

Signal Denoising with Wavelets

Selin Aviyente
Department of Electrical and Computer Engineering
Michigan State University

March 30, 2010

Introduction

- Assume an additive noise model: $x[n] = f[n] + w[n]$
- The goal is to estimate f from x and minimize the estimation error.
- We will consider estimators computed from an expansion of the signal in an orthogonal basis, in our case the wavelet transform.
- Wavelet transform of a noisy signal yields small coefficients that are dominated by noise, large coefficients carry more signal information.
- White noise is spread out equally over all coefficients.
- Orthogonal wavelets have a decorrelation property. Wavelet transform of white noise is white.

Wavelet Transform of Noisy Signal

- Write wavelet transform as a matrix, B , $B(x = f + w)$.
- Assume w is white, Gaussian noise with covariance, K .
- Covariance of transformed noise
 $Bw: S = E[(Bw)(Bw)^T] = BE[ww^T]B^T = BKB^T$.
- For white noise, $K = \sigma^2 I$, then $S = \sigma^2 I$.
- Since noise remains white in all bases, it does not influence the choice of basis. When the noise is not white, the noise can have an important impact on the basis choice.

Diagonal Operator Estimation

- A diagonal operator estimates independently each $f_B[m]$ from $x_B[m]$ with a function $d_m(x)$. The resulting estimator is:

$$\tilde{F} = DX = \sum_{m=0}^{N-1} d_m(X_B[m])g_m \quad (1)$$

where g_m s are the elements of the basis set, B .

- We will first consider estimators of the form $d_m(X_B[m]) = a[m]X_B[m]$ for $0 \leq m < N$ where $a[m]$ depends on $X_B[m]$. The operator D is linear when $a[m]$ is a constant.

Attenuation with Oracle

- Find $a[m]$ that minimizes the risk $r(D, f)$ of the estimator where the risk is the MSE:

$$\begin{aligned} r(D, f) &= \sum_{m=0}^{N-1} E[|f_B[m] - X_B[m]a[m]|^2] \\ &= |f_B[m]|^2(1 - a[m])^2 + \sigma^2 a^2[m] \end{aligned} \quad (2)$$

- The risk is minimum for $a[m] = \frac{|f_B[m]|^2}{|f_B[m]|^2 + \sigma^2}$.
- This is not practical since $a[m]$ depends on $f_B[m]$ and in practice we don't know it. Therefore, this estimator gives a lower bound on the risk which is not reachable.

Projection with Oracle

- The nonlinear projector that minimizes the risk is defined by:

$$a[m] = \begin{cases} 1 & \text{if } |f_B[m]| \geq \sigma \\ 0 & \text{if } |f_B[m]| < \sigma \end{cases} \quad (3)$$

- This projector cannot be implemented since $a[m]$ depends on $|f_B[m]|$.
- The risk of this oracle projector is $\sum_{n=0}^{N-1} \min(|f_B[m]|^2, \sigma^2)$.

Thresholding Estimator

- Hard Thresholding: A hard thresholding estimator is implemented with

$$d_m(x) = \rho_T(x) = \begin{cases} x & |x| > T \\ 0 & |x| \leq T \end{cases} \quad (4)$$

- The risk of this thresholding is :

$$|f_B[m] - \rho_T(X_B[m])|^2 = \begin{cases} |W_B[m]|^2 & |X_B[m]| > T \\ |f_B[m]|^2 & |X_B[m]| \leq T \end{cases} \quad (5)$$

- The risk is larger than the risk of an oracle projector.

Soft Thresholding

- Decrease by T amplitude of all noisy coefficients:

$$d_m(x) = \rho_T(x) = \begin{cases} x - T & x \geq T \\ x + T & x \leq -T \\ 0 & |x| \leq T \end{cases} \quad (6)$$

- The threshold is chosen that there is a high probability that it is just above the maximum level of noisy coefficients.

Theorem (Donoho and Johnstone)

- Let $T = \sigma\sqrt{2\log N}$. The risk $r_t(f)$ of a hard or soft thresholding estimator satisfies for all $N \geq 4$,
$$r_t(f) \leq (2\log N + 1)(\sigma^2 + r_p(f)).$$
- This is known as the universal threshold and is optimal for white Gaussian noise.

Thresholding Refinements

- SURE (Stein Unbiased Risk Estimator) Thresholds: To study the impact of the threshold on the risk, denote by $r_t(T, f)$ the risk of a soft thresholding with threshold T . An estimate $\tilde{r}_t(f, T)$ is calculated from the noisy data and T is optimized by minimizing $\tilde{r}_t(f, T)$.
- If $|X_B[m]| < T$ then the coefficient is set to zero and the risk is equal to $|f_B[m]|^2$. One can estimate this with $|X_B[m]|^2 - \sigma^2$.
- If $|X_B[m]| \geq T$, the expected risk is the sum of the noise energy plus the bias introduced by the reduction of the amplitude of $X_B[m]$ by T ($\sigma^2 + T^2$).
- The resulting estimator $\tilde{r}_t(f, T) = \sum_{m=0}^{N-1} \Phi(|X_B[m]|^2)$ with

$$\Phi(u) = \begin{cases} u - \sigma^2 & u \leq T^2 \\ \sigma^2 + T^2 & u > T^2 \end{cases}$$

SURE

- For a soft thresholding, the risk estimator is unbiased, $E[\tilde{r}_t(f, T)] = r_t(f, T)$.
- To find the threshold that minimizes the SURE estimator, the N data coefficients $X_B[m]$ are sorted in decreasing amplitude order. Let $X_B^r[k]$ be the coefficient of rank k . Let l be the index such that $|X_B^r[l]| \geq T > |X_B^r[l+1]|$, then
$$\tilde{r}_t(f, T) = \sum_{k=1}^N |X_B^r[k]|^2 - (N-l)\sigma^2 + l(\sigma^2 + T^2)$$
- To minimize the risk, choose the smallest possible $T, |X_B^r[l+1]|$.

Noise Variance Estimation

- We need to estimate the noise variance from the data $X_B[m]$.
- The signal X of size N has $N/2$ wavelet coefficients at the finest scale. The coefficient $|\langle f, \psi_{l,m} \rangle|$ is small if f is smooth over the support of $\psi_{l,m}$ in which case $\langle X, \psi_{l,m} \rangle \approx \langle W, \psi_{l,m} \rangle$. $|\langle f, \psi_{l,m} \rangle|$ is large if f has a sharp transition. A piecewise regular signal has few sharp transitions and hence produces a small number of large coefficients.
- At the finest scale, the signal f thus influences the value of a small portion of large amplitude coefficients $\langle X, \psi_{l,m} \rangle$. All others are approximately equal to $\langle W, \psi_{l,m} \rangle$ which are independent Gaussian random variables of variance σ^2 .

- A robust estimator of σ^2 is estimated from the median of $\{\langle X, \psi_{l,m} \rangle\}_{0 \leq m < N/2}$.
- If M is the median of the absolute value of P independent Gaussian random variables of zero-mean and variance σ^2 then it can be shown that $E[M] \approx 0.6745\sigma$.
- Therefore, the variance of the noise is estimated from the median of the wavelet coefficients, $\tilde{\sigma} = \frac{M}{0.6745}$.

Hard vs. Soft Thresholding

- Soft thresholding is better for removing noise but also reduces the amplitude of the signal and distorts the discontinuities in the signal.
- Hard thresholding is better at preserving discontinuities but may not be the best for removing noise.
- To obtain comparable SNR values, the threshold of the soft thresholding must be about half the size of the hard thresholding one.

Multiscale Thresholds

- Piecewise regular signals have a proportion of large coefficients that increases when the scale decreases. A singularity creates the same number of large coefficients at each scale, whereas the total number of coefficients increases when the scale increases.
- Adapt the threshold to the scale.
- At low scale, the threshold T_j should be smaller to avoid setting to zero too many large amplitude signal coefficients.
- At each scale, a different threshold is calculated from the noisy coefficients using SURE.

Best Basis Thresholding

- If we wish to choose a basis adaptively, we must use a higher threshold T than the universal threshold $\sigma\sqrt{2\log N}$ used when the basis is set.
- For a dictionary including P distinct vectors, $T = \sigma\sqrt{2\log P}$, $P > N$.