SHIFT INVARIANT WAVELET-BASED STATISTICAL MODELS AND $1/F$ PROCESSES

Robert D. Nowak

Michigan State University, East Lansing, MI, 48824-1226, nowak@egr.msu.edu

ABSTRACT

Two new approaches to signal processing based on the discrete wavelet transform (DWT) have recently received great attention. *Gaussian mixture densities* have been used to model non-Gaussian wavelet coefficient distributions, and show great promise in multiscale statistical signal and image processing [1]. *Shift-invariant DWTs* are very useful for signal processing since the usual shift dependency of the DWT is circumvented [2, 3]. This paper unifies these two ideas in a novel Bayesian modeling framework and draws a connection between non-Gaussian wavelet coefficient models and $1/f$ processes.

1. HAAR WAVELET-BASED MULTISCALE MODELS

Wavelet-based multiscale models exploit two key properties of the DWT [1]. The *Compression* property suggests that wavelet coefficients tend to have peaky, heavy-tailed distributions. The *Decorrelation* property has been used to justify modeling wavelet coefficients as independent random variables. Based on these two properties, several wavelet-based models have been developed in which the wavelet coefficients are modeled as independent random variables with Gaussian mixture densities (*e.g.,* [1]). This model is called the DWT independent mixture model (DWT-IMM). The DWT-IMM leads to very simple and elegant statistical signal processing algorithms.

The Haar DWT-IMM is described as follows. Let $\{\omega_{i,k}\}$ denote the Haar wavelet coefficients of a length 2^j signal s, where $0 \leq i \leq j$ denotes the scale and k denotes the position of the associated wavelet function. Each coefficient is modeled as a two-component, zero-mean, Gaussian mixture

$$\omega_{i,k}|\epsilon_{i,k} \sim \epsilon_{i,k} N(0, \sigma^2_{i,L}) + (1-\epsilon_{i,k}) N(0, \sigma^2_{i,H}) \qquad (1)$$

where $N(0,\sigma^2)$ denotes a Gaussian density with variance σ^2, and $\epsilon_{i,k}$ is a Bernoulli random variable with $\Pr(\epsilon_{i,k} = 1) = p_i$ and $\Pr(\epsilon_{i,k} = 0) = 1 - p_i$. Setting $\sigma^2_{i,L} \ll \sigma^2_{i,H}$, produces a peaky, heavy-tailed distribution representative of the Compression property of the DWT.

Supported by the National Science Foundation, grant no. MIP–9701692.

Now suppose that we observe this signal in additive Gaussian white noise with variance σ^2, and let $\{d_{i,k}\}$ denote the Haar wavelet coefficients of the noisy observation. It follows from (1) that the observed coefficients (of the signal plus noise) are distributed

$$d_{i,k}|\epsilon_{i,k} \sim \epsilon_{i,k} N(0, \sigma^2_{i,L} + \sigma^2) + (1-\epsilon_{i,k}) N(0, \sigma^2_{i,H} + \sigma^2)$$

This model suggests a Bayesian approach to wavelet denoising. Wavelet denoising refers to the well-known wavelet thresholding and shrinkage methods of noise removal [2]. Using the DWT-IMM as a prior for the signal, straightforward calculations show that the minimum mean-square error estimator of the wavelet coefficient $\omega_{i,k}$ is

$$\mathrm{E}[\omega_{i,k}|d_{i,k}] = \left(\frac{q_1 \sigma^2_{i,L}}{\sigma^2_{i,L} + \sigma^2} + \frac{q_0 \sigma^2_{i,H}}{\sigma^2_{i,H} + \sigma^2} \right) d_{i,k} \qquad (2)$$

where

$$\begin{aligned} q_1 &= \Pr(\epsilon_{i,k} = 1|d_{i,k}) \\ &= \frac{p_i f(d_{i,k}|\epsilon_{i,k} = 1)}{p_i f(d_{i,k}|\epsilon_{i,k} = 1) + (1-p_i) f(d_{i,k}|\epsilon_{i,k} = 0)} \\ q_0 &= \Pr(\epsilon_{i,k} = 0|d_{i,k}) \\ &= \frac{(1-p_i) f(d_{i,k}|\epsilon_{i,k} = 0)}{p_i f(d_{i,k}|\epsilon_{i,k} = 1) + (1-p_i) f(d_{i,k}|\epsilon_{i,k} = 0)} \end{aligned}$$

$f(d_{i,k}|\epsilon_{i,k} = 1)$ is $N(0, \sigma^2_{i,L} + \sigma^2)$ and $f(d_{i,k}|\epsilon_{i,k} = 0)$ is $N(0, \sigma^2_{i,H} + \sigma^2)$. The estimator (2) is parameterized by $\{\sigma^2_{i,L}, \sigma^2_{i,H}, p_i\}_{i=0}^{j}$, and σ^2, and it can be viewed as a Bayesian approach to wavelet denoising. The parameters can be directly estimated from the data using an EM algorithm [1].

A major drawback to using the DWT-IMM as prior for the signal is that the DWT is not a shift-invariant transform, and hence the signal prior is non-stationary. Not only is this undesirable from a modeling perspective, but also the non-stationarity of the DWT-IMM can lead to blocking artifacts in the signal reconstruction based on the wavelet coefficient estimates in (2).

2. SHIFT-INVARIANT MODELS

The shift-invariant DWT (SIDWT) introduced in [2, 3] was developed to reduce artifacts introduced in wavelet denoising schemes due to the shift-dependent nature of the DWT. The shift-invariant denoising approaches are quite simple. In effect, the shift-invariant denoising schemes simply denoise all possible circular shifts of the data independently, then shift the denoised signals back to the original alignment, and average the results. This can be very efficiently carried out using the undecimated wavelet transform [3].

The DWT-IMM and the SIDWT are easily unified in a new Bayesian framework. Specifically, the shift of the data can be viewed as an additional degree of freedom in the model and a probability measure on the shift parameter can be introduced. It is assumed that the DWT is the periodic (cyclic) DWT and all shifts are circular. Let n denote the shift of the data with respect to the wavelet system. If we regard the original signal as "unshifted" ($n = 0$) then the standard DWT-IMM is denoted $f(s|0)$. A SIDWT-IMM is obtained by placing a uniform prior distribution on the shift parameter. Let $p(n) = \frac{1}{N}$ denote the uniform probability mass function of the shift, where N is the length of the signal. The SIDWT-IMM is then given by

$$f(s) = \sum_n f(s|n)p(n) \tag{3}$$

The SIDWT-IMM provides shift-invariant (stationary) signal priors. Furthermore, shift-invariant signal estimation is easily accomplished by computing the undecimated DWT of the noisy signal, computing the Bayes estimates of each wavelet coefficient according to (2), and then computing the inverse undecimated DWT (for details on the undecimated DWT see [3]).

3. CORRELATION FUNCTIONS OF MULTISCALE MODELS

The Gaussian mixture densities model the heavy-tailed character of wavelet coefficient distributions. However, from a signal modeling perspective we are interested also in the signal structure generated by the DWT-IMM and SIDWT-IMM. This structure is revealed through a careful analysis of the autocorrelation functions of the two models.

Let $c_{i,k}$ and $\omega_{i,k}$ denote the Haar scaling and wavelet coefficient, respectively, of length $N = 2^j$, 1-d, discrete-time signal s. Again, $0 \leq i \leq j$ denotes the scale and k denotes the position of the associated scaling function or wavelet. Let us adopt the convention that $i = 0$ is the coarsest scale and $i = j$ denotes the finest scale of analysis. This means that the signal samples are equal the scaling coefficients at scale j; that is, $c_{j,k} = s(k)$. Assume that the signal has zero mean and that the single scaling coefficient at the coarsest scale $c_{0,0}$ is identically zero. The wavelet coefficients are modeled as independent Gaussian mixture random variables with identical distributions within each scale. That is, all wavelet coefficients at scale i are independent and identically distributed. Furthermore, assume that the variance of the wavelet coefficients at scale i is

$$\mathrm{E}[\omega_{i,k}^2] = 2^{-\gamma i}, \;\; 0 < \gamma < 2, \; \gamma \neq 1 \tag{4}$$

The variable γ controls the rate at which the variance decays across scale. The restrictions on the range of γ are assumed throughout the paper, and their necessity will become obvious in later analysis. Similar variance models have been used in the analysis and synthesis of $1/f$ processes using wavelets [4]. We will show that the SIDWT-IMM synthesizes a stationary signal prior with $1/f$ characteristics.

3.1. Autocorrelation of DWT-IMM

Now consider the correlations between scaling coefficients (signal samples) at the finest scale j. To keep the notation as simple as possible, we drop the position subscript of scaling and wavelet coefficients. Let c_j be a scaling coefficient at scale j (arbitrary position). The Haar coefficient recursions allow us to express c_j in terms of the wavelet coefficients at coarser scales:

$$c_j = \sum_{i=0}^{j-1} \omega_i 2^{-(j-i)/2} \tag{5}$$

where ω_i denotes a wavelet coefficient at scale i (again we drop the position subscript for convenience). Note that in (5) we also make use of the fact that $c_{0,0} = 0$, otherwise we would have one additional term in the series, namely $2^{-j/2}c_{0,0}$. According to (4), the variance of c_j is

$$\begin{aligned} \mathrm{E}[c_j^2] &= 2^{-j}\sum_{i=0}^{j-1}\mathrm{E}[\omega_i^2]2^i \;=\; 2^{-j}\sum_{i=0}^{j-1}2^{-\gamma i}2^i \\ &= \frac{2^{-\gamma j} - 2^{-j}}{2^{(1-\gamma)} - 1} \end{aligned} \tag{6}$$

The correlation between two scaling coefficients at scale j and different positions is deduced as follows. Since the two scaling coefficients, say c_j and c_j', have different positions, their wavelet coefficient expansions (5) differ by at least one wavelet coefficient. That is, at least one wavelet coefficient is different in the expansion of c_j' compared to that of c_j. The number of differing wavelet coefficients in the expansions of c_j and c_j' depends on the spatial displacement between the two and on their alignment with respect to the Haar wavelet basis functions. In general, the wavelet expansions of the two coefficients will be identical up to some scale $j - m$, $m > 0$, and then different from that point on. In this case, we say that the coefficients *split* at scale $j - m$.

Let us suppose that c_j and c'_j that split at scale $j-m$. The expansions of these coefficients can be written as

$$\begin{aligned} c_j &= 2^{-\frac{m}{2}}\left(c_{j-m}+\omega_{j-m}\right)+\sum_{i=1}^{m-1}\omega_{j-m+i}2^{\frac{i-m}{2}} \\ c'_j &= 2^{-\frac{m}{2}}\left(c_{j-m}-\omega_{j-m}\right)+\sum_{i=1}^{m-1}\omega'_{j-m+i}2^{\frac{i-m}{2}} \end{aligned}$$

Both coefficients have the first two terms in common, but the remaining terms are independent and zero-mean. Hence, their correlation is

$$\begin{aligned} \rho_m := \mathrm{E}[c_j c'_j] &= 2^{-m}\left(\mathrm{E}[c_{j-m}^2]-\mathrm{E}[\omega_{j-m}^2]\right) \\ &= 2^{-m}\mathrm{E}[c_{j-m}^2]-2^{-\gamma(j-m)}2^{-m} \quad (7) \end{aligned}$$

Substituting the expression for $\mathrm{E}[c_{j-m}^2]$ given in (6), we find

$$\rho_m = a\,2^{(\gamma-1)m}-b, \quad (8)$$

where $a = 2^{-\gamma j}\frac{2-2^{(1-\gamma)}}{2^{(1-\gamma)}-1}$ and $b=\frac{2^{-j}}{2^{(1-\gamma)}-1}$. Note that a and b are constants, depending on the length of the signal 2^j and the γ. Furthermore, we see from (8) that the correlation depends on the displacement between the two scaling coefficients through m, number of scales above the finest scale that the two split. As the displacement between the two scaling coefficients is increased, m is non-decreasing and it is easy to show that the correlation is therefore non-increasing. Moreover, the minimum correlation at $m=j$ is $\rho_j=-2^{-j}$, independent of γ. Two correlation functions (at two different locations $k=0$ and $k=55$) of a 256 length DWT-IMM signal prior are shown in Figure 1. Note that the DWT-IMM is a nonstationary signal model.

3.2. Autocorrelation of SIDWT-IMM

Now consider the autocorrelation function of the SIDWT-MMI model. Given a displacement of n between two scaling coefficients, c_j and c'_j, we can compute the probability of the scale j^* at which the two split, with respect to the uniform distribution over the shift parameter, as follows. We need to count the number of shifts that give rise to each possible value of j^*, or more precisely the probability of the set of shifts that give rise to each possible j^*. Given a displacement of $n=k$ between two scaling coefficients, the probability of the scale j^* of their split can be determined by inspecting the binary tree associated with the DWT. We need only consider $|k|\le 2^{j-1}$ due to the periodicity of the DWT. First note that we have $\Pr(j^*\ge 0\,|\,n=k)=1$. Next, notice that $(2^m-k)_+$ is the number of shifts of a length-k sequence that fit within a length-2^m sequence, where $(x)_+=\max(x,0)$. In other other words, $(2^m-k)_+$ is the total number of scaling coefficient pairs spaced k apart at the bottom of a m-level binary tree. Also notice that the number of m-level subtrees, $m<j$, at the bottom of a larger j-level binary tree is precisely 2^{j-m}. Hence, for $m<j$

$$\begin{aligned} \Pr(j^*\ge j-m\,|\,n=k) &= (2^m-|k|)_+\,2^{j-m}2^{-j} \\ &= (2^m-|k|)_+\,2^{-m} \quad (9) \end{aligned}$$

It follows that

$$p(m|k) := \Pr(j^*=j-m|n=k) = \begin{cases} 0, & m<\lceil\log_2(|k|+1)\rceil \\ 1-2^{-m}|k|, & m=\lceil\log_2(|k|+1)\rceil \\ 2^{-m}|k|, & \lceil\log_2(|k|+1)\rceil<m<j \\ |k|2^{-j+1}, & m=j \end{cases}$$

where $\lceil\log_2(|k|+1)\rceil$ denotes the smallest integer greater than or equal to $\log_2(|k|+1)$.

With these probabilities defined, the autocorrelation function of the SIDWT-IMM is

$$r(k) := \mathrm{E}\left[c_j c'_j\right] = \sum_{m=1}^{j}\rho_m\,p(m|k) \quad (10)$$

where the expectation is now taken over *both* the wavelet coefficient distributions and the shift prior, and ρ_m is given in (8). After some algebra, we have

$$\begin{aligned} r(k) &= a(2^{(\gamma-1)\lceil\log_2(|k|+1)\rceil}-\beta|k|2^{(\gamma-2)j} \\ &\quad +\beta|k|2^{(\gamma-2)\lceil\log_2(|k|+1)\rceil})-b \quad (11) \end{aligned}$$

for $|k|>0$, where

$$\beta=\frac{1-2^{(\gamma-1)}}{1-2^{(\gamma-2)}} \quad (12)$$

In the case $k=0$, $r(0)$ is given by (6). Finally, making use of the approximation $2^{\lceil\log_2(|k|+1)\rceil}\approx|k|$, we have

$$r(k)\approx a\left[(1-\beta)|k|^{(\gamma-1)}+\beta|k|2^{(\gamma-2)j}\right]-b,\quad |k|>0 \quad (13)$$

For large j and $\gamma<2$, the term $\beta|k|2^{(\gamma-2)j}$ is negligible, and hence the correlation function behaves like $|k|^{(\gamma-1)}$ and the power spectrum decays like $\frac{1}{|f|^\gamma}$ (see [4] for the relationship between autocorrelation functions and power spectrums of $1/f$ processes). That is, the SIDWT-IMM produces a stationary, non-Gaussian[1] process that is approximately $1/f$. In fact, the approximation in (13) is very accurate and coincides with the true correlation at points $|k|=2^i$, i integer. The (stationary) autocorrelation function for a 256 length SIDWT-IMM signal prior is shown in Figure 1. Note that the SIDWT-IMM correlation function is piecewise linear instead of piecewise constant like the DWT-IMM correlation. Hence, we gain stationarity and a degree a regularity from the SIDWT-IMM.

[1] If $p_i=1$ or $p_i=0$ in (1), then we have a Gaussian process. In this case, the SIDWT-IMM provides a method for synthesizing *stationary*, Gaussian processes with $1/f$ characteristics, extending previously proposed non-stationary synthesis methods [4].

4. APPLICATION: DENOISING $1/F$ PROCESSES

Suppose that we observe a signal in noise, and let us model the signal as a (non-Gaussian) $1/f$ process with parameter γ. In this model, the variance of the wavelet coefficients at each scale i is

$$\mathrm{E}[\omega_{i,k}^2] = \sigma_H^2 2^{-\gamma i} \tag{14}$$

where $\sigma_H^2 > 0$. Furthermore, assume that the distribution of each $\omega_{i,k}$ is a two-component mixture with variances σ_H^2 and $\sigma_L^2 = 0$. With $\sigma_L^2 = 0$, we are setting the low-variance component to be simply be a point mass at the origin. Note that (14) implies that $\omega_{0,0}$ is a zero-mean, Gaussian random variable with variance σ_H^2 (a reasonable model for the wavelet coefficient at the coarsest scale in light of the Central Limit Theorem). Furthermore since $\sigma_L^2 = 0$, (14) implies that the mixing probability at each scale is

$$p_i = 1 - 2^{-\gamma i} \tag{15}$$

The analysis in the previous section shows that the SIDWT-IMM with these parameters produces a *non-Gaussian* $1/f$ process with parameter γ. Moreover, in this case the model (and hence the estimator) is completely parameterized by only three parameters: σ_H^2, γ, and the noise power σ^2.

In practice the signal characteristics are unknown *a priori*, and we have only a noisy observation available. Selection of useful model parameters is a crucial issue. The $1/f$ model significantly reduces the parameterization (instead of two variances and a mixing probability at each scale and the noise variance as in (2), the $1/f$ model has only three parameters in total). A simple method for setting the three parameters in practice is to first estimate σ using the robust estimator $\widehat{\sigma} = \frac{\mathrm{Median}|d_{j,k}|}{0.6745}$. Next, the parameters γ and σ_H^2 can be estimated by the slope and intercept of a line fit to the points $\log_2(\widehat{\sigma_i^2} - \widehat{\sigma}^2)$ $0 < i \leq j$, where $\widehat{\sigma_i^2}$ is the sample variance of the wavelet coefficients at scale i. The results of a denoising experiment that used this strategy are shown in Figure 2.

5. CONCLUSIONS

Like the previously proposed DWT-IMM, the SIDWT-IMM captures the key statistics of wavelet coefficients. However, unlike the DWT-IMM which produces non-stationary priors with blocky correlation structures, the SIDWT-IMM is stationary and has piecewise linear correlation functions. Hence, I feel that the SIDWT-IMM is much more reasonable as prior signal model. Moreover, it was shown that with appropriately decaying wavelet coefficient variances the SIDWT-IMM is approximately $1/f$ in nature. The connection between $1/f$ processes and this model was also exploited to develop a simple, three-parameter, Bayesian denoising algorithm for estimating a non-Gaussian $1/f$ process in additive Gaussian white noise.

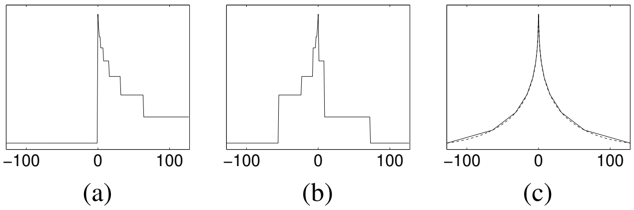

Figure 1: Autocorrelations of DWT-IMM and SIDWT-IMM 256 sample signal priors. DWT-IMM autocorrelation (a) $r(0,k) = \mathrm{E}[s(0)s(k)]$ (b) $r(55,k+55) = \mathrm{E}[s(55)s(55+k)]$. (c) Stationary SIDWT-IMM autocorrelation $r(k)$ (solid), approximation in eqn. (13) (dashed). For both the DWT-IMM and SIDWT-IMM in this example $\mathrm{E}[\omega_{i,k}^2] = 2^{-\gamma i}$, $\gamma = 1.25$.

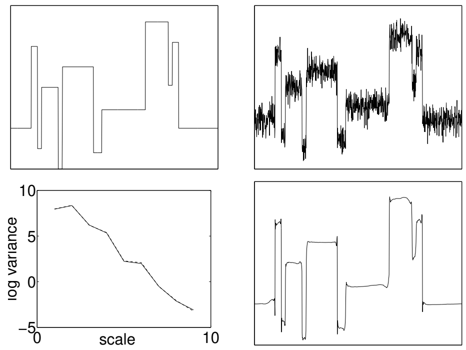

Figure 2: Denoising a $1/f$-like signal. Clockwise from upper-left: ‘Blocks’ test signal, noisy Blocks, denoised Blocks, $1/f$-like decay of variance across scale. Estimated γ (slope) $= 1.38$.

6. REFERENCES

[1] M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, “Wavelet-based statistical signal processing using hidden Markov models,” *IEEE Trans. Signal Processing*, vol. 46, pp. 886–902, 1998.

[2] R. Coifman and D. Donoho, “Translation invariant denoising,” in *Lecture Notes in Statistics: Wavelets and Statistics*, vol. New York: Springer-Verlag, pp. 125–150, 1995.

[3] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, “Noise reduction using an undecimated discrete wavelet transform,” *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10–12, 1996.

[4] G. W. Wornell, *Signal Processing with Fractals. A Wavelet-Based Approach.* Englewood Cliffs, New Jersey: Prentice Hall, 1996.