Directional step models for multiscale correlation thresholding

Eric Chicken

Department of Statistics, Florida State University, Tallahassee, Florida, USA

The wavelet based thresholding techniques described in Chicken et al. [2007] are studied via simulated data in this article. Forward, backward, and multidirectional step algorithms are examined with respect to the number of wavelet coefficients selected, mean squared errors (MSE), and increase in correlation.

The simulated data varies several parameters of interest. In the model

$$Y_i = f(t_i) + \varepsilon_i = g(t_i) + h(t_i) + \varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

h is linearly related to a known external signal h' and $\int h = 0$. Various levels of the noise, $\varepsilon \sim$ normal $(0, \sigma^2)$, is added to the composite function $f = g + h$. Also, normal noise is added to the external function h'. The noise changes the relation between h and h' from linearly related to correlated (positive in these simulations). The level of the noise is varied so that the signal to noise ratio (SNR) of the functions $f = g + h$ and h' compared to σ is an integer value from 2 to 8. In the simulations, n is taken to be 512.

The condition of orthogonality between g and h is also varied. The functions g and h are created to meet the requirements specified in Chicken et al. [2007]. There are three cases considered. First is the case where g and h reside in different resolution levels of the multiresolution analysis (MRA) of the discrete wavelet transform with respect to the least asymmetric wavelet with support length 8 ($la8$). The second case considered is when the functions g and h are still orthogonal as above, but the functions do not necessarily reside in different resolution levels. The third case simulates when g and h are nearly orthogonal.

In addition to varying the SNR and the form of the orthogonality between g and h, the wavelet used for the algorithm is also varied. The functions h and g were orthogonalized using the wavelet $la8$. The algorithms were run using this wavelet, as well as the best localized wavelet with support length 20 ($bl20$). When using $bl20$, the functions are nearly orthogonal for all three cases. They are only exactly orthogonal in cases 1 and 2 when the wavelet $la8$ is used for analysis.

Table 1 shows the results of applying the forward searching algorithm to the simulated data. 100 runs of the algorithm were completed in order to estimate the MSE. The level-by-level cutoff used is $p' = 0.01$, and the minimum increase in correlation is $p = 0$. For errors as percentages, the measure of the sizes of the functions h (corresponding to the MSE as the measure of the error) is $\int (h - \int h)^2$. For the h in cases 1 through 3 of this table, these sizes are 35.72714, 10.63935, and 19.33601, respectively. The MSE in case 1 with SNR 2 as a percentage of the size of h is 0.00588%, for instance.

Table 1. Forward step algorithm, $bl20$ wavelet.

SNR	MSE	ρ_f	ρ_h	n_c
		Case 1		
2	0.00210	0.4458	0.7806	20.96
3	0.00183	0.5056	0.8315	21.12
4	0.00162	0.5391	0.8616	21.40
5	0.00154	0.5631	0.8814	21.64
6	0.00145	0.5761	0.8949	21.99
7	0.00139	0.5907	0.9052	21.85
8	0.00138	0.5973	0.9132	21.94
		Case 2		
2	0.01579	0.3988	0.7335	13.20
3	0.01269	0.4526	0.7966	15.94
4	0.01196	0.4820	0.8250	17.50
5	0.01097	0.5031	0.8483	19.87
6	0.01122	0.5151	0.8572	19.49
7	0.00976	0.5317	0.8782	23.31
8	0.00940	0.5356	0.8862	23.62
		Case 3		
2	0.00713	0.2311	0.7455	5.03
3	0.00625	0.2614	0.8046	5.14
4	0.00594	0.2794	0.8302	5.32
5	0.00582	0.2897	0.8487	5.17
6	0.00563	0.3012	0.8660	6.29
7	0.00551	0.3081	0.8724	6.32
8	0.00564	0.3120	0.8807	6.46

Several patterns are present with respect to increasing SNR. For the lowest simulated SNR, the MSE is largest because of the large amount of noise in the functions. Reducing this noise allows the estimator to more accurately model the function h. The table also displays averages of the initial value of the correlation ρ_f between h' and f, the correlation ρ_h between h and h', and the number of coefficients used to model h (n_c). As the SNR increases, the MSE decreases, and the remaining quantities increase. The final correlation values ρ_h approach 1 as expected if we are accurately modeling h. The number of coefficients is increasing as well. However, n_c does not approach the actual number of coefficients needed for h due to the sparseness property of wavelets, especially for case 1. Orthogonality requirements of the algorithm are not met when using $bl20$ for analysis, but the results in terms of error are still excellent.

Figure 1 displays a typical run of the forward step algorithm. Here the wavelet $bl20$ is used for analysis, the SNR is 5, and the functions g and h follow case 3. The noisy observed signals f (solid) and h' (dotted) are shown in the top panel. These are the data the algorithm uses in analysis. Note the two scales for the different signals. The bottom panel shows the true function h (dotted) and its estimate h' (solid). The MSE of the estimate shown is 0.00341, or, 0.01764%. The initial correlation between the noisy h' and noisy f is $\rho_f = 0.2977$ and the final correlation between $\hat{h}$ and the noisy h' is $\rho_h = 0.8889$.

Other directional models (backward, forward-backward, backward-forward) were implemented, also. Results were comparable to those discussed here. One notable difference is that the backward algorithm always used more coefficients than the forward method (not surprising since the backward model starts with all 512 wavelet coefficients). Multidirectional models behaved similarly, the forward-backward method behaving like the forward method, and the backward-forward method mimicking the backward method. The multidirectional method is the most computationally intensive. However, it may be useful (and necessary) in a situation where the initial configuration map of the wavelet coefficients is not empty or full. For each step method, case 1 shows best results. This is unsurprising since here the functions h and g are well separated in resolution spaces.

For the variant estimator, where wavelet coefficients are added in levels rather than individually, the results again are similar to those in Table 1. The only important difference is in the number of coefficients used, n_c. The variant estimator always uses more coefficients than the original estimator. This is not surprising since it adds them in groups rather than individually.

Changing the wavelet used for analysis, *la*8 vs *bl*20, has little effect on the results. For more visual and numerical results regarding wavelet choice, test function choice, step method, and the variant estimator, see Chicken [2006].

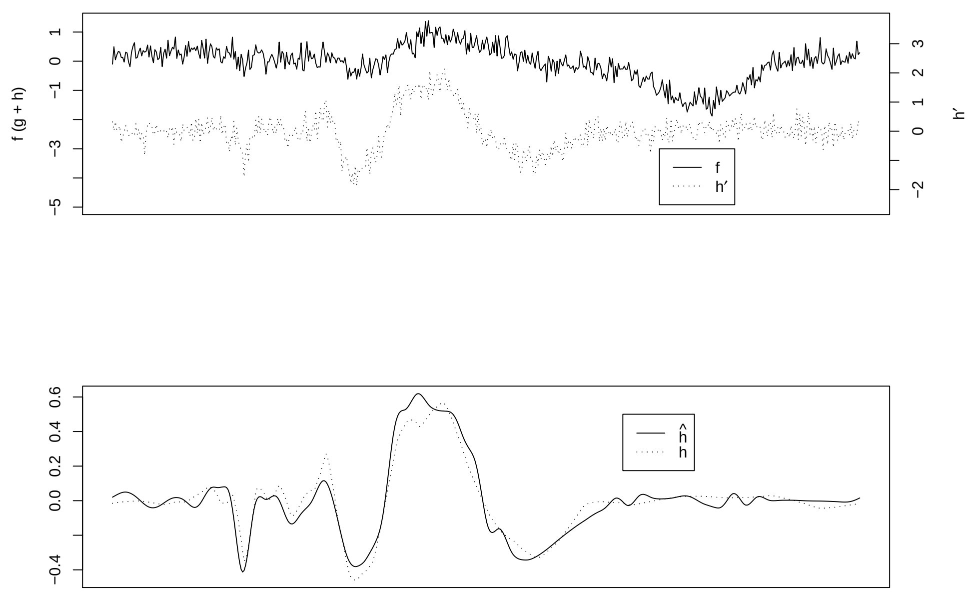

Figure 1. Results for case 3, SNR 5, wavelet *bl*20, forward step algorithm.

References

Chicken, E. (2006). Directional step models for multiscale correlation thresholding. Tech. Rep. M977, Florida State University Department of Statistics. http://stat.fsu.edu/techreports.

Chicken, E., Loper, D. and Werner, C.(2007). Estimating tidal effects in spring discharge: a multiscale method using correlated phenomena. *Water Resour. Res.* **43**, W05414 doi:10.1029/2006WR005117.

E. Chicken, Department of Statistics, Florida State University, Tallahassee, FL 32306, USA. (chicken@stat.fsu.edu)