Robust Nonlinear Wavelet Transform based on Median-Interpolation

David L. Donoho & Thomas P.Y. Yu
Department of Statistics
Stanford University
Stanford, CA 945305, USA
donoho@stat.stanford.edu
pyu@stat.stanford.edu

Abstract

It is well known that wavelet transforms can be derived from stationary linear refinement subdivision schemes [2, 3]. In this paper we discuss a special nonlinear refinement subdivision scheme – median-interpolation. It is a nonlinear cousin of Deslauriers-Dubuc interpolation and of average-interpolation. The refinement scheme is based on constructing polynomials which interpolate median functionals of the underlying object. The refinement scheme can be deployed in multiresolution fashion to construct nonlinear pyramid schemes and associated forward and inverse transforms. In this paper we discuss the basic properties of this transform and its possible use in wavelet de-noising schemes for badly non-Gaussian data. Analytic and computational results are presented to show that the nonlinear pyramid has very different performance compared to traditional wavelets when coping with non-Gaussian data.

1. Introduction

A number of recent theoretical studies [6, 5] 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

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

where $(z_i)_{i=1}^n$ is a Gaussian white noise. In this setting, they show that one removes a noise successfully by applying a wavelet transform, then applying thresholding in the wavelet domain, then inverting the transform.

A key principle underlying this approach is the fact that an orthogonal transform takes Gaussian white noise into Gaussian white noise. It follows from this, and the Gaussian assumption on the noise z, that in the transform domain, all wavelet coefficients suffer noise with the same probability distribution; and so some form of homogeneous treatment (i.e. thresholding all coefficients in a standard way) is plausible.

Now consider (1) in cases when the white noise $(z_i)_{i=1}^n$ is highly non-Gaussian. An example of such setting arises in target tracking in radar [9, 13], in which one encounters data contaminated by *impulsive glint noise*. As a model, we consider $(z_i)_{i=1}^n$ to be i.i.d. 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. It therefore offers an extreme example of non-Gaussian data.

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, the wavelet transform of i.i.d. Cauchy noise does not result in independent, nor identically distributed wavelet coefficients. In fact, coefficients at larger scales are more likely to be affected by the perturbing influence of the few large noise values, and so have a systematically larger dispersion at large scales. Invariance of distribution under orthogonal transforms and $O(\sqrt{\log(n)})$ behavior of maxima are fundamental to the results of [5, 6]. This extremely non-Gaussian situation lacks key quantitative properties which were used in de-noising in the Gaussian case.

In this paper we study a wavelet-like approach to deal with such extremely non-Gaussian data which is based on abandoning linear orthogonal transforms and using instead a kind of *nonlinear* wavelet transform called *Median-Interpolating Pyramid Transform* (MIPT.) The nonlinear transforms we propose combine ideas from the theory of robust estimation [10, 8] and ideas from wavelets and the theory of interpolating subdivisions [1, 7]. The forward transform (FMIPT) is based on deploying the *median* in a multiresolution pyramid; this gives an analysis method which is highly robust against extreme observations. The transform itself is based on computation of “block medians” over triadic cells, and the recording of block medians in a “résumé-detail" array. The inverse transform (IMIPT) is based on deploying *median-interpolation* in a multiresolution fashion. At its center is the notion of "median-interpolating refinement scheme", a two-scale refinement scheme based on imputing behavior at finer scales from coarse-scale information.

Our method of building transforms from refinement schemes is similar to the way interpolating wavelet transforms are built from Deslauriers-Dubuc interpolating schemes in [2] and in the way biorthogonal wavelet transforms are built from average-interpolating refinement in [3]. However, despite structural similarities there are 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).

It would be more accurate to call the approach a *nonlinear pyramid transform*.

2. Median-Interpolating Refinement

In this section, we describe the mechanism of *median-interpolating refinement*, which is the major building block of MIPT.

Given a function f on an interval I, let $\mathrm{med}(f|I)$ denote a median of f for the interval I, for definiteness we follow John Tukey's suggestion and use the lowest median, defined by $\mathrm{med}(f|I) = \inf\{\mu : m(t \in I : f(t) \geq \mu) \geq m(t \in I : f(t) \leq \mu)\}$. 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 $I_{j,k} = [k3^{-j}, (k+1)3^{-j})$: $m_{j,k} = \mathrm{med}(f|I_{j,k}) \quad 0 \leq k \leq 3^j$. 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 will 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*:

$$\mathrm{med}(\pi_{j,k}|I_{j,k+l}) = m_{j,k+l} \quad \text{for} \quad -A \leq l \leq A. \tag{2}$$

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

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

Unlike Lagrange interpolation or average-interpolation, median-interpolation is a nonlinear procedure and we are not aware of any algorithms available in the literature. In the following we briefly describe our attempt in solving this problem.

$D = 0$. It is the simplest by far; in that case one is fitting a constant function $\pi_{j,k}(t) = Const$. Hence $A = 0$, and (2) becomes $\pi_{j,k}(t) = m_{j,k}$. The imputation step (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.

$D > 0$. Let $I_i = [i-1, i]$, $x_i = i - 1/2$, $i = 1, \ldots, (D+1)$. Define the map $F : \mathbb{R}^{D+1} \to \mathbb{R}^{D+1}$ by $(F(\mathbf{v}))_i = \mathrm{med}(\pi|I_i)$ where π is the polynomial specified by the vector $\mathbf{v} \in \mathbb{R}^{D+1}$ and the Lagrange interpolation basis,

$$\pi(x) = \sum_{i=1}^{D+1} v_i l_i(x) \text{ where } l_i(x) = \prod_{j=1, j\neq i}^{D+1} \frac{(x - x_j)}{(x_i - x_j)}. \tag{4}$$

We have the following bisection algorithm to apply F to a given vector v.

Algorithm $m =$`BlockMedian`**(**v,*tol***):**

1. $m^+ = \max(\pi|I)$, $m^- = \min(\pi|I)$, $m = \pi(\frac{a+b}{2})$, where π is formed from v as in (4).
2. Calculate all the roots of $\pi(x) - m$, collect the ones that are (i) real, (ii) inside (a,b) and (iii) with odd multiplicities into the ordered array $\{r_1, \ldots, r_k\}$ (i.e. $r_i < r_{i+1}$). Let $r_0 = a$ and $r_{k+1} = b$.
3. If $\pi(a) \geq m$, $M^+ = \sum_{i \text{ odd}} r_i - r_{i-1}$ and $M^- = \sum_{i \text{ even}} r_i - r_{i-1}$; else $M^- = \sum_{i \text{ odd}} r_i - r_{i-1}$ and $M^+ = \sum_{i \text{ even}} r_i - r_{i-1}$.
4. If $|M^+ - M^-| < tol$, terminate, else if $M^+ > M^- + tol$, $m^- = m, m = (m + m^+)/2$, goto 2 ; else $m^+ = m, m = (m + m^-)/2$, goto 2.

To search for a degree D polynomial with prescribed block median $\{m_i\}$ at $\{I_i\}$, we need to solve the nonlinear system

$$F(\mathbf{v}) = \mathbf{m}. \tag{5}$$

There are general nonlinear solvers that can be used to solve (5). We recommend below a more tailored method which is based on fixed point iteration.

The main empirical observation is that median-interpolation can be well approximated by Lagrange interpolation. We first observe that $\mathrm{med}(\pi|[a,b]) \approx \pi((a+b)/2)$. The above approximation is exactly, for example, when π is monotonic on $[a,b]$, which is the case when $[a,b]$ is away from the vicinity of the extremum points of π. The linear approximation can be combined with the idea of fixed point iteration to give the following iterative algorithm for median-interpolation.

Algorithm v =`MedianInterp`(m, tol)**:**

1. *Input.* $\mathbf{m}$ - vector of prescribed median at intervals $I_i = [i-1, i], i = 1, \ldots, (D+1)$.
2. *Initialization.* Set $\mathbf{v}^0 = \mathbf{m}$.
3. *Iteration.* Iterate $\mathbf{v}^{k+1} = \mathbf{v}^k + \big(\mathbf{m} - F(\mathbf{v}^k)\big)$ until $||\mathbf{m} - F(\mathbf{v}^k)||_\infty < tol$.
4. *Output.* $\mathbf{v}$ - solution of (5).

Experiments suggest that the above algorithm has an exponentially fast convergence rate, which is an expectable phenomenon due to the contraction mapping theorem. However it requires more work to obtain a complete proof of convergence.

$D = 2$: A surprise. Although algorithm `MedianInterp` can be applied to any $D = 2, 4, 6, \ldots$, it suffers from two pitfalls. First, it is an iterative procedure and is harder to implement in practical setting (e.g. in DSP hardware.) Second, more analysis on the convergence of the algorithm is needed to be carried out. These problems go away completely in the special case of $D = 2$, because closed-form solution of median-interpolating refinement can be carried out in the quadratic case. The reason of our success in obtaining such closed-form expressions is due to the fact that quadratic polynomials all have a simple shape, namely, each of them has a unique extremum point and is symmetric and monotonic on each side of the extremum. This geometric property happens to make manipulation with medians much easier, and hence a closed-form solution can be derived. For the formulae and details and proofs, see [14].

For the rest of the paper, we will mainly focus on the $D = 2$ case, although most of the results can be extended to higher degrees.

2.1 Multiscale Refinement

The two-scale refinement scheme described in the last section applied to an initial median sequence $\{\tilde{m}_{j_0,k}\} = \{m_{j_0,k}\}$ implicitly defines a (generally nonlinear) refinement operator $R_{MI} = R\ R(\{\tilde{m}_{j,k}\}) = \{\tilde{m}_{j+1,k}\} \qquad j \geq j_0$. We can associate resulting sequences $(\tilde{m}_{j,k})_k$ to piecewise constant functions on the line via $\tilde{f}_j(\cdot) = \sum_{k=-\infty}^{\infty} \tilde{m}_{j,k}\, 1_{I_{j,k}}(\cdot)$ for $j \geq j_0$. We can recursively apply this operator, and associate the sequence of piecewise constant functions defined on successively finer and finer meshes.

In case $D = 0$, we have $\tilde{f}_{j_0+h} = f_{j_0}$ 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 more than a direct application of theory in linear refinement schemes. In [14], it was shown that $\{\tilde{f}_j(\cdot)\}_j$ defined above converges always to a uniformly continuous limit. Moreover, the limit is Hölder continuous and a lower bound of the Hölder exponent was derived.

3. Median-Interpolating Pyramid Transform

We now apply the refinement scheme to construct a nonlinear pyramid and associated nonlinear multiresolution analysis.

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 $y_i, i = 0, ..., n-1$ where $n = 3^J$, is a triadic number, such as $729, 2187, 6561, \ldots$. 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 $m_{j,k} = \mathrm{med}(y_i : i/n \in I_{j,k})$.
3. *Formation of Refinements.* Calculate $\tilde{m}_{j,k} = R((m_{j-1,k}))$ 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 [14]. Important details for the implementation concern the treatment of boundary effects and efficient calculation of block medians.

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 Median-Interpolating Pyramid Transform 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.

4. Properties

The pyramid just described has several basic 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

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

It follows that $Dim(\theta) = 3^J(1+1/3+1/9+\ldots) \sim 3/2\cdot n$. 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

$$|\alpha_{j,k}| \le C' C 3^{-j\alpha}. \quad (6)$$

If 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 $MIPT\ D = 2$,

$$|\alpha_{j,k}| \le C' C 3^{-j(r+\alpha)}. \quad (7)$$

P4. Gaussian Noise. Suppose that $y_i = \sigma z_i$, $i = 0, \ldots, n-1$, and that z_i is i.i.d Gaussian white noise. Then

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

where the $C_i > 0$ are absolute constants.

While *P1* and *P2* are rather obvious, *P3* and *P4* require proof. See [14].

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.

A key property of MIPT but *not* linear wavelet transforms 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 Cauchy white noise. Then

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

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

For typical wavelet transforms, the coefficients of Cauchy noise have Cauchy distributions, and such exponential bounds cannot hold. In contrast, it is known that the median of a sample of Cauchy random variables obeys exponential inequalities over a wide range of arguments. Again, see [14] for proof of *P5*.

5. De-Noising via MIPT Thresholding

We now consider applications of pyramid transforms to multi-scale de-noising. In general, we act as we would in the wavelet de-noising case.

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

In this approach, coefficients 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 [4], 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 is very successful, the thresholds can be set 'as if' the noise is Gaussian and the transform is the AIPT, even when the noise is very non-Gaussian. This would mean that the use of the median cancels out many of the 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 $\sqrt{3^{J-j}}\alpha_{j,k}$ is the L^2-normalized MIPT coefficients of the 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 as $J \to \infty$. It was derived in [14] that, when F is a symmetric law, a valid choice is

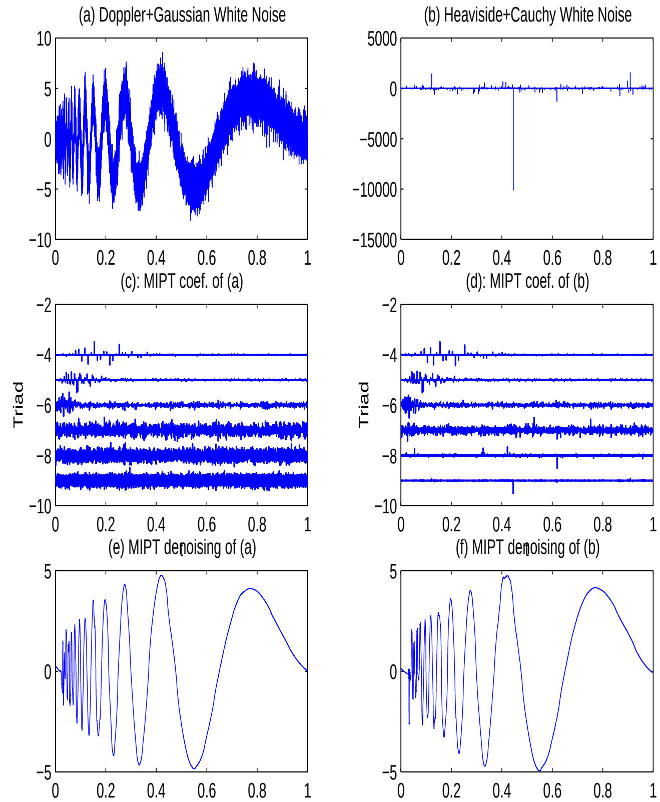
$$t_j := t_j^F = \sqrt{3^{J-j}}\, F^{-1}\left(\frac{1}{2} + \frac{1}{2}\sqrt{1 - \left(\frac{1}{2J3^J}\right)^{\frac{2}{3^{J-j}}}}\right). \qquad (8)$$

We claim that the thresholding rule (8) does not depend much on F despite its appearance. For simplicity we will consider only distributions that are symmetric. Since $\alpha^{1/3^{J-j}}$ increases very rapidly to 1 as j decreases, (8) suggests that as long as we are away from the finest scales, the magnitude of t_j is pretty much governed by the behavior of F^{-1} at $1/2$. This observation can be turned into the following thoerem.

Theorem 1 *Let* $\mathcal{C}(M,\eta) := \{F : (F^{-1})'(1/2) = \sqrt{2\pi},\ |(F^{-1})''(p)| \le M \quad \forall 1/2 - \eta \le p \le 1/2 + \eta\}$ *where* $M > 0$ *and* $0 < \eta < 1/2$ *are absolute constants. For any* $\epsilon > 0$ *and* $\theta \in (0,1)$*, there exists* $J^* = J^*(\epsilon, \theta, M, \eta)$ *such that if* $J \ge J^*$ *then* $\max_{j \le \lfloor \theta J \rfloor} |t_j^{F_1} - t_j^{F_2}| \le \epsilon$, $\forall F_1, F_2 \in \mathcal{C}(M,\eta)$.

Proof. See [14].

It was also shown that alpha stable laws and, in particular, Cauchy and Gaussian distributions, are members of $\mathcal{C}(M,\eta)$ (with M and η appropriately calibrated.) The following figure shows a Doppler signal contaminated by Gaussian noise and Cauchy noise in Panel (a) and (b) respectively. Panel (c) and (d) depict their MIPT coefficients plotted scale by scale. Notice the similar distributions in magnitude when comparing Panel (c) and (d) in the three corasest scales; and the distinguish differences when comparing the two finest ones – a phenomenon described mathematically by Theorem 1. In Panel (e) and (f) we show the denoised results by MIPT thresholding, the thresholds t_j in both cases were set by (8). The performance in the Gaussian case (Panel (a) and (c)) is comparable to results obtained by orthonormal wavelet thresholding. However, in the Cauchy case there is no hope to obtain results comparable to that in Panel (d) using a simple thresholding procedure in linear (orthogonal or bi-orthogonal) wavelet transform domains.



References

[1] I. Daubechies. *Ten Lectures on Wavelets*. Number 61 in CBMS-NSF Series in Applied Mathematics. SIAM, Philadelphia, 1992.

[2] D. L. Donoho. Interpolating wavelet transform. Technical Report 408, Department of Statistics, Stanford University, 1992. Available at ftp://stat.stanford.edu/reports/donoho/interpol.ps.Z.

[3] D. L. Donoho. Smooth wavelet decompositions with blocky coefficient kernels. In L. Schumaker and G. Webb, editors, *Recent Advances in Wavelet Analysis*, pages 259–308. Boston: Academic Press, 1993.

[4] D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. on Information Theory.*, 41(3):613–27, 1995.

[5] D. L. Donoho and I. M. Johnstone. Adapting to unknown smoothness by wavelet shrinkage. Preprint Department of Statistics, Stanford University, 1992.

[6] D. L. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage: Asymptopia? *Journ. Roy. Stat. Soc.*, B, 1995.

[7] N. Dyn, J. Gregory, and D. Levin. Analysis of uniform binary subdivision schemes for curve design. *Constr. Approx.*, 7:127–147, 1991.

[8] F. R. Hampel. *Robust Statistics : the approach based on influence functions*. Wiley, 1986.

[9] G. A. Hewer, R. D. Martin, and J. Zeh. Robust prepocessing for kalman filtering of glint noise. *IEEE Trans. on Aerospace and Electron Systems*, AES-23(1):120–128, Jan. 1987.

[10] P. Huber. *Robust Statistics*. Wiley, 1981.

[11] C. L. Nikias and M. Shao. *Signal Processing with Alpha-Stable Distributions and Applications*. Wiley, 1995.

[12] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Process: Stochastic Models with Infinite Variance*. Chapman & Hall, 1994.

[13] W.-R. Wu. Target tracking with glint noise. *IEEE Trans. on Aerospace and Electron Systems*, 29(1):174–185, Jan. 1993.

[14] T. P.-Y. Yu. *New Developments in Interpolating Wavelet Transforms*. PhD thesis, Program of Scientific Computing and Computational Mathematics, Stanford University, August 1997. Available at http://www-sccm.stanford.edu/Students/yu/thesis.ps.gz.