Bayesian Doubly Blocked Wavelet Thresholding, Oscillation Spaces and Varying Bandwidth Kernel Smoothers.

Pierpaolo Brutti

University "La Sapienza" of Rome
pierpaolo.brutti@uniroma1.it

Abstract

In a variety of real–life signals, significant wavelet coefficients often occur in cluster at adjacent scales and locations. Irregularities, like a discontinuity for example, in general tend to affect the whole block of coefficients corresponding to wavelet functions whose "support" contains them. For this reason, heuristically, it is reasonable to expect that the risk of "blocked" thresholding rules might compare quite favorably with other classical estimators based on level–wise or global thresholds.

Successful examples of "horizontally" (within scales) blocked rules include the nonoverlapping thresholding of Cai (1999) and Hall et al. (1999), the *NeighBlock* and *NeighCoeff* methods of Cai and Silverman (2001), as well as their Bayesian counterparts explored by Abramovich et al. (2000), Wang and Wood, (2003) and De Canditiis (2001), just to mention a few.

Recently, an increasing amount of work has been devoted to study a new class of "vertically" (across scales) blocked rules. For example, in Romberg et al. (2001), a new framework to capture the statistical dependencies by using wavelet–domain hidden Markov Models is developed. On the other hand in Autin et al. (2004) and Brutti (2004) a vertical (hereditary) and not disjoint thresholding method inspired by Cohen et al. (1999), is introduced and its similarity to existing variable bandwidth kernel smoothers established. Finally in Sendur (2002) and Fryzlewicz (2004), a Bayesian parent–child vertical rule is proposed and theoretically explored.

In this work we basically try to fill in the gaps. We start reviewing the non-disjoint vertical wavelet thresholding methodology outlined in Brutti (2004), and its interpretation as a computationally efficient alternative to existing varying bandwidth kernel smoother (see Gijbels and Mammen, 1998, or Lepski et al., 1997). Then we extend this procedure introducing a within–scale blocking parameter. At this point, as in Antoniadis and Fan (2001), we rewrite the problem as a penalized regression model, characterized by the particular non–separable penalty function coming from our double–block policy. Exploiting the Bayesian interpretation of the derived optimization problem, we are left, as usual, with a vector of smoothing parameters to be set. A convenient way to get a data based estimate of it, is by using the well known *generalized cross validation* criterion as suggested, for example, in Amato et al. (2004) or Jansen et al. (1997). In this work, we follow a slightly different path. We interpret the classical GCV criterion as a particular data–dependent (second level) prior over the smoothing parameter vector, to be calibrated, as in George and Foster (2001) or Autin et al. (2004), in such a way that the resulting MAP estimator achieves the minimax rate over some functional class of interest or, by duality, has that functional class as maxiset with a prescibed (possibly optimal) rate. Note that by applying this same reasoning to other common risk estimators, such as SURE for example, one can obtain a whole family of very flexible smoothing techniques.

Finally we close the circle. Noting that the wavelet characterization of classical Besov, or weak Besov spaces, cannot capture existing across–scale correlations, given its invariance with respect to within scale permutations, we switch setting, and establish a link between the penalty arising from our vertical shrinkage methodology and the semi–norm over the so called *Oscillation spaces* introduced and studied in Jaffard (2004), so that, as in Amato et al. (2004), we are eventually able to cast the whole procedure within some sort of wavelet–based reproducible kernel Hilbert space framework.

Comparison with existing methods and possible extensions to non-Gaussian (heteroscedastic) noise and non–equispaced, high–dimensional designs, are also briefly sketched.

References

Abramovich, F., Besbeas, P. and Sapatinas, T. (2002). Empirical Bayes approach to block wavelet function estimation, *Computational Statistics and Data Analysis*, 39, 435–451.

Amato U., Antoniadis A. and Pensky M. (2004). Wavelet kernel penalized estimation for non-equispaced design regression. Submitted to *Statistics and Computing*.

Antoniadis A. and Fan J. (2001). Regularization of wavelet Approximations. *Journal of the American Statistical Association*, **96**, 939–967.

Antoniadis, A. and Fryzlewicz, P. (2004). Parametric modelling of thresholds across scales in wavelet regression. *Technical Report TR-04-04*, Department of Mathematics, Imperial College London, UK.

Autin F. , Picard D. , Rivoirard V. (2004). Maxiset comparisons of procedures, application to choosing priors in a Bayesian nonparametric setting. *Prépublication PMA-931*, Laboratoire de Probabilités et Modéles Aléatoires, Université Paris 7.

Brutti, P. (2004). Variable bandwidth schemes for local polynomial smoothers via vertical wavelet thresholding In R.G. Aykroyd, S. Barber, and K.V. Mardia (Eds.), *Bioinformatics, Images, and Wavelets*, pp. 119–121. Department of Statistics, University of Leeds.

Cai, T. (1999). Adaptive wavelet estimation: a block thresholding and oracle inequality approach. *Annals of Statistics*, 27, pp. 898–924.

Cai, T. and Silverman, W. (2001). Incorporating information on the neighboring coefficients into wavelet estimation. *Sankhyā, Ser. B*, 63, pp. 127–148, Special issue on wavelets.

Cohen, A., Dahmen, W., Daubachies, I, and DeVore, R. (2001). Tree approximation and optimal encoding. *Applied Computational and Harmonic Analysis*, **11**(2), pp. 167–191.

De Canditiis, D. and Vidakovic, B. (2001). Wavelet Bayesian block shrinkage via mixtures of normal–inverse gamma priors. *Research report*.

Fan, I. and Gijbels, I. (1996). Local Polynomial Modelling and Its Applications. London, Chapman and Hall.

Fryzlewicz, P. (2004). Bivariate hard thresholding in wavelet function estimation. *Technical Report TR-04-03*, Department of Mathematics, Imperial College London, UK.

George, E.I. and Foster D.P. (2000). Calibration and empirical Bayes variable selection. *Biometrika*, **87**(4), pp. 731–747.

Gijbels, I. and Mammen, E. (1998). Local adaptivity of kernel estimates with plug-in local bandwidth selectors. *Scandinavian Journal of Statistics*, **25**, 503-520.

Hall, P., Kerkyacharian, G. and Picard, D. (1999). On the minimax optimality of block thresholded wavelet estimators. *Statistica Sinica*, 9, 33–50.

Jaffard, S. (2004). Beyond Besov spaces, Part 2: Oscillation Spaces, *Constructive Approximation*.

Jansen M., Malfait M. and Bultheel A. (1997). Generalized Cross Validation for wavelet thresholding. *Signal Processing*, **56**(1), pp. 33-44.

Lepski, O. V., Mammem, E. and Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, **7**(3), 929-947.

Romberg J., Choi H. and Baraniuk R. (2001), Bayesian tree-structured image modeling using wavelet domain hidden markov models, *IEEE Transactions on Image Processing*, **10**(7), pp. 1056-1068.

Sendur, L. and Selesnick, I. (2002). Bivariate shrinkage functions for wavelet-based denoising exploiting interscale despendency. *IEEE Trans. Sig. Proc.*, 50, pp. 2744–2756.

Wang, X.W. and Wood, A.T.A. (2003). Empirical Bayes block shrinkage of wavelet coefficients via the non–central χ^2 distribution. Tech. Report 03–01, University of Nottingham, Division of Statistics.