Nonparametric Regression with Random Sampling

E. K. Chicken

Department of Statistics, Florida State University, Tallahassee, FL 32306-4330, USA

ABSTRACT

Wavelet methods have been applied to nonparametric regression problems with much success. Usually, the problem assumes fixed, equispaced sampling points. When the number of data points is dyadic, the discrete wavelet transform (DWT) works especially well, providing fast convergence to the unknown underlying function in a computationally efficient manner. This poster shows that the DWT method can be applied to the problem when the data are random placed in an interval or follow a random process. Specific conditions on the distributions and processes are given that enable the DWT, through the use of block thresholding, to come within a constant factor of the optimal minimax rate of convergence.

1 Introduction

When the positions of the sampling points are equispaced, wavelets excel in the areas of spatial adaptivity, optimality, and low computational cost. For fixed, nonequispaced sample points, Cai and Brown (1998) investigated wavelet methods via an approximation approach. Hall and Turlach (1997) used interpolation methods to deal with samples with random design. Unfortunately, these methods are more complex from a computational standpoint than their equispaced counterparts.

To maintain computational efficiency, it is desirable to use equispaced methods to analyze the data. Cai and Brown (1999) have examined convergence rates when the positions of the sample points are distributed as independent uniform random variables. Using term-by-term thresholding, they showed that equispaced wavelet methods can be directly applied to the nonequispaced data without a loss in the rate of convergence, i.e., to within a logarithmic factor of the optimal minimax convergence rates found by Donoho and Johnstone (1994). This method maintains the computational efficiency and simplicity of the equispaced algorithm.

Here, the results of Cai and Brown (1999) for nonparametric regression with randomly placed sampling points are improved upon.

- By using the block thresholding of wavelet coefficients method proposed by Hall et al. (1999) and refined by Cai (1998), the logarithmic penalty in the convergence rate associated with term-by-term thresholding has been removed.
- The scope of the function spaces has been enlarged to include not only the Hölder spaces studied by Cai and Brown, but many Besov spaces as well.
- Moment conditions are given that define a class of random distributions for which these results hold.
- This fast convergence rate is also attained when the sample points come from a random process. Conditions are given defining the processes for which this method will result in optimal convergence rates.
- Spatial adaptivity is maintained, and the computational cost remains low since the equispaced algorithm is used.

2 Model

The usual model for nonparametric regression is

$$y_i = f(x_i) + \sigma\varepsilon_i, \qquad i = 1, 2, \ldots, n,$$

where the noise ε_i is iid standard normal random variables and σ is known. To use the DWT, we require that $n = 2^J$. We will modify this to the more general problem

$$y_i = f(X_i) + \sigma\varepsilon_i, \qquad i = 1, 2, \ldots, N.$$

There are two schemes.

- If the sample points X_i are randomly placed in the interval, then N is a constant and we may assume it to be dyadic. The X_i in this case are iid.
- If the X_i follow a random process, then N is a random variable.

In each case, we assume the distribution of the X_i (and hence N in the random process case) are independent of the noise ε_i.

3 Wavelets and the Estimator

ϕ and ψ will represent the father and mother wavelets, respectively. Both are assumed to be compactly supported and periodized to the interval $[0,1]$. The wavelet coefficients of a function f are the usual inner product: $\xi_{jk} = \langle f, \phi_{jk}\rangle, \quad \theta_{jk} = \langle f, \psi_{jk}\rangle$. In terms of resolution, the ξ_{j_0k} coefficients represent the coarsest, smoothest portions of f, and the θ_{jk} are the coefficients representing the detailed structure of f. Rather than compute the coefficients in the above manner (f is unknown), we use the DWT to estimate them.

By ordering the sample points and performing the equivalent reordering of the y_i's and ε_i's, the signal now looks like

$$y_i = f(X_{(i)}) + \sigma\varepsilon_i,$$

where the $X_{(i)}$ are either the ordered random variables or the arrival times of the random process. Using the appropriately relabeled y_i's, the observed data vector is now $(X_{(1)}, y_1), (X_{(2)}, y_2), \ldots, (X_{(N)}, y_N)$. These randomly spaced data points will be treated as equidistant in the wavelet algorithm. Each $X_{(i)}$ will be replaced with $\frac{i}{n+1}$. Provided that $N = 2^J$, the DWT can be applied to the vector $y = \{y_1, y_2, \ldots, y_N\}$. In the random process case, if $N \neq 2^J$, the signal will be extended to the next multiple of 2 by reflecting the signal about its endpoint. The underlying, unknown function is then assumed to be extended similarly on $[1, 1+\frac{k}{N}]$, where $k = 2^J - N$ is the number of points added to the signal.

The DWT coefficients $\tilde{\theta}_{jk}$ are thresholded in blocks using the James-Stein threshold. At each resolution level j, the $\tilde{\theta}_{jk}$ are grouped into non-overlapping blocks of length L. Let B_{jb} be the indices of the $\tilde{\theta}_{jk}$ in the bth block in resolution level j. Given $N = n$, the James-Stein threshold rule is

$$\hat{\theta}_{jk} = \left(1 - \frac{\lambda L \sigma^2}{n\sum_{i\in B_{jb}} \tilde{\theta}^2_{ji}}\right)_+ \cdot \tilde{\theta}_{jk},$$

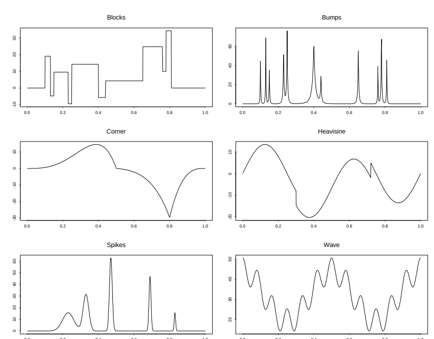
for all k in the bth block of resolution level j. Cai (1998) has shown that the optimal rate of convergence for equispaced samples is attained when $L = \log(n)$ and $\lambda = 4.50524$. The inverse DWT is applied to the thresholded coefficients to give the estimator $\hat{f}$ of f.

4 Function Spaces

The functions to be studied reside in Besov and Hölder spaces. These may be defined by their wavelet coefficients as in Meyer (1990). If ξ_{j_0k} and θ_{jk} are the wavelet coefficents of f, then $f \in B^\alpha_{p,q}$ whenever its norm is finite:

$$\|f\|_{B^\alpha_{p,q}} = \|\xi_{j_0\cdot}\|_p + \left(\sum_{j=j_0}^{\infty}\left(2^{j\left(\alpha+\frac{1}{2}-\frac{1}{p}\right)}\|\theta_{j\cdot}\|_p\right)^q\right)^{\frac{1}{q}},$$

when $p, q < \infty$. (For $p = \infty$ of $q = \infty$, make the appropriate modifications.) The Hölder space Λ^α is a special case of a Besov space. Specifically, $\Lambda^\alpha = B^\alpha_{\infty,\infty}$. This paper considers functions whose Besov (or Hölder) norms are bounded by a finite M. Define $F^\alpha_{p,q}(M)$ to be $B^\alpha_{p,q}(M) \cap \Lambda^{\frac{2\alpha}{2\alpha+1}}(M)$. For large ranges of p and q, the assumption that f be in $\Lambda^{\frac{2\alpha}{2\alpha+1}}(M)$ is redundant. These spaces cover a very wide range irregluar functions. In simulation, the following functions from Donoho and Johnstone (1994) will be used.



5 Rates of Convergence

Donoho and Johnstone (1994) showed that the best rate (in terms of minimax) for a diagonal projection estimator such as the one presented here is

$$E\|f - \hat{f}\|_2^2 \sim n^{-2\alpha/(2\alpha+1)},$$

where n is the number of places the function is sampled, and α is the smoothness parameter of the function space. Using term-by-term thresholding on equispaced samples, the DWT estimate is within a $\log n$ term of this rate. By applying block thresholding, this rate is improved, and not just for equispaced designs.

Theorem 1 *Let X_i be iid positive random variables with*

$$EX_i = 1/n, \quad \textit{var } X_i \le c/n^2, \textit{ and } S_i = \sum_i X_i.$$

Suppose a sample $\{(S_1,y_1),(S_2,y_2),\ldots,(S_N,y_N)\}$ is collected with $y_i = f(S_i) + \sigma\varepsilon_i$. Further, ψ has r vanishing moments, S_i and ε_i are independent, $\alpha \in [\frac{1}{2}, r]$, $2 \le p \le \infty$, $1 \le q \le \infty$, and $0 < M < \infty$. Then for $\hat{f}$ as constructed above, and on each interval,

$$\sup_{f\in F^\alpha_{p,q}(M)} E\|\hat{f} - f\|_2^2 \le Cn^{-\frac{2\alpha}{2\alpha+1}}.$$

Theorem 2 *Let X_i be iid random variables on $[0,1]$ whose density is bounded,*

$$\left|EX_{(i)} - i/(n+1)\right| \le c/\sqrt{n}, \textit{ and var } X_{(i)} \le c/n.$$

Suppose a sample $\{(X_{(1)},y_1),(X_{(2)},y_2),\ldots(X_{(n)},y_n)\}$ is collected with $y_i = f(X_{(i)}) + \sigma\varepsilon_i$. Further, ψ has r vanishing moments, $\alpha \in [\frac{1}{2}, r]$, $2 \le p \le \infty$, $1 \le q \le \infty$, and $0 < M < \infty$. Then for $\hat{f}$ as constructed above, and on each interval,

$$\sup_{f\in F^\alpha_{p,q}(M)} E\|\hat{f} - f\|_2^2 \le Cn^{-\frac{2\alpha}{2\alpha+1}}.$$

We see that the minimax rate is attained to within a constant in each case. Note also that if the space of functions is just $\Lambda^\alpha(M)$, this exceeds the performance of previous estimators due to the removal of the $\log$ penalty term.

6 Simulations

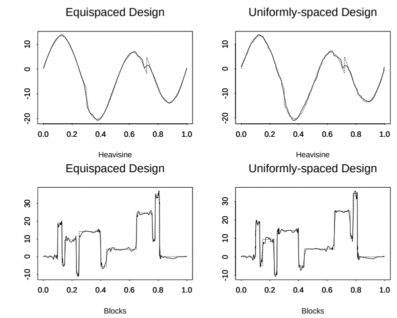
6.1 Numerical Results - I

Here we specifically look at design points uniformly placed on an interval and design points following a Poisson process. "B" refers to the block thresholded wavelet estimator, "V" to the term-by-term estimator, "U" to the uniform scheme, "P" to the Poisson process scheme.

Function n	BE	BU	BP	VU	BE vs BU	VU vs BU	BE vs BP	BU vs BP
Blocks								
512	3.02	3.02	3.05	5.33	0%	-43%	0%	0%
1024	1.85	1.85	1.83	3.85	0%	-52%	0%	0%
2048	1.15	1.09	1.07	2.67	-5%	-59%	-7%	0%
4096	0.64	0.64	0.64	1.82	0%	-65%	0%	0%
8192	0.37	0.37	0.37	1.20	-2%	-70%	-1%	0%
Bumps								
512	3.15	3.24	3.29	8.33	0%	-61%	4%	0%
1024	1.66	1.98	2.01	6.03	20%	-67%	21%	0%
2048	0.93	1.21	1.20	4.01	31%	-70%	29%	0%
4096	0.51	0.70	0.70	2.51	38%	-72%	38%	0%
8192	0.29	0.38	0.39	1.51	35%	-75%	35%	0%
Corner								
512	0.40	0.51	0.51	0.44	27%	17%	27%	0%
1024	0.20	0.28	0.28	0.25	39%	10%	38%	0%
2048	0.10	0.14	0.14	0.15	47%	-5%	46%	0%
4096	0.05	0.07	0.08	0.10	52%	-25%	57%	0%
8192	0.03	0.04	0.04	0.06	40%	-36%	48%	6%

6.2 Numerical Results - II, Graphical Results

Function n	BE	BU	BP	VU	BE vs BU	VU vs BU	BE vs BP	BU vs BP
Heavisine								
512	0.59	0.74	0.82	0.62	26%	19%	40%	11%
1024	0.40	0.45	0.48	0.43	14%	6%	21%	7%
2048	0.24	0.26	0.28	0.28	6%	-9%	16%	10%
4096	0.14	0.15	0.15	0.20	5%	-24%	9%	4%
8192	0.08	0.08	0.09	0.13	7%	-37%	11%	3%
Spikes								
512	1.05	1.75	1.80	3.34	67%	-48%	71%	0%
1024	0.56	0.96	0.96	2.11	72%	-55%	72%	0%
2048	0.33	0.50	0.50	1.28	50%	-61%	50%	0%
4096	0.18	0.26	0.26	0.75	38%	-66%	38%	0%
8192	0.09	0.13	0.13	0.44	39%	-71%	40%	0%
Wave								
512	0.98	1.46	1.49	2.56	50%	-43%	52%	0%
1024	0.39	0.66	0.71	1.65	70%	-60%	83%	8%
2048	0.18	0.31	0.34	1.00	69%	-69%	87%	10%
4096	0.10	0.16	0.17	0.59	61%	-73%	72%	7%
8192	0.06	0.09	0.10	0.34	46%	-73%	54%	5%



7 Conclusions

- Want to use DWT because of its adaptivity and speed
- Previously shown to be within a log term of minimax for uniform design
- Now, have shown to be within a constant of minimax for design points
  - Randomly distributed on an interval (theorem 2)
  - Following a random process (theorem 1)
- Results hold on more general space of functions than previously considered

References

Cai, T. (1998). Adaptive wavelet estimation: A block thresholding and oracle inequality approach. *Ann. Statist.* **27** 898–924.

Cai, T. and Brown, L. (1998). Wavelet shrinkage for nonequispaced samples. *Ann. Statist.* **26** 1783–1799.

Cai, T. and Brown, L. (1999). Wavelet estimation for samples with random uniform design. *Statist. Probab. Lett.* **42** 313–321.

Donoho, D. and Johnstone, I. (1994). Ideal spatial adaptation via wavelet shrinkage. *Biometrika* **81** 425–455.

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

Hall, P. and Turlach, B. (1997). Interpolation methods for nonlinear wavelet regression with irregularly spaced design. *Ann. Statist.* **25** 1912–1925.

Meyer, Y. (1990). *Ondelettes et Opérateurs: I. Ondelettes.* Hermann et Cies, Paris.

Some of this work to appear in *The Journal of Statistical Planning and Inference*.