Multiscale interpretation of taut string estimation and its connection to Unbalanced Haar wavelets

Haeran Cho · Piotr Fryzlewicz

Abstract We compare two state-of-the-art non-linear techniques for nonparametric function estimation via piecewise constant approximation: the taut string and the Unbalanced Haar methods. While it is well-known that the latter is multiscale, it is not obvious that the former can also be interpreted as multiscale. We provide a unified multiscale representation for both methods, which offers an insight into the relationship between them as well as suggesting lessons both methods can learn from each other.

Keywords Multiscale · Unbalanced Haar wavelets · Taut string · Nonparametric function estimation

1 Introduction

A canonical problem in nonparametric regression is the estimation of a one-dimensional function \( f \) from noisy observations \( y \) in the additive model

\[
y_t = f \left( \frac{t}{n} \right) + \epsilon_t, \quad t = 1, \ldots, n,
\]

where the observations \( \{y_t\}_{t=1}^n \) are taken on an equispaced grid. In the simplest version of (1), the noise \( \{\epsilon_t\}_{t=1}^n \) is assumed to be iid Gaussian, which is not necessarily a realistic assumption in many applied problems, but serves as an excellent benchmark for comparing estimation techniques and gauging their potential performance in more complex models, in the sense that if a method misperforms in the model (1) with iid Gaussian noise, there is normally little chance of it performing well in more complex settings.

In particular, the problem of estimating \( f \) using piecewise constant estimators has attracted considerable attention. The class of piecewise constant functions is flexible in approximating a wide range of function spaces (e.g. see DeVore 1998). Also, piecewise constant estimates are easy to interpret, as breakpoints in the estimate indicate significant changes in the mean of the data, while the constant intervals between the breakpoints represent regions where the mean remains approximately the same. It is well-known that if the underlying function \( f \) is spatially inhomogeneous, non-linear piecewise constant estimators perform better than linear estimators. Therefore, in what follows, we discuss non-linear approaches.

Without attempting to be exhaustive, we mention a few recent, well-performing estimation techniques. Wavelet thresholding estimation was first introduced in Donoho and Johnstone (1994), where the thresholded estimator was shown to be theoretically tractable and perform well. By using Haar wavelets, piecewise constant estimators are obtained. The CART methodology (Breiman et al. 1983, Classification and Regression Trees) performs greedy binary splitting to grow a partition, whose terminal nodes yield a piecewise constant estimator. In Engel (1997), a method for locally adaptive histogram construction was introduced, which is based on a tree of dyadic partitions and hence obtains a multiscale, piecewise constant estimator. Polzehl and Spokoiny (2000) presented Adaptive Weight Smoothing, a data-driven local averaging procedure with an adaptive choice of weights, which iteratively produces a piecewise constant estimator. More recently, methods involving the complexity-penalized likelihood op-
timization were proposed for estimating an unknown function by piecewise polynomials (Comte and Rozenholc 2004; Kolaczyk and Nowak 2005), which can be adopted to produce piecewise constant estimators.

In this paper, we are particularly interested in two methodologies, the Unbalanced Haar (UH) technique (Fryzlewicz 2007) and the taut string (TS) based estimation (see e.g. Barlow et al. 1972 and Davies and Kovac 2001). Both techniques are computationally fast, achieve theoretical consistency, and exhibit excellent performance in numerical simulation studies. The former involves the decomposition of the data with respect to orthonormal Haar-like basis vectors with jumps not necessarily in the middle of their support, while the latter finds a piecewise constant estimator via penalizing its total variation.

Our aim in this paper is to compare these two methods and discover links between them. The UH technique is multiscale by nature (Fryzlewicz 2007), yet the multiscale character of the TS technique is less obvious, and has not been noted in the literature before. Thus, in order to establish links between the two methods, we first provide an interesting multiscale interpretation of the TS technique. This then enables us to better understand similarities and differences between the UH and TS techniques, and establish a unified estimation methodology, which both the UH and the TS technique are instances of. Finally, taking advantage of this common framework, we derive lessons which either method can learn from the other.

The rest of paper is organized as follows. In Sect. 2, we provide a description of the UH and TS techniques, as well as flowcharts of their algorithms, which offer an insight into the relationship between their physical interpretations. Then follows the comparison study, including the understanding of the two techniques in the framework of breakpoint detection (Sect. 3). We conclude the discussion in Sect. 4 by listing some ways of possible improvement and extension for both techniques, which suggest avenues for further research.

2 Unbalanced Haar and taut string techniques

In this section, we give an overview of the UH and TS techniques. In particular, we emphasize the explicit multiscale nature of the UH methodology. One contribution of this paper is to cast a new light on the TS technique via its new multiscale interpretation, which is achieved by introducing multiscale algorithms for both methods in Sect. 2.3. These new algorithms are key to understanding and comparing the two techniques.

2.1 Unbalanced Haar technique

The UH technique consists of three steps: the transformation of \( \{ y_i \}_{i=1}^{n} \) with respect to an adaptively chosen UH wavelet basis, hard-thresholding of the wavelet coefficients, and the inverse UH transformation of the thresholded coefficients to yield an estimate of \( f \). For the principles of traditional wavelet thresholding estimation (without the adaptive basis selection), the reader is referred to Vidakovic (1999).

The UH wavelet basis vectors were first studied in Girardi and Sweldens (1997) as an extended version of classical Haar wavelet vectors, the extension being that the breakpoint was permitted to occur anywhere within their support. Let \( s \) and \( e \) denote the start and end of a generic interval, respectively, and let \( b \) denote the location of the breakpoint. Then, a UH vector on the interval \([s, e]\) with breakpoint \( b \), \( \psi_{s,b,e} \), is defined as \( \psi_{s,b,e}(l) = \{ \frac{1}{e-b} - \frac{1}{e-s+1} \}^{1/2} \psi_{[s,b]}(l) - \{ \frac{1}{e-b} - \frac{1}{e-s+1} \}^{1/2} \psi_{[b+1,e]}(l) \), for \( s \leq l \leq e \). Classical Haar wavelet vectors are a special case with \( b = (s + e - 1)/2 \).

Denote the vector of observations as \( \tilde{y} = (y_1, \ldots, y_n)^{T} \) and its sub-vector on a generic support \( \{s, \ldots, e\} \) as \( \tilde{y}^{s,e} = (y_s, \ldots, y_e)^{T} \). Noting that on a given support, the choice of breakpoints \( b \) defines the choice of a UH basis, one way of UH basis selection is presented in Fryzlewicz (2007). The first breakpoint \( b_{1,1} \) is chosen from \( \{1, \ldots, n\} \) such that the inner product between \( \tilde{y} \) and \( \tilde{\psi}_{1,1,n} \) is maximized in absolute value, i.e. \( b_{1,1} = \arg \max_{b \in \{1, \ldots, n\}} |\langle \tilde{y}, \tilde{\psi}_{1,1,n} \rangle| \). The explicit expression for the UH wavelet coefficient is given in (5). The next breakpoints are chosen similarly on the supports defined by the previously chosen breakpoint, \( \{1, \ldots, b_{1,1}\} \) and \( \{b_{1,1}+1, \ldots, n\} \), and the same procedure is repeated until it is no longer possible to divide any support into two. Then \( \tilde{y} \) is transformed with respect to the orthonormal basis defined by the selected breakpoints. The next step is the hard-thresholding of the wavelet coefficients by setting to zero those which fall below the universal threshold \( \sigma \sqrt{2 \log n} \). In practice the standard deviation of the noise is unknown but can be estimated as the median of the sequence \( \{|y_{i+1} - y_{i}|/\sqrt{2}\}_{i=1}^{n-1} \) divided by the 0.75-quantile of the standard normal distribution (which is approximately equal to 0.6745). Finally the inverse transform is taken to obtain the final estimate \( \hat{f}_{UH} \), which is shown to be a mean-square consistent estimator for a wide range of functions, uniformly over those UH bases (however they have been selected) which are not “too unbalanced” in the sense that each basis vector should satisfy

\[
\max \left\{ \frac{b - s + 1}{e - s + 1}, \frac{e - b}{e - s + 1} \right\} \leq p, \tag{2}
\]

for a fixed \( p \in [1/2, 1) \). Thus, in practice, the maximisation of the inner products as described above is performed in such a way that each time, the maximum is only taken over those wavelets which satisfy condition (2), to ensure mean-square consistency of the resulting estimator.

We note that at the outset of the UH basis selection procedure, the entire observation vector is scanned in the search