Rejoinder: Time-Threshold Maps: using information from wavelet reconstructions with all threshold values simultaneously

Piotr Fryzlewicz*

February 22, 2012

I am very grateful to all four sets of discussants: Jean-Marc Freyermuth & Rainer von Sachs, Thomas Lee, Yaeji Lim & Hee-Seok Oh, and Hernando Ombao, for their interesting and stimulating comments, questions and insights. I offer my thoughts on some of them below.

One issue discussed by Freyermuth & von Sachs is that of the difference between what they refer to as 'diagonal' ('separable') and 'tree-structured' ('vertical') wavelet thresholding rules. In the former, the decision on whether to keep or kill, or alternatively whether and how to modify, each wavelet coefficient only depends on the magnitude of the coefficient and on the value of a scalar parameter, referred to as a threshold. In the latter, extra factors in the decision are the parent and/or children of the given coefficient in the wavelet decomposition tree. It is worth mentioning that reconstructions from such a tree-structured scheme could also be visualised via a Time-Threshold Map, as long as the thresholding decisions for the coefficients of the decomposition could be jointly parameterised by a single real-valued parameter which would be assigned to the y-axis of the TTM. For example, in Fryzlewicz (2007), the simplest such decision rule for each wavelet coefficient $d_{j,k}$ is: keep if $(d_{j,k})^2 + (d^{(P)}_{j,k})^2 > \lambda^2$ and kill otherwise, where $d^{(P)}_{j,k}$ is the parent of $d_{j,k}$. In this example, the parameter appearing on the y-axis of the corresponding TTM could be e.g. λ, or λ^2.

Freyermuth & von Sachs also ask a series of stimulating questions regarding the derivative TTM, and in particular the Orthogonal Feature Decomposition. I believe that the analysis of the properties of the OFD as a tool for signal approximation and feature extraction can only be meaningfully performed in conjunction with the analysis of the properties of the underlying wavelet transform: the OFD is best interpreted as acting on an existing transform (by grouping the wavelet basis functions according to the magnitudes of the corresponding coefficients) rather than serving as a separate transform in itself. For this reason, approximation properties of the wavelet basis (whether selected adaptively or not) will have a substantial impact on those of the resulting OFD. Thus it is my belief that the

*Department of Statistics, London School of Economics, Houghton Street, London WC2A 2AE, UK. Email: p.fryzlewicz@lse.ac.uk.

most obvious route to investigating the approximation properties of the OFD is to develop an understanding of the approximation properties of the underlying wavelet basis (a useful resource for the latter is e.g. DeVore (1998)).

From the algorithmic point of view, similar "grouping according to coefficient magnitudes" could be performed for any decomposition of a signal into a linear combination of simpler functions, such as that of Chen et al. (1998), and similarly yield a resolution of the signal into what can be interpreted as its "more important" (i.e. those corresponding to larger coefficients) and "less important" components.

Extension of the TTM approach to image data would be possible: the (three-dimensional) object of interest would then be a sequence of image reconstructions via wavelets, performed with increasing (or decreasing) threshold values and 'stacked' one on top of another.

Further, the TTM approach could be used to compare two or more decompositions of the same signal, each in a different wavelet basis or system. The comparison could be done according to a pre-specified criterion expressible as an appropriate functional of the TTM. This could lead to the choice of the 'best' transform for the signal, as surmised by Freyermuth & von Sachs.

Freyermuth & von Sachs's final question regarding the TTM's potential capacity for signal classification, especially in the context of a competing method by Timmermans et al. (2011), also based on Unbalanced Haar wavelets, is a very interesting one. In the latter work, the authors define a distance between two signals by comparing the magnitudes and the locations of their respective Unbalanced Haar coefficients, arranged in the order in which the corresponding Unbalanced Haar vectors are selected. I conjecture that the natural construction of a TTM-based distance using Unbalanced Haar wavelets would be to compare the OFD's of both signals, in which case one would most likely be comparing the orthogonal features corresponding to the same or similar threshold values in both signals. I leave this thought-provoking question for future investigation.

The discussion contribution by Lim & Oh proposes a number of interesting possible extensions, modifications and improvements to the TTM methodology.

As regards the detection of peaks in the variogram, rather than using the TTM method based on Unbalanced Haar wavelets (which 'specialise' in changepoint detection rather than peak detection, see Cho and Fryzlewicz (2011)), it might be advantageous to use the taut string methodology of Davies and Kovac (2001) in this context. The latter technique can lead to consistent estimation of peaks and could be embedded in the TTM framework by setting the parameter on the y-axis of the TTM to be the 'tube width' (see Davies and Kovac (2001) for a description of this parameter) rather than the 'threshold'. The reader is referred to Cho and Fryzlewicz (2011) for a description of differences and similarities between the Unbalanced Haar and taut string function estimation techniques.

Another interesting aspect of Lim & Oh's contribution is an attempt to use the TTM technique to aid the visualisation of the dependence structure between two signals or time series. Here, at least two alternative TTM-based techniques might be possible. The first possibility would be to construct a TTM of a 'raw' measure of dependence between two signals X_t and Y_t; the simplest such measure, taken under the assumption of both X_t and Y_t having mean zero, would be the product $Z_t = X_t Y_t$ (note that $E(Z_t) = \mathrm{Cov}(X_t, Y_t)$), but more complex measures such as various local cross- periodograms, might also be an option.

Another possibility would be, for example, to define an empirical multiscale measure of dependence between X_t and Y_t by computing the sequence of sample correlations between the corresponding orthogonal features of X_t and Y_t. This, I believe, could serve as a basis for constructing a TTM-based distance between X_t and Y_t, an issue also mentioned in Freyermuth & von Sachs's contribution, as described above. This could potentially be an example of what Freyermuth and von Sachs refer to as "going beyond the representational benefits of the derivative TTM".

Lim & Oh also propose, and Lee mentions, a possible robust version of the TTM. It is worth mentioning that naturally robust signal processing techniques such as the median filter (see e.g. Huang and Lee (2006), where a data-adaptive span selection technique is proposed for such a filter) could also be embedded in the TTM framework, by simply assigning the parameter of the given technique (e.g. the span in median filtering) to the y-axis of the TTM.

To respond to the final issue noted in Lim & Oh's contribution, it is worth noting that Haar and Unbalanced Haar wavelets do not suffer from the boundary problems, which is one of the reasons why the present paper focuses on these two families. In addition, the Unbalanced Haar transform extends naturally to data vectors of any length, whereas there is no unique way of extending 'classical' discrete wavelet transforms to data of length different from an integer power of two.

Lee, I believe, is correct in stating that SiZer and TTM should be viewed as complementary, rather than competing, techniques. However, I was interested to read about the potential advantage of the TTM over SiZer in terms of their computational speeds.

I also agree with Lee that iterating the construction of the artificial signal $\tilde{X}_t$ might be a good idea in certain situations. However, it must be borne in mind that each such iteration affects the structure of the noise – as I imagine, not always in easy to quantify ways. Also, $\tilde{X}_t$, constructed by averaging the TTM across thresholds, is but one example of a functional of the TTM. Other, more complex, functionals, are possible, including ones that attach more importance to certain thresholds than to others, perhaps using Bayesian formalism, as suggested by Ombao.

Indeed, the 'averaging' aspect is the main focus of Ombao's contribution. In fact, he goes further and proposes an interesting aggregated wavelet function estimate, which at any point of the domain is defined as a convex combination of the rows of the TTM; but in which the weights are permitted to vary over the domain. This, I believe, is an appealing idea which is worth investigating in more detail.

Once again, I would like to thank all four sets of contributors for the extremely interesting discussion, and the Editors for accepting this work for publication in the Journal.

References

S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20:33–61, 1998.

H. Cho and P. Fryzlewicz. Multiscale interpretation of taut string estimation and its connection to unbalanced haar wavelets. *Statistics and Computing*, 21:671–681, 2011.

P.L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution (with discussion). *Annals of Statistics*, 29:1–65, 2001.

R. DeVore. Nonlinear approximation. *Acta Numerica*, 7:51–150, 1998.

P. Fryzlewicz. Bivariate hard thresholding in wavelet function estimation. *Statistica Sinica*, 17:1457–1481, 2007.

H.C. Huang and T. Lee. Data adaptive median filters for signal and image denoising using a generalized SURE criterion. *IEEE Signal Processing Letters*, 13:561–564, 2006.

C. Timmermans, L. Delsol, and R. von Sachs. Bases giving distances. A new semimetric and its use for nonparametric functional data analysis. In F. Ferraty, editor, *Recent Advances in Functional Data Analysis*, Contributions to Statistics, pages 307–313. Springer, 2011.