Application of Lifting Transform to Detect Change Points

Chatterjee, Arunendu
University of Wisconsin River Falls, Department of Mathematics
410 South Third Street
River Falls(54022),USA
E-mail: arunendu.chatterjee@uwrf.edu

Change point is the problem of identifying sudden change at a particular time point in the pattern of the data collected over time. Change point problem has many diverse applications not only in statistics but also in other disciplines such as hydrology, climatology etc. where change point problem occurs regularly. There are various aspects to the change point problem, namely detection of change point, estimation of the time at which the change occurred and finally, modeling the data before and after change. A substantial literature exists on models that combine detection, estimation and modeling using a statistical framework. In this paper, we discuss a Bayesian procedure to detect change points in the presence of missing or irregular data points. Modeling change points may become complicated in the presence of missing data. We discuss here a detection method using second generation wavelet transforms.

In statistics literature, Discrete wavelet transform (DWT) is used to detect change points. DWT, however, can not be used to detect change points in the presence of missing data. We thus introduce second generation wavelet transform technique or lifting technique to detect change points. The use of lifting technique to detect change points is a new and timely procedure.

We introduce an algorithm based on lifting transform to detect change points. Our method is easy to implement and can be applied to any data size. Bayesian procedure is used to find the posterior distribution of the change point and the position of the change point can be determined by the mode of the skewed posterior distribution.

The objective of this paper is to compare the existing DWT method with the method of lifting transform to show the relative importance of these two methodologies in the context of real data problems.

Overview of Change Point Methods

A change point is the time at which some features of the distribution of a variable changes; the most common features usually considered are changes in the mean structure in the form of shifts in trends, or changes in the variance structure. Detection of change points is a complicated problem in practice as neither the occurrence nor the possible multiplicity of change points is known.

The change point problem was originally addressed in Bayesian statistics by Smith (1975), followed by Carter and Blight (1981). Bayesian methods were applied considering single or multiple changes in conjunction with a known or an unknown number of change points. Gelfand et al. (1990) considered a known number of change points and discussed Bayesian analysis of a variety of normal data models, including regression and ANOVA, which allowed some unequal variances. Stephens (1994) carried out a Bayesian analysis of a multiple change point problem where the number of change points was assumed to be known, but the times of occurrence of the change points remained unknown. Examples of such approaches using a known number of change points include Carlin et al.(1992), Tanner (1993) and Rasmussen (2001).

Overview of Wavelets

Wavelets are special functions consisting of dilation and translation indices. Larger values of dilation indices correspond to higher frequency components, and larger values of translation indices correspond to rightward shifts. In practice we use a discrete wavelet transform (DWT) to map a data vector $y = (y_1, ..., y_n)$, for $n = 2^J$, to a vector of wavelet coefficients $w = (w_1, ..., w_n)$ via an orthogonal matrix W. Choice of wavelet functions determines W. Since higher frequency components occur for larger values of dilation, detection of change points involves examination of these higher frequency coefficients in w.

Computation of the discrete wavelet transform is carried out using the popular Mallat's pyramid algorithm which consists of low-pass and high-pass filters through which, at each stage, the input values of the function are decimated. When the data are of size $n = 2^J$ the DWT requires J levels of decomposition. Denoting $n_j = n/2^j$, the output of a DWT is a set of 'detail' coefficients $d_j = (d_{j,1}, d_{j,2}, ..., d_{j,n_j})$ at levels $j = 1, 2, ..., J$ along with 'smooth' coefficients $s_J = (s_{J,1}, s_{J,2}, ..., s_{J,n_J})$, corresponding to the high-pass and low-pass filters, respectively. The detail wavelets coefficients contain the high-frequency content and are used in the change point detection procedure.

A classical approach towards detecting a change point is to choose a level j and examine the corresponding detail coefficients, d_j. Specifically, one can choose a threshold value based on the sample size and an estimated sample variance, compare the coefficients in d_j to this threshold value and decide whether any coefficients are 'large' enough to indicate the existence of a change point. Such a procedure is similar to wavelet-based outlier detection procedures available in Wang (1995) and Bilen and Huzurbazar (2002).

All the applications discussed above suggest a very essential first step, that is, to transform the data into empirical wavelet coefficients through the discrete wavelet transform (DWT). DWT has its own limitations. Such a transformation is only possible under the following conditions:

(a) the time points should be equally spaced, and

(b) total number of observations n should be a power of 2.

Various methods have been proposed for adjusting irregularly spaced data. To address the shortcomings of DWT, the most popular of these methods is considered as lifting technique. We describe below the method of lifting in detail.

Adaptive Lifting

Lifting transform is relatively new in statistics literature and there are no applications of lifting in change point detection. The algorithm was first introduced by Sweldens (1996) which facilitated for a wavelet construction of non-standard data, including irregular data on a grid. Jansen et al. (2004) introduced a new lifting algorithm which was modified by Nunes et al. (2006). An imputation method based on lifting was later introduced by Heaton and Silverman (2008). Lifting essentially consists of three steps: splitting the data, predicting the removed data and updating the remaining data. In this section we describe the different lifting algorithms.

As a first step, the data is subsampled into two disjoint subsets: points representing the even positions in the grid, and points representing the odd positions in the grid (Swelden 1996). Jansen (2004) and Jansen et al. (2001) proposed generating just one wavelet coefficient at each step. Nunes et al. (2006) used a flexible lifting scheme that suggests removing one coefficient at a time in order to build up adaptive prediction steps and embedding them into the lifting algorithm.

The second step is to predict the function values corresponding to the odd positions by using polynomial regression of the corresponding function values to the even positions. The error in prediction (the difference between the true and predicted functional values of odd positions) is then quantified in a vector referred to as the set of wavelet (or detail) coefficients.

In the final step, the function values of the even positions are updated by using linear combinations of the current function values of the even positions and the vector of detail coefficients obtained at the previous step.

The split-predict-update steps can be re-iterated on the updated data set, and the initial data set is replaced by the remaining updated subsample (which reproduces the coarse scale features of the initial signal). The detail coefficients are accumulated throughout this process. This is similar to the DWT method, discussed above, which replaces the initial signal by a set of scaling and wavelet coefficients. The procedure is easily inverted by undoing the update stage, the prediction stage and then merging the subsamples.

Jansen (2004) and Jansen et al. (2001) introduced the concept of lifting just one coefficient in each step. The split step of the lifting algorithm suggests choosing a point which can be removed. The odd/even split, however, poses some problems in higher dimensions. Jansen (2004) and Jansen et al. (2001) proposed removing the points in an order guided by the configuration, namely, that the points belonging to denser areas be removed first. Again, each location is supposed to be associated with an interval values: the shorter the interval, higher the densely sampled area around the location. Once a point has been selected for removal, identify its set of neighbors. The next step is to predict scaling coefficients by using regression over the neighboring locations. The prediction error will be the detail coefficient corresponding to that location. In the update step, only the function-values of the neighboring points are updated by using a linear combination with the detail coefficients, At this stage, the lengths of the intervals associated with the neighboring points are also getting updated, accounting for the decreasing number of scaling points within the range of same interval. The procedure is then repeated on the updated data set, and with each repetition a new wavelet coefficient is added. Hence after say $(n - L)$ removals, we have L scaling coefficients and $(n - L)$ wavelet coefficients.

Nunes et al. (2006) proposed a modification to the LOCAAT (Lifting One Coefficient At a Time) lifting scheme which is called “Adaptive Lifting”. This algorithm is primarily based on lifting one coefficient at a time without imposing any restriction on a choice of prediction or a choice of neighborhood at the beginning of the procedure. The adaptive lifting scheme is called “adaptive” because at each step a choice of prediction or a choice of neighborhood is made. Two types of neighbor choices are used in adaptive lifting. We can choose symmetrical neighbors that is, same number of neighbors on the left and right of the removed points or we can choose closest neighbors to the removed points irrespective of which side they lie. This algorithm is computationally efficient than the LOCAAT lifting schemes suggested by Jansen et al (2001).

We know that the DWT scale is a discrete dyadic quantity. In “one coefficient at a time” lifting, this scale becomes more of a continuous type. In this adaptive lifting scheme the finest level contains half of the wavelet coefficients, the next coarser level represents a quarter and so on.

A few modifications of the resulting coefficients are needed to apply adaptive lifting in Bayesian change point detection procedure. In the framework of nonparametric regression, use of wavelets $\epsilon_{j,k}$ follows independent normal distribution with constant variance. This may not be true in case of lifting. In fact, in this case of lifting there will be a correlation between coefficients and different coefficients will have different variances. To overcome this, we normalize the resulting coefficients by multiplying $[diag(\tilde{W}\tilde{W}^T)]^{-1/2}$ for each step where $\tilde{W}$ is the the matrix associated with the transform.

Lifting Transformation in Bayesian framework

In this section, we derive the marginal posterior distribution of change point using lifting-based wavelet coefficients. It is clear from our above discussion that the lifting coefficients are not independently distributed with constant variance as we expect in case of wavelet coefficients derived from DWT. If we assume initial independence of data then the resulting lifting coefficients remain to be independent but do not have homogeneous variances. To make the variance constant, we have to normalize coefficients at each step by multiplying a transformation matrix. Even if the initial data has a covariance structure, normalizing lifting coefficients at each step by multiplying a transformation matrix will marginally affect the detection of change points.

In the general change point problem, as mentioned by Ogden and Lynch (1998), $Y_1, \ldots\ldots, Y_n$ are the ordered observed data such that $Y_i \sim N(f(i/n), \sigma^2)$ for some function f defined on [0,1]. Hence, f has the form

$$f(u) = \begin{cases} \mu, & \text{if } 0 \leq u < \tau, \\ \mu + \triangle, & \text{if } \tau \leq u < 1. \end{cases}$$

for some change point $\tau \in$ [0,1]. Our objective is to estimate the parameters τ, μ, $\triangle$ and σ^2. Estimation of change-point τ is important here. Later we will choose suitable priors for μ, $\triangle$ and σ^2.

We have already defined lifting transformation. In the lifting-based procedure after $(n-L)$ removals, we have L scaling coefficients and $(n-L)$ wavelet coefficients. These lifting coefficients are independent but with a non constant variance. To overcome this, we normalize the resulting coefficients by multiplying $[diag(\tilde{W}\tilde{W}^T)]^{-1/2}$ for each step where $\tilde{W}$ is the matrix associated with the transform. After normalization we denote these transformed empirical lifting coefficients as w. Since the empirical lifting coefficients are independent with constant variance σ^2, the joint distribution can be written as:

$$p(w|\tau, \triangle, \sigma^2) = \prod_j \prod_k p(w_{j,k}|\tau, \triangle, \sigma^2)$$

where w represents the vector of coefficients. As before, the posterior distribution of the parameter is

$$p(\tau, \triangle, \sigma^2|w) \propto \prod_j \prod_k p(w_{j,k}|\tau, \triangle, \sigma^2) \cdot \pi(\tau, \triangle, \sigma^2)$$

for a joint prior distribution π on $\tau, \triangle$ and σ^2. The distribution of a single lifting-based wavelet coefficient was taken to be

$$w_{j,k}|\tau, \triangle, \sigma^2 \sim N(\triangle q_{j,k}(\tau), \sigma^2/n).$$

Ogden and Lynch (1998) defined the mean function $q_{j,k}(\tau)$ as

$$q_{j,k}(\tau) = 2^{j/2}/n \times \begin{cases} 2^{-j}k - [n\tau], & \text{if } 2^{-j}k \leq \tau < 2^{-j}(k+1/2), \\ -(n2^{-j}(k+1) - [n\tau]), & \text{if } 2^{-j}(k+1/2) \leq \tau < 2^{-j}(k+1). \end{cases}$$

for $\tau \in [0,1]$. The $q_{j,k}$ function is continuous and piecewise linear, shaped like an inverted hat; it is zero outside the support of the corresponding wavelet and goes to a peak at the midpoint in support of $\psi_{j,k}$. To get the values of $q_{j,k}$ for all j and k we can apply lifting scheme to the vector of 0 and 1 ($[n\tau]-1$ zeros and $n-[n\tau]+1$ ones).

Hence, the posterior density will be

$$p(\tau, \triangle, \sigma^2|w) \propto \sigma^{-(n-1)/2}\pi(\tau, \triangle, \sigma^2) \times exp(-\sum_j \sum_k (w_{j,k} - \triangle\, q_{j,k}(\tau))^2/(2\sigma^2)) \tag{1}$$

The analytical result may vary according to the choice of the prior distribution π on $\tau, \triangle$ and σ^2.

Each empirical wavelet coefficient contains information regarding the changes of a function in a localized region. For this reason, we use more localized or higher level coefficients to compute the posterior density. The posterior density is thus computed using only the wavelet coefficients with dilation index $j \geq j_0$. Ogden and Lynch (1998) used Jeffreys' non-informative joint prior for $\tau, \triangle$ and σ such that

$$\pi(\tau, \triangle, \sigma^2) \propto 1/\sigma.$$

Thus the joint posterior density is given by

$$p(\tau, \triangle, \sigma^2|w) \propto \sigma^{-(n+1)/2} \times \\ exp(-\sum_{j \geq j_0} \sum_k (w_{j,k} - \triangle\, q_{j,k}(\tau))^2/(2\sigma^2))$$

Integrating out $\triangle$ and σ, the marginal posterior of change point τ will be

(2) $$p(\tau|w) \propto C^{-1/2}(A - B^2/C)^{-(n+9)/4}$$

where $A = \sum_{j \geq j_0} \sum_k w_{j,k}^2$, $B = \sum_{j \geq j_0} \sum_k w_{j,k}\, q_{j,k}(\tau)$, and $C = \sum_{j \geq j_0} \sum_k q_{j,k}^2(\tau)$. Mode of above posterior probability distribution of τ will give us the change point.

Simulation Results

Statistical simulation studies provide powerful tools for the analysis of many mathematical models and real data problems when analytical solution is not possible. To recommend a suitable choice of lifting-based coefficients, it is essential to explore different cases of variable sample size, variable noise variances, variable jump sizes and variable time series structures. The results of the study will provide us a guideline to choose lifting-based coefficients.

As mentioned earlier, there is no specific rule of selecting a specific number of coefficients for detection of change points. Due to lifting scheme of resolution L, we have L smooth and $(n-L)$ detail coefficients. These $(n-L)$ detail coefficients are assigned with artificial levels into fine and coarse levels. To detect change points we use only the fine level detail coefficients.

We decided on 1500 simulations. The principle was that for a fixed set of lifting coefficients, the percentage of detection should stay constant even if we increase the number of simulations. Hence, we run 1500 simulations for different choices of sample sizes, resolution levels and different sets of lifting coefficients.

(a) Comparison between DWT and Lifting:

For 1500 simulations we first compare DWT-based coefficients and lifting-based coefficients. In this case our sample size is $n = 2^7$ and the resolution level is two. There are no missing observations. We consider $d7$, $d6$ and $d5$ wavelet coefficients to calculate posterior probabilities. We divide our grid into equally spaced 128 points. We will choose five highest posterior probabilities matching with any five points on the grid. We consider it as a successful detection if the change point is detected among those five grid points. If the grid point corresponding to the highest posterior probability matches the change point then we will disregard the next four highest posterior probabilities. If the grid point matches the second highest posterior probability but not the first one then we will also count it as a successful detection and disregard next three highest posterior probabilities and so on. For $n = 2^7$, $d7$, $d6$ and $d5$ are the finest level coefficients of DWT which should have information about jumps. For lifting-based scheme we have four artificially assigned levels of detail coefficients. Out of these 126 detail coefficients, we choose all coefficients of two finest levels and 22 of 31 coefficients from the third finest level. These sets of lifting coefficients correspond to our choice of coefficients in DWT. It has a combination of detail coefficients at fine levels and coarse levels. This combination should be able to detect both small and large jumps. We simulate data from normal distribution with error terms for three different choices of variances. As the variance term increases, noise in the data also increases and detection of change points is expected to be difficult for the reason that it will be harder to judge as to whether jump is due to high level of noise or due to the presence of an actual change. We considered two different jump sizes for simulation. We can expect that with the big jump size, change point detection becomes relatively easier. In the following table we report the percentages of detections for different combinations of error variances and jump sizes. Increase in error variance implies that our simulated data are getting more noisy. Two different jump sizes are taken into account which in comparison to the noise variance of our generated data are small (i.e. jump size 1) and large (i.e. jump size 3) respectively.

Table 1: Comparison between DWT and lifting, $n = 2^7$***, no missing observations***

	Variance					
	0.5		1		1.5	
Size of jump	DWT	Lifting	DWT	Lifting	DWT	Lifting
1	54.1 %	93.7%	36.3%	89.8%	28.9%	87.4%
3	98.0%	93.7%	91.9%	91.1%	84.9%	90.7%

Table 1 shows that, as expected, coefficients based on lifting procedure perform consistently better than DWT coefficients under the presence of different noise variances. In fact, when the noise variance is high and jump size is small (i.e. when the detection is difficult), DWT coefficients can detect change points only for 28.9% cases whereas the detection is possible for 97.4% cases by lifting. Hence it may be recommended that the lifting-based procedure coefficients should be chosen for detecting change points.

We may now turn to provide a real data example to substantiate our observations based on simulation results.

St. Lawrence Streamflow Data

Our case study, for purposes of comparison, uses annual streamflow data from the St. Lawrence River at Ogdensbourg, New York from the years 1860 to 1950. These data were first analyzed by Rasmussen (2001) and recently by Seidou et al.(2007), both using the framework of Bayesian linear models. A description of the data is given in Rasmussen (2001). Here we note that both the methods i.e. DWT and Lifting found the mode of the posterior distribution for the time of the change point to be 1891. The original data had 90 observations, so we augmented it to n=128 or 2^7. We get the same year 1891 as the change point. Figure 1 is the plot with the original data and posterior distribution. Now, from the plot we find that the posterior density picks up at both the ends. Our procedure clearly detects a change point at 1891 if we consider years between 1860 and 1950.

Using lifting-based coefficients, we find the mode of the posterior distribution for the time of the change point to be 1891 (Figure 2). Since the original data had 90 observations, we do not need to augment it to 2^J to apply wavelet transform. Lifting transform can perform with any number of data points. Using 45% of fine detail coefficients we get our change point as 1891. Plot of the posterior pdf (Figure 1)shows multiple modes. Among them lifting-based coefficients detect the correct change point. It is not unusual to have a posterior pdf with multiple modes using lifting-based coefficients. There are two possible reasons for the multiple modes in the plot. Our procedure is useful for finding single change point in the data set but if there exists multiple change points, we can not be able to find them. Secondly, lifting based coefficients are based on polynomial regression which does not take into account the time dependence of the data which may contribute to the calculation of lifting coefficients and hence to the posterior pdf.

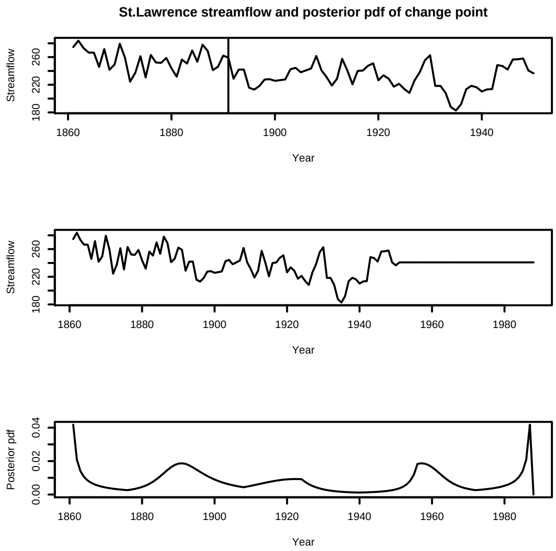

Figure 1:Plot of the St. Lawrence streamflow data (top), with change point at year 1891, augmented data (middle) and posterior pdf (bottom) with the augmented data on both sides

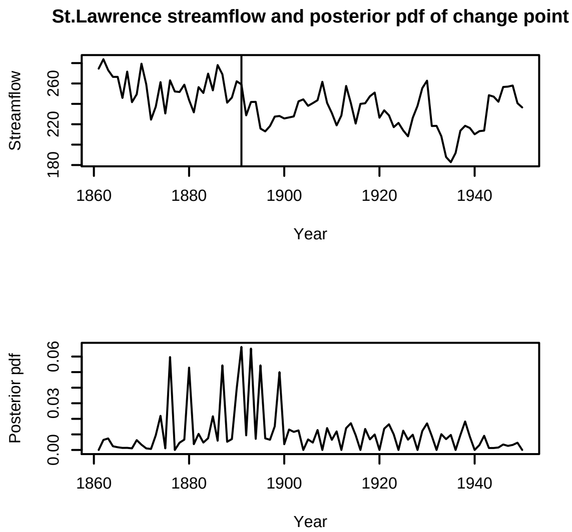

Figure 2:Plot of the St. Lawrence streamflow data (top) with change point at year 1891 and posterior pdf (bottom)

REFERENCES

Antoniadis, A., and Fan, J. (2001), "Regularization of wavelets approximations" (with discussion), *J. Amer. Statist. Assoc.*, 96, 939-967.

Bilen, C. and Huzurbazar, S.(2002), "Wavelet-based detection of Outliers in Time Series," *Journal Computational and Graphical Statistics*, 11(2), 311-327.

Carlin, B. P., Gelfand, A. E., and Smith, A. F. M. (1992), "Hierarchical Bayesian analysis of changepoint problems", *Appl. Stat.*, 41, 389-405.

Carter, R. L., and Blight, B. J. N. (1981), "A Bayesian Change Point Problem with an Application to the Prediction and Detection of Ovulation in Women", *Biometrics*, 37, 743-751.

Gelfand, A. E., Hills, S. E., Racine-Poon, A., and Smith, A. F. M. (1990), "Illustration of Bayesian inference in normal data models using Gibbs sampling", *Journal of the American Statistical Association*, 85, 972-85.

Heaton, T. J., and Silverman, B. W. (2008), "A wavelet- or lifting-scheme-based imputation method," *Journal of the Royal Statistical Society, Series B: Statistical Methodology*, 70, 567-587.

Jansen, M., Nason, G. P., and Silverman, B. W. (2001), "Scattered data smoothing by empirical Bayesian shrinkage of second generation wavelet coefficients", *Proc. SPIE*, 4478, 87-97.

Jansen, Mark J. (2004), "Semiparametric Bayesian inference of long-memory stochastic volatility models", *Journal of Time Series Analysis*, 25, 895-922.

Nunes, M., Knight, M., and Nason, G. P. (2006), "Adaptive lifting for nonparametric regression", *Statist. Comput.* , 16, 143-159.

Ogden, R.Todd, and Lynch, James D. (1998), Bayesian analysis of Change-point models, *Bayesian Inference in Wavelet Based Models*, P. Muller and B. Vidokovic eds., Springer-Verlag, New York.

Rasmussen, P. (2001), "Bayesian estimation of change points using the general linear model", *Water Resour. Res.*, 37, 2723-2731.

Richwine, J. E. (1996), "Bayesian estimation of change-points using Haar wavelets," Master's Thesis at the University of South Carolina.

Smith, A.F.M., (1975), "A Bayesian approach to inference about change-point in sequence of random variables," *Biometrika*, 62, 407416.

Stephens, D. A. (1994), "Bayesian retrospective multiple-changepoint identification", *Appl. Stat.*, 43, 159 178.

Sweldens, W. (1996), "The lifting scheme: a custom-design construction of biorthogonal wavelets", *Ap-*

plied and Computational Harmonic Analysis, 3, No. 2, 186-200.

Tanner, M. (1993), *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions*, Springer Verlag, New York.

Wang, Y. (1995), "Jump and Sharp Cusp Detection by Wavelets," *Biometrika*, 82, 385-397.

(ABSTRACT) — optional