On “Multi-Resolution Nonparametric Regression for Spatially Adaptive Image De-Noising”

Project for ELE856
Yanjun Yan, Fall 2004

Reading List

Outline
- Wavelet and MRA review
- Wavelet image de-noising
- Local Polynomial Approximation (LPA) based de-noising [1],[2],[3]
- Experimental results

Wavelet and MRA Review
- Haar wavelet: the simplest in form, yet very good in understanding the concept
- Definitions of MRA
- Several other wavelets with analytical expressions
- Daubechies wavelet: the construction of wavelets from its required properties

Haar Wavelet Family
- Father wavelet \( \phi(t) = \begin{cases} 
1 & 0 \leq t < 1 \\
0 & \text{else}
\end{cases} \)

Father wavelet is also called the **scaling function** for the wavelet family. All other family members can be obtained by combining the scalings and translations of the father wavelet.

Haar Wavelet Family
- Mother wavelet can be derived from the father wavelet

\[
\psi(t) = \phi(2t) - \phi(2t-1) = \begin{cases} 
1 & 0 \leq t < \frac{1}{2} \\
-1 & \frac{1}{2} \leq t < 1 \\
0 & \text{else}
\end{cases}
\]
The Haar wavelet family can be viewed as the basis of an inner product space. Originally, $\phi(t)$, a rank-1 or $V_0$ space with father & mother

$\phi(t) \rightarrow [1,1]$, $\psi(t) \rightarrow [1,-1]$.

The inner product $\int_0^1 \phi(t) \psi(t) dt = 0$ Thus $\phi(t), \psi(t)$ are orthogonal. The space is extended to $V_1$ space.

Daughter wavelets are derived from the mother wavelet

$\psi_{2,0}(t) = \psi(2t)$

with $0 < t < \frac{1}{2}$

$\psi_{2,1}(t) = -\psi(2t-1)$

with $0 < t < 1$

$\psi_{2,2}(t) = \psi(2t-1)$

with $\frac{1}{2} < t < 1$

$\psi_{2,3}(t) = -\psi(2t-2)$

with $0 < t < \frac{1}{2}$

Similarly, we can extend the family, and also the dimension of the space, by more and more daughters:

For the $n$th generation, there are $2^n$ daughters $\psi_{2^n,0}(t) = \psi(2^n t - k)$, $k = 0, 1, ..., 2^n - 1$.

The family members (within the same generation or between generations) are orthogonal to each other.

Let’s come back to the $V_2$ space spanned by $\{\phi(t), \psi(t), \psi_{1,0}(t), \psi_{1,1}(t)\}$

$$
\begin{bmatrix}
\phi(t) \\
\psi(t) \\
\psi_{1,0}(t) \\
\psi_{1,1}(t)
\end{bmatrix} = 
\begin{bmatrix}
1 & 1 & 1 & 1 \\
1 & 1 & -1 & -1 \\
1 & -1 & 0 & 0 \\
0 & 0 & 1 & -1
\end{bmatrix}
$$

A nature basis for such $V_2$ space is

$$
\begin{bmatrix}
\phi_{2,0}(t) \\
\phi_{2,1}(t) \\
\phi_{2,2}(t) \\
\phi_{2,3}(t)
\end{bmatrix} = 
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{bmatrix}
$$

which are the sons of the 2nd generation.
Haar Wavelet Family

- For the $n$th generation, there are $2^n$ sons.
  \[ \{ \phi_{n,k}(t) = \phi(2^n t - k) \} \quad k = 0,1,...,2^n - 1 \]
- These sons of the $n$th generation alone can span the $V_n$ space.
- If by parents and daughters, we need both parents and all generations of the daughters up to the $(n-1)$th generation to span the $V_n$ space.
- Sons and Daughters are orthogonal too.

Haar Wavelet Family

- Let's take a look on the relations between daughters and sons.
- Given the $n$th generation of sons, we can span the $V_n$ space. If we want to extend it into $V_{n+1}$ space, we may use the orthogonal decomposition rule to obtain the difference between the current space and the higher level space.

Haar Wavelet Family

- Example: $V_1 = V_0 \oplus V_1$
  Given $V_0$ spanned by $\{ \phi_0(t), \psi_0(t) \}$
  By orthogonal decomposition, we can derive that $V_1 \leftrightarrow \{ \psi_{10}(t), \psi_{11}(t) \}$
  There are several possible wavelet decomposition
  $V_2 \leftrightarrow \{ \phi(t), \psi(t), \phi_{10}(t), \psi_{10}(t) \}$
  $\leftrightarrow \{ \phi_0(t), \phi_1(t), \phi_{10}(t), \psi_{11}(t) \}$
  $\leftrightarrow \{ \phi_0(t), \phi_0(t), \phi_{10}(t), \psi_{11}(t) \}$

Haar Wavelet Decomposition

- With the Haar wavelet family, we can approximate an arbitrary $f(x) \in L^1([0,1))$ on Haar family's compact support $[0,1)$ with an arbitrary precision given as many generations as needed.
- By scaling and translation, we can get more general compact support in $R$.  

Multi-Resolution Analysis (MRA)

- In order to approximate any function as closely as we like by functions in the Haar spaces $V_n$, we need to have access to wavelets $\psi_{m,k}$ in $V_n$ for arbitrarily large $n$. Rather than examine $V_n$ for each $n$ to determine the precision of an approximation, it is useful to consider the space $V = \bigcup_{n \in \mathbb{N}} V_n$.  

Dense

- A subset \( B \) of \( A \) is dense in \( A \) if any given element in \( A \) can be approximated as closely as we like by an element in \( B \).
- Example: the set of rational numbers \( \mathbb{Q} \) is dense in real number set \( \mathbb{R} \).
- \( V = \bigcup_{n \in \mathbb{Z}} V_n \) is dense in \( L^2[\mathbb{R}] \), which is the collection of functions \( f : \mathbb{R} \to \mathbb{R} \) such that \( \|f\|^2 = (\int_{\mathbb{R}} f(t)^2 \, dt)^2 < \infty \).

Separation Property

- Whenever we have a nested sequence of sets satisfying \( \ldots \subseteq V_2 \subseteq V_1 \subseteq V_0 \subseteq V \subseteq \ldots \) of sets satisfying \( \bigcap_{n \in \mathbb{Z}} V_n = \{0\} \), the collection \( \{V_n\} \) is said to have the separation property.
- The spaces spanned by generations of wavelets are separable.

Definition of MRA

- A MRA is a nested sequence of subspaces of \( L^2(\mathbb{R}) \) with a scaling function \( \phi(t) \) such that
  1. \( V = \bigcup_{n \in \mathbb{Z}} V_n \) is dense in \( L^2[\mathbb{R}] \),
  2. \( \bigcap_{n \in \mathbb{Z}} V_n = \{0\} \),
  3. \( f(t) \in V_n \) iff \( f(2^{-n}t) \in V_n \),
  4. \( \phi(t-k) \) is an orthonormal basis for \( V_k \).

Dilation Equation

- For most wavelets used in practice, there’s no simple formula for the scaling function. Instead a critical property of each scaling function follows from (4) and (3).
- Since \( \{\phi(2^{-k}t)\} \) is an orthonormal basis for \( V_0 \), the set \( \{\phi(2^{-k}t)\} \) is an orthonormal basis for \( V_k \). It means \( \phi(t) \in V_k \) can be written in the form \( \phi(t) = \sum c_k \phi(2^{-k}t-k) \) with \( c_k \) are the refinement coefficients.
- The dilation equation of the Haar wavelets
  \[
  \phi(t) = \phi(2t) + \phi(2t-1)
  \]

Relation between \( \phi(t) \) and \( \psi(t) \)

- Normalized refinement coefficient \( h_0 = \frac{1}{\sqrt{2}} \)
  \[
  \phi(t) = \sum c_k \phi(2t-k) = \sum h_k \sqrt{2} \phi(2t-k) \]
- Suppose \( \psi(t) = \sum g_k \sqrt{2} \phi(2t-k) \).
- With some derivation
  \[
  \langle \phi(t), \psi(t) \rangle = \sum h_k g_k = 0 \\
  \langle \psi(t-k), \psi(t-m) \rangle = \sum h_k h_{2t-2k-m} = \delta_{km}
  \]
- The generally accepted relation is
  \[
  g_k = (-1)^k h_k 
  \]
- which is wavelet function coefficient.

Quadradic Battle-Lemarie Scaling Function (one B-spline)

- \[
  \phi(t) = \begin{cases} 
  \frac{1}{8} t^2 & 0 \leq t < 1 \\
  -t^2 + \frac{1}{2} & 1 \leq t < 2 \\
  \frac{1}{2} (t-3)^2 & 2 \leq t < 3 \\
  0 & \text{otherwise}
  \end{cases}
  \]
- Its dilation equation is
  \[
  \phi(t) = \frac{1}{2} \phi(2t) + \frac{1}{2} \phi(2t-1) + \frac{1}{2} \phi(2t-2) + \frac{1}{2} \phi(2t-3)
  \]
- The corresponding mother wavelet is
  \[
  \psi(t) = \frac{1}{2} \phi(2t) - \frac{1}{2} \phi(2t+1) + \frac{1}{2} \phi(2t-2) - \frac{1}{2} \phi(2t-1)
  \]
- The \( \phi(t-k) \) is not orthogonal basis for space \( V_k \). It violates condition (4).
Mexican Hat (Maar) Wavelet

- There’s a simple representation of the mother wavelet, but there’s no simple form for the scaling function.
  \[ \psi(t) = (1 - t^2)e^{-\frac{1}{2}t^2} \]

Morlet Wavelet

- It’s widely used in time-frequency analysis.

![Morlet Wavelet Image]

Shannon Scaling Function

- \( \phi(t) = \begin{cases} \frac{\sin(\pi t)}{\pi t} & t \neq 0 \\ 1 & t = 0 \end{cases} \)

In contrast to Haar father wavelets, it is smooth - all of its derivatives exist and continuous - but its support is all of \( \mathbb{R} \).

- We want to seek a wavelet family with compact support and some sort of smoothness in between of the two extreme cases.

Daubechies Wavelets

- Dr. Ingrid Daubechies

  "I guess we just have to live with the way they look"

  example above: D4

Daubechies Wavelets

- Compact Support Condition
  \[ \phi(t) = c_0 \phi(2t) + c_1 \phi(2t-1) + c_2 \phi(2t-2) + c_3 \phi(2t-3) \]

  Correspondingly,
  \[ \psi(t) = c_0 \phi(2t + 2) - c_1 \phi(2t + 1) + c_2 \phi(2t) - c_3 \phi(2t - 1) \]

- Orthogonality Condition (Parseval’s Theorem)
  \[ \sum c_k^2 = c_0^2 + c_1^2 + c_2^2 + c_3^2 = 2 \]  
  \[ \sum c_k c_{k+2m} = c_0 c_{2m} + c_1 c_{2m} + 0 \quad \forall m, m = 1 \text{ here} \]  
  \[ \sum c_k^2 = 2(\phi(1), \phi(2t-k)) \]

Daubechies Wavelets

- Regularity Condition (by and dilation equation)
  \[ c_0 + c_1 + c_2 + c_3 = 2 \]  
  \[ \int_0^1 \phi(t) dt = \frac{1}{2} \int_0^1 \phi(t) dt \]  
  Notice \[ \sum c_k^2 = 0 \text{ here} \]  
  \[ \sum c_k c_{k+2m} = c_0 c_{2m} + c_1 c_{2m} + 0 \quad \forall m, m = 1 \text{ here} \]  
  \[ \sum c_k^2 = 2(\phi(1), \phi(2t-k)) \]
**Daubechies Wavelets**

- Moment Condition (mother wavelet has vanishing moments)
  
  \[
  \int_{-\infty}^{\infty} \psi(t) dt = 0, \quad \int_{-\infty}^{\infty} t \psi(t) dt = 0.
  \]
  
  Therefore
  
  \[
  -c_0 + c_1 - c_2 + c_3 = 0 \quad (4)
  \]
  
  \[
  -2c_1 + c_2 - c_3 = 0 \quad (5)
  \]
  
  Notice
  
  \[
  \int_{-\infty}^{\infty} t \phi(2t - k) dt = 4 \int_{-\infty}^{\infty} \phi(t) dt + 4 \int_{-\infty}^{\infty} \phi(t) dt.
  \]

- With the refinement coefficients
  
  \[
  c_4 = \frac{\sqrt{2}}{\sqrt{3}}, c_2 = \frac{\sqrt{2}}{\sqrt{3}}, c_0 = \frac{\sqrt{2}}{\sqrt{3}}, c_1 = \frac{\sqrt{2}}{\sqrt{3}}.
  \]
  
  \[
  \phi(t) = c_4 \phi(2t) + c_2 \phi(2t - 1) + c_0 \phi(2t - 2) + c_1 \phi(2t - 3).
  \]

  The D4 wavelet is approximated as below by the Cascade Algorithm.

**High and Low Pass Filters**

- \(\phi(t) = \sum h_k \sqrt{2} \phi(2t - k)\)
  
  \(\psi(t) = \sum g_k \sqrt{2} \phi(2t - k)\)

- The sequences \(h_k\) and \(g_k\) above can be used to process signals and are typically called **lowpass** and **highpass** filters respectively.

- In our textbook, \(h_k(n) \leftrightarrow \{h_k\}\) \(g_k(n) \leftrightarrow \{g_k\}\)

  Refer to eg 7.10 on page 382.

**2D Image Transform**

- Implement the transform row-wise and column-wise.

**Wavelet Image De-Noising**

- Peak SNR 20dB

- Improvement of SNR

  ISNR 3.9715dB

  ISNR 4.5798dB
Local Polynomial Approximation

- Model the noise corrupted image as 
  \[ g(x) = f(x) + \alpha(x) \]
  where \( x = (x, y) \)
- The underlying true image is \( f(x) \)
- Assume \( f(x) \) belongs to a nonparametric class of piecewise continuous \( m \)-differentiable functions
- The objective of the de-noising is to minimize the point-wise mean squared error risk

The following criteria function is applied in LPA

\[ J_s(x) = \sum w_s(x-x)[g(x) - C^\top \phi(x-x)]^2 \]
where \( w_s(t) = w(t/h) / h \), the window function, formalizes the localization of fitting with respect to the center \( t \), while the scale \( h \) determines the size of the window.
- \( w(t) \geq 0, \max_{t} w(t), \int w(x, y) dx \, dy = 1 \)

\[ f = \hat{C}(x, h) \]
\[ \hat{C}(x, h) = \arg \min_{C} J_s(x) \]
Gives \( f(x) = \hat{C}(x, h) \) as an estimate of \( f(x) \), and \( \hat{C}_k(x, h), k = 2, \ldots, M \) are used in the estimates of the derivatives of \( f(x) \).

For the regular infinite grid and \( x \) belonging to this grid, (6) & (7) can be written as a homogeneous transform

\[ f(x, h) = \sum r_i(x, x_x, h) g(x) \]
where \( r_i \) is the first element of the vector \( r \) given by the equation

\[ r = \Phi^{-1} w_s(x-x)[\phi(x-x)]^2 \]

\[ \Phi = \sum w_s(x-x) [\phi(x-x)]^2 \]

It can be verified that for any polynomial \( f(x) \) of the power \( m \), the estimate above is accurate.
Especially \( \sum r_i(x, x_x, h) [\phi(x-x)]^2 = 0 \) for \( k = 1, \ldots, M \)

\[ f(x, h) = \sum \phi(x-x) g(x) \]
\[ r = \Phi^{-1} w_s(x-x)[\phi(x-x)]^2 \]

\[ \Phi = \sum w_s(x-x)[\phi(x-x)]^2 \]
Local Polynomial Approximation

- The window size selection is a crucial point of the efficiency of the local estimators.
- The best choice of scale involves a trade-off between the bias and variance of the estimator.

Optimal scale in [2]

- Paper [2] has derived an expression for the optimal scale by minimizing the MSE risk and [2] also gives the upper bound on the risk, but it needs lots of information on unknown parameters.
- Then [2] uses the Intersection of Confidence Intervals (ICI) rule to iteratively determine the scale from set

\[ H = \{h_0 > h_1 > ... > h_j\} \]

by only the estimate and its variance.

Example in [2]

- LPA de-noising with ICI adaptive scale

Example in [2]

- Small windows are dark, large windows are bright
- Window sizes delineate the true image

Nonparametric Spectrum & MRA

- Assume \( H \) be a finite or infinite set of positive scales

\[ H = \{h_0 > h_1 > ... > h_j\} \]

starting from \( h_0 \) and decreasing to small \( h_j \).
- For the smallest scale \( h_J \), the LPA kernel \( r_J(x) \) gives the identical operator

\[ \lim_{h \to 0} r_J(x) = \delta_{x,0} \]

- In [2], find a range from \( h_0 \) to a small \( h_j \)
- In [1], combine all estimates in all scales

Nonparametric Spectrum & MRA

- By Gram-Shmidt orthogonalization, the linearly independent kernels for different scales can be transformed to

\[ Qw(x) = r(x) \]

where \( r(x) = [r_{h_0}(x), r_{h_1}(x), ..., r_{h_J}(x)]^T \)

\[ w(x) = [w_{h_0}(x), w_{h_1}(x), ..., w_{h_J}(x)]^T \]

\( w_{h_j}(x) \) are orthonormal, \( Q \) is nonsingular lower triangular matrix

\[ \hat{f}(x, h) = \sum r_{h_j}(x, h) g(x - x_j) \] (8) Reprint from p42
Nonparametric Spectrum & MRA

- For any $f(x) \in L^2(\mathbb{R})$
- $\alpha_j(x) = (w_j \circ f)(x)$
  
  Defines a spectral analysis with components varying from coarse scale base image $\alpha_0$ to finer scale image $\alpha_j$.
- $f(x) = \sum_{j=0}^{J} \alpha_j(x)w_j(0)$

It's shown in [1] that above formulae yield accurate MR spectral expansion.

Pros

- The scale defined by the set $H$ are quite arbitrary ⇔ the dyadic scale is a crucial point of the wavelets implementation.
- Only the inter-scale orthogonality is used ⇔ the wavelets require both inter-scale and intra-scale orthogonality.
- Extension to multi-dimension is easy.

"In contrast"

LPA Nonparametric De-Noising

- True spectra $\alpha_j(x) = (w_j \circ f)(x)$
- Noisy estimate $\hat{\alpha}_j(x) = (w_j \circ g)(x)$
- $\hat{\alpha}_j(x) = \alpha_j(x) + n_j(x)$
  
  where $n_j(x) = (w_j \circ e)(x)$ is normal noise and $n_j$ are uncorrelated for different scale $j = 0,1,..., J$.
- Goal: estimate $\alpha_j(x)$ from $\hat{\alpha}_j(x)$ and use the estimate to reconstruct the image.

As in wavelet MRA, the common underlying assumption is that the function to estimate has some redundancy, then the heuristic for the use is that the expansion of such a function in some basis is sparse.

So one tends to estimate large coefficients and discard the small ones.

Thresholding is used in getting the estimates of the LPA coefficient to reconstruct image (de-noising).

$\tilde{f}(x) = \sum_{j=0}^{J} \gamma_j \hat{\alpha}_j(x)w_j(0)$

LPA Nonparametric De-Noising

(1) Hard thresholding $\gamma_j = 1(\hat{\alpha}_j(x) > t\sigma)$
  
  where $t$ is certain threshold, $\sigma$ is noise variance. $1(x) = 1$

(2) Soft thresholding $\gamma_j = (1 - \frac{t\sigma}{|\hat{\alpha}_j(x)|})$. $\gamma_j$ where $(x)_+ = x$, if $x > 0$
(3) Stein’s thresholding
\[
y_j = (1 - t\sigma / |\tilde{\alpha}_j(x)|^2).
\]
(4) Blockwise Stein’s thresholding
\[
y_j = (1 - t\sigma / |\tilde{h}_j(x)|^2),
\]
where \(\tilde{h}_j(x)\) is the moving mean of \(\tilde{\alpha}_j(x)\) in pixel’s square neighborhood.

LPA Nonparametric De-Noising

LPA Nonparametric De-Noising

LPA Nonparametric De-Noising

Summary

- Wavelet theory review
- LPA estimate introduction
- LPA scale selection [2]
- LPA multi-scale based perfect reconstruction [1]
- De-noising by different thresholding schemes [1]

References

The Wavelet Transform (WT)

The Wavelet Transform (WT) is defined as

$$WT_a^d(x) = \int \frac{x(t)}{a^d} \psi \left( \frac{t - a^d}{a} \right) dt$$

$\psi(\cdot/a)$ here is the analysis wavelet.

The WT preserves time shift and time scaling, but it doesn’t preserve frequency shift:

$$\tilde{x}(t) = x(t-t_0) \Rightarrow WT_a^d(t, f) = WT_a^d(t-t_0, f)$$

Set analysis scale $a = \gamma$, then

$$\tilde{x}(t) = \sqrt{a} x(at) \Rightarrow WT^d_a(t, f) = WT^d_a(at, \gamma)$$

Signal 1 in simulation

$$x(t) = \delta(t-t_0) + \delta(t-t_2) + e^{i2\pi ft_0} + e^{i2\pi ft_2}$$

$$X(f) = e^{-i2\pi ft_0} + e^{-i2\pi ft_2} + \delta(f-f_1) + \delta(f-f_2)$$

Parameters:

$t_0 = 20 s \quad t_2 = 80 s \quad f_1 = 0.2 f_s \quad f_2 = 0.4 f_s$

STFT of signal 1 in Contour Format

WT of Signal 1