543	8.1534		6	243	6.9149	7.4261
	7	27	6.7058	12.9767		7	81	6.9350	8.6533
	8	9	6.8393	67.3342		8	27	6.9928	14.4074
	9	3	7.0693	19659.0616		9	9	7.1406	88.0447
	10	1	7.2704	1.4168×10^{12}		10	3	7.3833	43575.6704
						11	1	7.5864	1.5470×10^{13}
3^{12}					3^{13}				
	3	19683	7.1696	7.1763		3	59049	7.4232	7.4257
	4	6561	7.1699	7.1900		4	19683	7.4233	7.4308
	5	2187	7.1707	7.2312		5	6561	7.4237	7.4460
	6	729	7.1732	7.3573		6	2187	7.4246	7.4918
	7	243	7.1808	7.7555		7	729	7.4274	7.6320
	8	81	7.2032	9.1531		8	243	7.4358	8.0765
	9	27	7.2676	15.9479		9	81	7.4606	9.6545
	10	9	7.4294	114.8298		10	27	7.5318	17.6099
	11	3	7.6836	96054.9354		11	9	7.7074	149.4526
	12	1	7.8869	1.6131×10^{14}		12	3	7.9718	210754.2497
						13	1	8.2100	1.5790×10^{15}

6. Appendix: Proofs.

6.1. Preliminaries. Our proofs of P3–P5 rely on two basic facts about medians and median-interpolating refinement.

LEMMA 6.1. *Let I be a closed interval and $||f-g||_{L^\infty(I)} \le \epsilon$, then $|\operatorname{med}(f|I) - \operatorname{med}(g|I)| \le \epsilon$.*

Proof. $\operatorname{med}(\cdot|I)$ is a monotone functional: $f \le g \Rightarrow \operatorname{med}(f|I) \le \operatorname{med}(g|I)$. If $||f-g||_{L^\infty(I)} \le \epsilon$, then $f \le g + \epsilon$ and

$$\operatorname{med}(f|I) \le \operatorname{med}(g+\epsilon\,|I) = \operatorname{med}(g|I) + \epsilon.$$

By symmetry, we also get $\operatorname{med}(g|I) \le \operatorname{med}(f|I) + \epsilon$. □

LEMMA 6.2. *The operators $\Pi_{(2)}, Q_{(2)} : \mathbb{R}^3 \to \mathbb{R}^3$ are Lipschitz operators, i.e., if $\mathbf{m} = (m_1, m_2, m_3)^T$ and $\mathbf{m}' = (m_1', m_2', m_3')^T$, then $||\Pi_{(2)}(\mathbf{m}) - \Pi_{(2)}(\mathbf{m}')||_\infty \le C \cdot ||\mathbf{m} - \mathbf{m}'||_\infty$ and $||Q_{(2)}(\mathbf{m}) - Q_{(2)}(\mathbf{m}')||_\infty \le C' \cdot ||\mathbf{m} - \mathbf{m}'||_\infty$.*

Proof. We focus on the proof for $\Pi_{(2)}$; the proof for $Q_{(2)}$ is similar. The closed-form expressions (2.11) and (2.14–2.17) can be used to show that $a(\mathbf{m}), b(\mathbf{m}), c(\mathbf{m})$ are globally continuous, and even that they are analytic within each of the "branches" $\mathcal{N}_1 = \{\mathbf{m} : (m_3 - m_2)/(m_2 - m_1) \in [7/3, 5]\}$, $\mathcal{N}_2 = \{\mathbf{m} : (m_3 - m_2)/(m_2 - m_1) \in [1/5, 3/7]\}$, $\mathcal{N}_3 = \{\mathbf{m} : (m_3 - m_2)/(m_2 - m_1) \in [-3, -1/3]\}$, and $\mathcal{L} = \mathbb{R}^3 - \cup_{i=1}^3 \mathcal{N}_i$. In particular, $a(\mathbf{m}), b(\mathbf{m}), c(\mathbf{m})$ have bounded partial derivatives in each of $\mathcal{N}_i$ and constant partial derivatives in $\mathcal{L}$. Hence, there is a constant $C > 0$ such that if both $\mathbf{m}, \mathbf{m}'$ belong to one of $\mathcal{N}_i$ and $\mathcal{L}$, then

$$||\left(a(\mathbf{m}), b(\mathbf{m}), c(\mathbf{m})\right)^T - \left(a(\mathbf{m}'), b(\mathbf{m}'), c(\mathbf{m}')\right)^T||_\infty \le C \cdot ||\mathbf{m} - \mathbf{m}'||_\infty. \tag{6.1}$$

It remains to show that (6.1) holds also without the restriction that $\mathbf{m}, \mathbf{m}'$ both have to belong to one of $\mathcal{N}_i$ and $\mathcal{L}$. Notice that each $\mathcal{N}_i$ is a *convex* set in $\mathbb{R}^3$ because

$$d_1 \le \frac{m_3 - m_2}{m_2 - m_1} \le d_2 \text{ and } d_1 \le \frac{m_3' - m_2'}{m_2' - m_1'} \le d_2$$

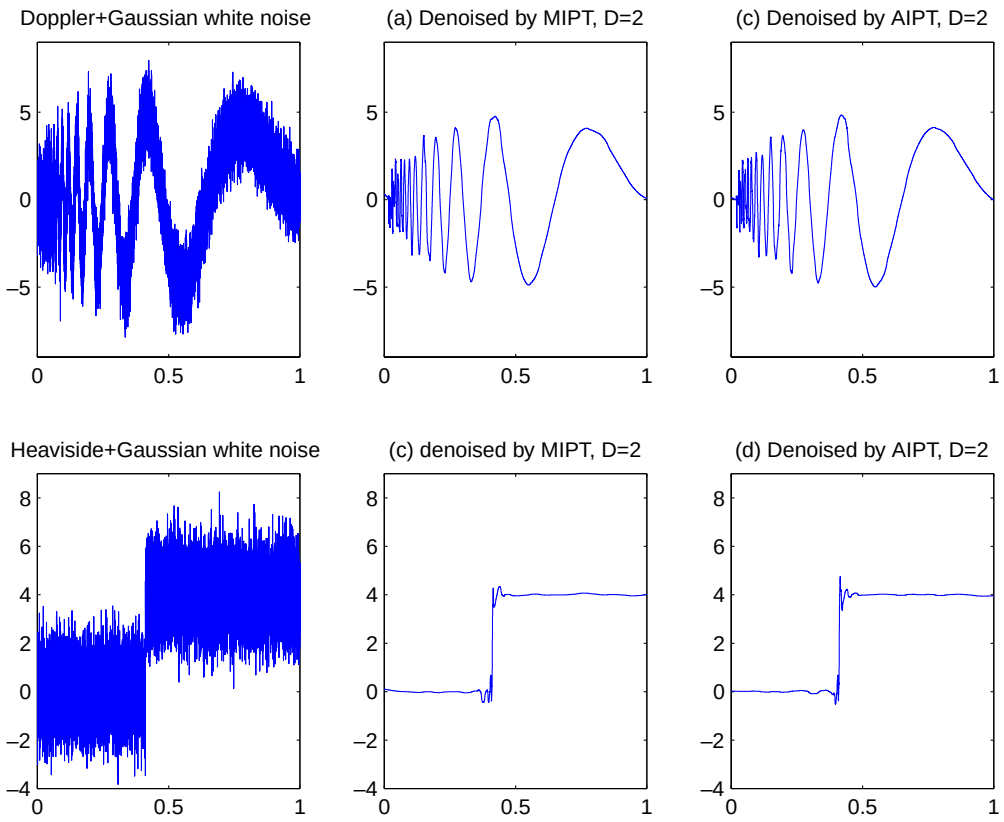

FIG. 5.1. *Denoising of Gaussian data with AIPT and MIPT thresholding.*

implies

$$d_1 \le \frac{(m_3 + t(m_3 - m_3')) - (m_2 + t(m_2 - m_2'))}{(m_2 + t(m_2 - m_2')) - (m_1 + t(m_1 - m_1'))} \le d_2 \quad \text{for all } t \in [0,1].$$

Also for each $d \in \mathbb{R}$, there is a unique $t \in \mathbb{R}$ such that $((m_3 + t(m_3 - m_3')) - (m_2 + t(m_2 - m_2')))/((m_2 + t(m_2 - m_2')) - (m_1 + t(m_1 - m_1'))) = d$. Since $\mathbb{R}$ is the disjoint union of the seven intervals $(-\infty, -3]$, $[-3, -1/3]$, $[-1/3, 1/5]$, $[1/5, 3/7]$, $[3/7, 7/3]$, $[7/3, 5]$, $[5, +\infty)$, we conclude that the line segment (in $\mathbb{R}^3$) joining $\mathbf{m}$ and $\mathbf{m}'$ is the disjoint union of at most seven subsegments each lying completely in one of the sets $\mathcal{N}_i$ and $\mathcal{L}$. Hence by replacing C by $7 \cdot C$ in (6.1) one makes the bound valid for all $\mathbf{m}$ and $\mathbf{m}'$. □

Comment. If $\mathbf{m}$, $\mathbf{m}'$ are associated with triadic intervals $I_{j,k+e}$, $e = -1, 0, 1$ and $\pi_{j,k}$ and $\pi'_{j,k}$ are the corresponding median-interpolants, then Lemma 6.2 also implies that $||\pi_{j,k} - \pi'_{j,k}||_{L^\infty(I_{j,k})} \le C||\mathbf{m} - \mathbf{m}'||_\infty$, where C is an absolute constant independent of $\mathbf{m}$, $\mathbf{m}'$, j, and k.

6.2. Proof of P3. We first recall a standard result in approximation theory.

LEMMA 6.3. *Let $f \in \dot{C}^{r+\alpha}$, $r = 0, 1, 2, \ldots$ and $0 \le \alpha \le 1$. Then there exists a constant C, proportional to the Hölder constant of f, so that for any small enough*

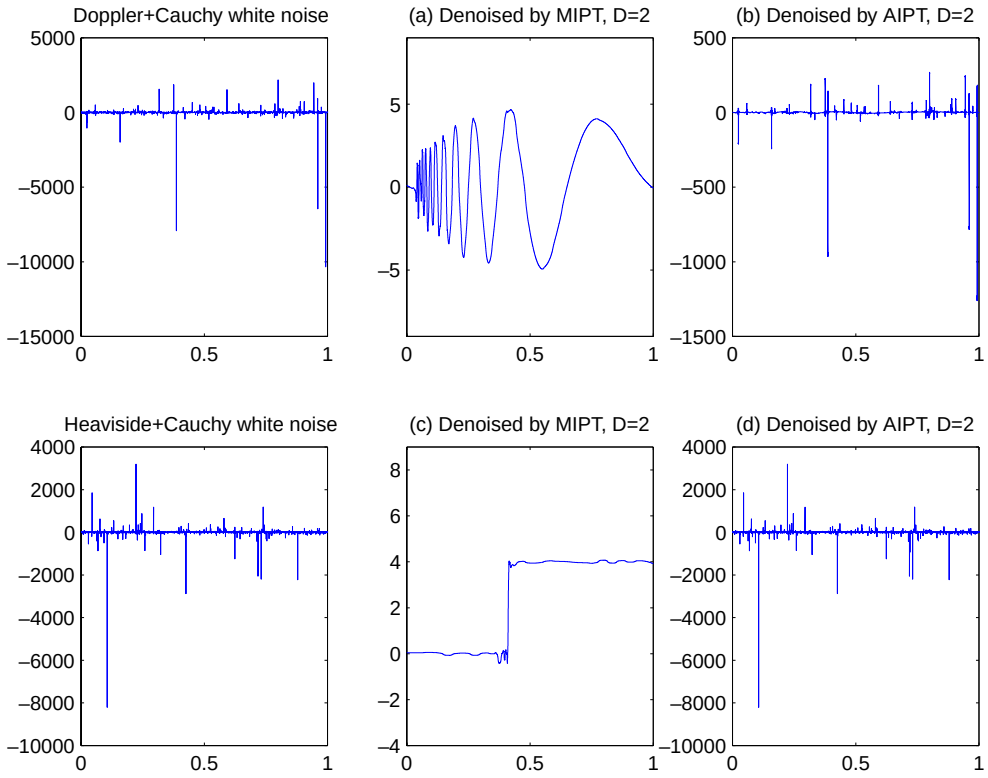

FIG. 5.2. *Denoising of Cauchy data with AIPT and MIPT thresholding.*

interval I, there is a polynomial π_I of degree r with

$$\|f - \pi_I\|_{L^\infty(I)} \le C|I|^{r+\alpha}.$$

Proof of P3. Let $f \in \dot{C}^{r+\alpha}$ ($r = 0, 1$, or 2, $0 \le \alpha \le 1$) and $I_{j,k}$ be an arbitrary triadic interval (with j large enough.) By Lemma 6.3, there exists a degree r polynomial, $\tilde{\pi}_{j,k}$, such that

$$||f - \tilde{\pi}_{j,k}||_{L^\infty(I_{j,k-1} \cup I_{j,k} \cup I_{j,k+1})} \le CC_1 3^{-(r+\alpha)j}. \tag{6.2}$$

Put for short $\epsilon = CC_1 3^{-(r+\alpha)j}$.

Recall the notation $m_{j,k} = \mathrm{med}(f|I_{j,k})$ and let $\pi_{j,k}$ be the degree $D = 2$ polynomial that interpolates the block medians $m_{j,k'}$, $k' = k-1, \ldots, k+1$. We want to show that $\pi_{j,k}$ is close to $\tilde{\pi}_{j,k}$. Denote $\tilde{m}_{j,k'} = \mathrm{med}(\tilde{\pi}_{j,k}|I_{j,k'})$, by (6.2) and Lemma 6.1,

$$|m_{j,k'} - \tilde{m}_{j,k'}| \le \epsilon \text{ for } k' = k-1, k, k+1. \tag{6.3}$$

By Lemma 6.2 and (6.3),

$$\left\|\Pi_{(2)}(m_{j,k-1}, m_{j,k}, m_{j,k+1}) - \Pi_{(2)}(\tilde{m}_{j,k-1}, \tilde{m}_{j,k}, \tilde{m}_{j,k+1})\right\|_{L^\infty(I_{j,k})} \le c\epsilon.$$

But $\Pi_{(2)}(m_{j,k-1}, m_{j,k}, m_{j,k+1}) = \pi_{j,k}$ and $\Pi_{(2)}(\tilde{m}_{j,k-1}, \tilde{m}_{j,k}, \tilde{m}_{j,k+1}) = \tilde{\pi}_{j,k}$, hence

$$||f - \pi_{j,k}||_{L^\infty(I_{j,k})} \le ||f - \tilde{\pi}_{j,k}||_{L^\infty(I_{j,k})} + ||\pi_{j,k} - \tilde{\pi}_{j,k}||_{L^\infty(I_{j,k})} \le c''\epsilon. \tag{6.4}$$

Finally, use Lemma 6.1 and (6.4) to conclude that for $e = 0, 1, 2$,

$$|d_{j+1,3k+e}| = |\operatorname{med}(f|I_{j+1,3k+e}) - \operatorname{med}(\pi_{j,k}|I_{j+1,3k+e})| \le c'''\epsilon$$

or to write it in a cleaner form, $|d_{j,k}| \le c^{(iv)} \cdot C3^{-(r+\alpha)j}$, where $c^{(iv)} = c'''3^{r+\alpha}$. □

6.3. Proof of P4 and P5. Since

$$\alpha_{j,3k+\epsilon} = m_{j,3k+\epsilon} - (Q_{(2)}(m_{j-1,k-1}, m_{j-1,k}, m_{j-1,k+1}))_\epsilon$$

and $Q_{(2)}$ is Lipschitz (Lemma 6.2), there is a constant $c > 0$, independent of j and k, such that $|\alpha_{j,3k+\epsilon}| \le c \cdot \max(|m_{j,3k+\epsilon}|, |m_{j-1,k-1}|, |m_{j-1,k}|, |m_{j-1,k+1}|)$. Boole's inequality gives for random variables W_i that

$$P(\max(W_1, \dots, W_4) > \xi) \le \sum_{i=1}^{4} P(W_i > \xi)$$

and so we can write

$$P(\sqrt{3^{J-j}}|\alpha_{j,k}| \ge \xi) \le 4 \cdot P(\sqrt{3^{J-j}}|m_{j,k}| \ge \xi/c). \tag{6.5}$$

Thus, P4 and P5 boil down to the calculation of $P(\sqrt{n}|\operatorname{med}(X_1, \dots, X_n)| \ge \xi)$ for $X_i \sim_{i.i.d.}$ Gaussian and Cauchy.

We first develop an inequality which derives from standard results in order statistics [6] and in Cramèr–Chernoff bounds on large deviations [7].

LEMMA 6.4. *Let $X_1, \dots, X_n$ be i.i.d. with cumulative distribution function (c.d.f.) $F(\cdot)$. We have the following estimate:*

$$P\{|(\operatorname{med}(X_1, \dots, X_n)| \ge x\} \le \min\left\{1, \left[2\sqrt{F(x)(1-F(x))}\right]^n + \left[2\sqrt{F(-x)(1-F(-x))}\right]^n\right\}.$$

Proof. It suffices to show that $P(\operatorname{med}(X_1, \dots, X_n) \ge x) \le \left[2\sqrt{F(x)(1-F(x))}\right]^n$ for any $x \ge 0$. Let $I_i = 1_{(X_i \ge x)}$,

$$P\left(\operatorname{med}(X_1, \dots, X_n) \ge x\right) \le P\left(\sum_{i=1}^{n} I_i \ge \frac{n}{2}\right).$$

Since $I_i \sim_{\text{i.i.d.}} \text{Binomial}(1, 1-F(x))$, $S_n := \sum_{i=1}^n I_i \sim \text{Binomial}(n, 1-F(x))$. By $1_{(S_n \ge \frac{n}{2})} \le e^{\lambda S_n} e^{-\lambda \frac{n}{2}}$ for all $\lambda > 0$, we have

$$\begin{aligned} P(S_n \ge \frac{n}{2}) &\le \min_{\lambda>0} \mathrm{E}(e^{\lambda S_n} e^{-\lambda \frac{n}{2}}) = \min_{\lambda>0} e^{-\lambda \frac{n}{2}} \mathrm{E}(e^{\lambda \sum_1^n I_i}) \\ &= \min_{\lambda>0} e^{-\lambda \frac{n}{2}} \left(F(x) + (1-F(x))e^{\lambda}\right)^n = \left[\min_{\lambda>0} F(x)e^{-\frac{\lambda}{2}} + (1-F(x))e^{\frac{\lambda}{2}}\right]^n \\ &= \left(F(x)e^{-\frac{\lambda}{2}} + (1-F(x))e^{\frac{\lambda}{2}}|_{\lambda=\ln\left(\frac{F(x)}{1-F(x)}\right)}\right)^n = \left(2\sqrt{F(x)(1-F(x))}\right)^n. \end{aligned}$$ □

COROLLARY 6.5. *Let $X_1, \dots, X_n$ be i.i.d. with c.d.f. $F(x)$ having symmetric density f. Given $\alpha \in (0, 1/2)$, define*

$$t_{\alpha,n} := F^{-1}\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 - \left(\frac{\alpha}{2}\right)^{\frac{2}{n}}}\right),$$

then

$$P\left(|\mathrm{med}(X_1,\ldots,X_n)| \geq t_{\alpha,n}\right) \leq \alpha.$$

We now apply Lemma 6.4 to the Gaussian and Cauchy distributions. Since they are both symmetric distributions, we have

$$\begin{aligned} P\left(\left|\sqrt{n}\ \mathrm{med}(X_1,\ldots,X_n)\right| \geq \xi\right) &\leq 2\cdot 2^n \sqrt{F\left(\frac{\xi}{\sqrt{n}}\right)\left(1-F\left(\frac{\xi}{\sqrt{n}}\right)\right)}^n \\ &= 2\cdot\left(\left[4F\left(y\right)\left(1-F\left(y\right)\right)\right]^{y^{-2}/2}\right)^{\xi^2} \\ &= 2\cdot\exp(\theta(y)\xi^2), \end{aligned}$$

where $y \equiv \xi/\sqrt{n}$ and $\theta(y) \equiv y^{-2}/2\cdot\log\left[4F\left(y\right)\left(1-F\left(y\right)\right)\right]$. Gaussian-type probability bounds will follow for a range $0 \leq \xi \leq X$ in P4 and P5, from an inequality $\sup_{[0,Y]}\theta(y) < 0$ on a corresponding range of values $0 \leq y \leq Y$, with $Y = X/\sqrt{n}$.

(i) Gaussian distribution: To establish P4, we need inequalities valid for $0 \leq \xi < \infty$, i.e., $0 \leq y < \infty$. Now

$$\sup_{y\in[2,\infty)}\theta(y) = \sup_{y\in[2,\infty)} y^{-2}/2\cdot\left[\log(1-F_2(y)) + \log(4F_2(y))\right].$$

From Mills' ratio $\int_y^\infty e^{-x^2/2}dx \leq \frac{1}{y}e^{-y^2/2}$ holding for all $y > 1$, we have $\log(1-F_2(y)) \leq -y^2/2 - \log(y)$. Hence

$$\sup_{y\in[2,\infty)}\theta(y) \leq \sup_{y\in[2,\infty)} y^{-2}/2\cdot\left[-y^2/2 - \log(y) + \log(4)\right] = -(2-\log(2))/2 < 0.$$

On the other hand, from symmetry of F_2 and unimodality of the density f_2 we get $4F_2(y)(1-F_2(y)) = 4F_2(y)F_2(-y) \leq 1-cy^2$ on $|y| \leq 2$, with $c > 0$; so

$$\sup_{y\in[0,2]}\theta(y) = \sup_{y\in[0,2]} y^{-2}/2\cdot\log(1-cy^2) \leq -c/2.$$

(ii) Cauchy distribution: $F_1(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\sqrt{\frac{\pi}{2}}x\right)$. To get P5 we aim only for an inequality valid on $y \in [0,Y]$, with $Y = 1$, which gives a Gaussian-type inequality for $\xi \in [0,\sqrt{n}]$.

$$\theta(y) = y^{-2}/2\cdot\log\left[1-\frac{4}{\pi^2}\arctan^2\left(\sqrt{\frac{\pi}{2}}y\right)\right] \leq -y^{-2}/2\cdot\frac{4}{\pi^2}\arctan^2\left(\sqrt{\frac{\pi}{2}}y\right).$$

However, as $\arctan(y) > c\cdot y$ for $y \in [0,1]$, with $c > 0$, this gives $\theta(y) < -\frac{c^2}{\pi}$ for $y \in [0,1]$.

6.4. Proof of Theorem 5.1. Let $p = 1/(2J3^J)^2$, $n_j = 3^{J-j}$. Let j be chosen such that

$$1/2\sqrt{1-p^{1/n_j}} \leq \eta. \tag{6.6}$$

Then there exist $0 < \eta_1, \eta_2 \leq \eta$ such that

$$\begin{aligned} |t_j(F_1) - t_j(F_2)| &= \sqrt{n_j}\left|F_1^{-1}\left(\frac{1}{2}+\frac{1}{2}\sqrt{1-p^{\frac{1}{n_j}}}\right) - F_2^{-1}\left(\frac{1}{2}+\frac{1}{2}\sqrt{1-p^{\frac{1}{n_j}}}\right)\right| \\ &= \sqrt{n_j}\left|(F_1^{-1})''(1/2+\eta_1) - (F_2^{-1})''(1/2+\eta_2)\right|\left(\frac{1}{2}\sqrt{1-p^{\frac{1}{n_j}}}\right)^2 \\ &\leq M/2\sqrt{n_j}\left(1-p^{\frac{1}{n_j}}\right), \end{aligned}$$

i.e., $|t_j(F_1) - t_j(F_2)| \le \epsilon$ if

$$\sqrt{n_j}\left(1 - p^{\frac{1}{n_j}}\right) \le \frac{2\epsilon}{M}. \tag{6.7}$$

Since $(1-x) \le \log(1/x)$ for all $x \in (0,1]$, (6.6) holds for large enough J because

$$\begin{aligned}\frac{1}{2}\sqrt{1-p^{1/n_j}} &\le \frac{1}{2}\sqrt{\log p^{-1/n_j}} = \frac{1}{2}\sqrt{\frac{1}{n_j}\log(4\cdot J^2\cdot 3^{2J})} \\ &\le \frac{1}{2}\sqrt{\frac{1}{3^{(1-\theta)J}}(\log(4)+2\log(J)+2J\log(3))} \to 0 \text{ as } J\to 0.\end{aligned}$$

Similarly, (6.7) holds for large enough J because

$$\begin{aligned}\sqrt{n_j}\left(1-p^{\frac{1}{n_j}}\right) &\le (1/\sqrt{n_j})\log p^{-1} \\ &\le \frac{1}{\sqrt{3^{(1-\theta)J}}}(\log(4)+2\log(J)+2J\log(3)) \to 0 \text{ as } J\to 0. \qquad \square\end{aligned}$$

6.5. Proof of Lemma 5.2. It suffices to find M, η such that

$$\sup_{\alpha\in[\alpha_0,2]}\sup_{p\in[\frac{1}{2}-\eta,\frac{1}{2}+\eta]}(F_\alpha^{-1})''(p) \le M.$$

Since

$$(F_\alpha^{-1})''(p) = -\frac{f_\alpha'(F_\alpha^{-1}(p))}{[f_\alpha(F_\alpha^{-1}(p))]^3},$$

we work with F_α^{-1}, f_α, and f_α' separately.

1. Since $|F_\alpha^{-1}(p)|$ is monotone increasing in p for fixed α, and is monotone increasing in α for fixed p, we have, for any $0<\eta<1/2$,

$$\sup_{0<\alpha\le 2}\sup_{1/2-\eta\le p\le 1/2+\eta}|F_\alpha^{-1}(p)| = F_2^{-1}(1/2+\eta).$$

2. Now $\sup_{|t|\le\epsilon_1}|f_\alpha(t)-f_\alpha(0)| \le |t|\cdot\{\sup_{|t|\le\epsilon_1}|f_\alpha'(t)|\}$. Also

$$\begin{aligned}\sup_{|t|\le\epsilon_1}|f_\alpha'(t)| &= \sup_{|t|\le\epsilon_1}\left|\frac{1}{\pi}\int_0^\infty e^{-\sigma_\alpha^\alpha\omega^\alpha}(-\omega)\sin(\omega t)d\omega\right| \\ &\le \frac{1}{\pi}\int_0^\infty e^{-\sigma_\alpha^\alpha\omega^\alpha}\omega d\omega = \frac{1}{\pi}\frac{1}{\sigma_\alpha^2}\int_0^\infty e^{-\omega^\alpha}\omega d\omega \le C_1(\alpha_0),\end{aligned}$$

where $C_1(\alpha) = \frac{1}{2}(\int_0^\infty e^{-\omega^\alpha}\omega d\omega)/(\int_0^\infty e^{-\omega^\alpha}d\omega)^2$ is defined on $(0,2]$ and is positive and monotone decreasing.

3. Similarly $\sup_{|t|\le\epsilon_1}|f_\alpha'(t)-f_\alpha'(0)| \le |t|\cdot\{\sup_{|t|\le\epsilon_1}|f_\alpha''(t)|\}$. Moreover

$$\begin{aligned}\sup_{|t|\le\epsilon_1}|f_\alpha''(t)| &= \sup_{|t|\le\epsilon_1}\left|\frac{1}{\pi}\int_0^\infty e^{-\sigma_\alpha^\alpha\omega^\alpha}(\omega^2)\cos(\omega t)d\omega\right| \\ &\le \frac{1}{\pi}\int_0^\infty e^{-\sigma_\alpha^\alpha\omega^\alpha}\omega^2 d\omega = \frac{1}{\pi}\frac{1}{\sigma_\alpha^3}\int_0^\infty e^{-\omega^\alpha}\omega^2 d\omega \le C_2(\alpha_0),\end{aligned}$$

where $C_2(\alpha) = \frac{1}{2}\sqrt{\frac{\pi}{2}}(\int_0^\infty e^{-\omega^\alpha}\omega^2 d\omega)/(\int_0^\infty e^{-\omega^\alpha} d\omega)^3$ is defined on $(0, 2]$ and is positive and monotone decreasing. Therefore

$$\begin{aligned}\sup_{\alpha\in[\alpha_0,2]} \sup_{p\in[\frac{1}{2}-\eta,\frac{1}{2}+\eta]} |(F_\alpha^{-1})''(p)| &= \sup_{\alpha\in[\alpha_0,2]} \sup_{p\in[\frac{1}{2}-\eta,\frac{1}{2}+\eta]} \frac{|f'_\alpha(F_\alpha^{-1}(p))|}{|f_\alpha(F_\alpha^{-1}(p))|^3} \\ &\le \frac{|F_2^{-1}(\frac{1}{2}+\eta)|C_2(\alpha_0)}{\left|\frac{1}{\sqrt{2\pi}} - F_2^{-1}(\frac{1}{2}+\eta)C_1(\alpha_0)\right|^3}. \end{aligned} \tag{6.8}$$

If we choose $\eta = \eta(\alpha_0) > 0$ small enough such that $F_2^{-1}(\frac{1}{2}+\eta)C_1(\alpha_0) < \frac{1}{\sqrt{2\pi}}$, then with the choice of $M = M(\alpha_0)$ defined by (6.8) we get $\{F_\alpha : \alpha_0 \le \alpha \le 2\} \subset \mathcal{F}(M(\alpha_0), \eta(\alpha_0))$. □

Reproducible research. In this paper, all computational results are reproducible, meaning that the code which generated the figures is available over the Internet, following the discipline indicated in [3]. Interested readers are directed to http://www-stat.stanford.edu/~wavelab/.

Acknowledgments. The authors would like to thank Andrew Bruce, Amir Dembo, Gary Hewer, and Charles Micchelli for helpful discussions and references.

REFERENCES

[1] B. Borden and M. Mumford, *A statistical glint/radar cross section target model*, IEEE Trans. Aerospace Electron. Systems, 19 (1983), pp. 781–785.
[2] A. G. Bruce, D. L. Donoho, H.-Y. Gao, and R. D. Martin, *Denoising and robust nonlinear wavelet analysis*, in SPIE Proceedings, Wavelet Appl. 2242, Orlando, FL, 1994.
[3] J. Buckheit and D. L. Donoho, *Wavelab and reproducible research*, in Wavelets in Statistics, A. Antoniadis and G. Oppenheim, eds., Springer-Verlag, New York, 1994, pp. 55–82.
[4] I. Daubechies, *Ten Lectures on Wavelets*, CBMS-NSF Regional Conf. Ser. in Appl. Math., SIAM, Philadelphia, 1992.
[5] I. Daubechies and J. Lagarias, *Two-scale difference equations* II. *Local regularity, infinite products of matrices and fractals*, SIAM J. Math. Anal., 23 (1992), pp. 1031–1079.
[6] H. A. David, *Order Statistics*, Wiley, New York, London, Sydney, 1970.
[7] A. Dembo and O. Zeitouni, *Large Deviations and Applications*, Bartlett and Jones, Boston, 1993.
[8] G. Deslauriers and S. Dubuc, *Symmetric iterative interpolation processes*, Constr. Approx., 5 (1989), pp. 49–68.
[9] D. L. Donoho, *Interpolating Wavelet Transforms*, tech. report, Department of Statistics, Stanford University, Stanford, CA, 1992. Available at http://www-stat.stanford.edu/~donoho/Reports/1992/interpol.ps.Z
[10] D. L. Donoho, *Smooth wavelet decompositions with blocky coefficient kernels*, in Recent Advances in Wavelet Analysis, L. Schumaker and G. Webb, eds., Academic Press, Boston, 1993, pp. 259–308.
[11] D. L. Donoho, *Minimum entropy segmentation*, in Wavelets Theory, Algorithms and Applications, C. Chui, L. Montefusco, and L. Puccio, eds., Academic Press, San Diego, 1994, pp. 233–269.
[12] D. L. Donoho, *De-noising by soft-thresholding*, IEEE Trans. Inform. Theory, 41 (1995), pp. 613–27.
[13] D. L. Donoho and I. M. Johnstone, *Adapting to unknown smoothness by wavelet shrinkage*, J. Amer. Statist. Assoc., 90 (1995), pp. 1200–1224.
[14] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, *Wavelet shrinkage: Asymptopia?*, J. Roy. Statist. Soc. Ser. B, 57 (1995), pp. 301–369.
[15] D. L. Donoho and T. P.-Y. Yu, *Nonlinear "Wavelet Transforms" Based on Median-Interpolation*, tech. report, Department of Statistics, Stanford University, Stanford, CA, 1997. Available at http://www-stat.stanford.edu/~donoho/Reports/1997/median.ps.Z.
[16] N. Dyn, J. Gregory, and D. Levin, *Analysis of uniform binary subdivision schemes for curve design*, Constr. Approx., 7 (1991), pp. 127–147.

[17] H.-Y. Gao., *Wavelet Estimation of Spectral Densities in Time Series Analysis*, Ph.D. Thesis, Department of Statistics, University of California, Berkeley, CA, 1993.
[18] T. N. T. Goodman and T. P.-Y. Yu, *Interpolation of medians*, Adv. Comput. Math., 11 (1999), pp. 1–10.
[19] D. Greer, I. Fung, and J. Shapiro, *Maximum-likelihood multiresolution laser radar range imaging*, IEEE Trans. Image Process., 6 (1997), pp. 36–47.
[20] P. Hall and P. Patil, *On the choice of smoothing parameter, threshold and truncation in nonparametric regression by wavelet methods*, J. Roy. Statist. Soc. Ser. B, 58 (1996), pp. 361–377.
[21] F. R. Hampel, *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York, 1986.
[22] P. Huber, *Robust Statistics*, Wiley, New York, 1981.
[23] H. Longbotham, *A class of robust nonlinear filters for signal decomposition and filtering utilizing the Haar basis*, in ICASSP-92, Vol. 4, IEEE Signal Processing Society, Piscataway, NJ, 1992.
[24] C. L. Nikias and M. Shao, *Signal Processing with Alpha-Stable Distributions and Applications*, Wiley, New York, 1995.
[25] O. Rioul, *Simple regularity criteria for subdivision schemes*, SIAM J. Math. Anal., 23 (1992), pp. 1544–1576.
[26] D. F. F. R. L. de Queiroz and R. Schafer, *Nonexpansive pyramid for image coding using a non-linear filter bank*, IEEE Trans. Image Process., 7 (1998), pp. 246–252.
[27] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random Process: Stochastic Models with Infinite Variance*, Chapman & Hall, New York, 1994.
[28] J. L. Starck, F. Murtagh, and A. Bijaoui, *Image Processing and Data Analysis: The Multiscale Approach*, Cambridge University Press, Cambridge, UK, 1998.
[29] J. Starck, F. Murtagh, B. Pirenne, and M. Albrecht, *Astronomical image compression based on noise suppression*, Proc. Astro. Soc. Pac., 108 (1996), pp. 446–455.
[30] B. Stuck and B. Kleiner, *A statistical analysis of telephone noise*, Bell System Tech. J., 53 (1974), pp. 1263–1320.
[31] T. P.-Y. Yu, *New Developments in Interpolating Wavelet Transforms*, Ph.D. thesis, Program of Scientific Computing and Computational Mathematics, Stanford University, Stanford, CA, 1997.