

REJOINDER OF: TREELETS—AN ADAPTIVE MULTI-SCALE BASIS FOR SPARSE UNORDERED DATA

BY ANN B. LEE, BOAZ NADLER AND LARRY WASSERMAN

*Carnegie Mellon University, Weizmann Institute of Science
and Carnegie Mellon University*

We are grateful to all of the discussants for their thoughtful comments. Their remarks have added significant insight and perspective on the work. As a variety of issues have been raised, we have organized our rejoinder according to main topics that have been brought up by the discussants.

1. A multiresolution transform guided by the second-order statistics of the data. The treelet transform is a multiresolution transform that allows one to represent the original data in an alternative form. Rather than describe the data in terms of the original set of covariates, we perform a series of rotations which gradually reveal the hierarchical grouping structure of the covariates. The idea is very similar to the Grand Tour by [Asimov \(1985\)](#). The treelet transform is a tour “guided” by the covariance structure of the data.

Once the treelet transform has been completed, there are multiple ways of choosing an orthogonal basis (see Section 2.2). We never directly discard residual terms as noise. These terms are in fact an integral part of the final representation. In the simulated example of Section 4.2, most of the detail variables represent noise with small expansion coefficients; consequently, only certain coarse-grained variables are chosen for regression. In general, however, detail variables may convey crucial information. The latter point is illustrated in Sections 5.1 and 5.3, where we use the standard choice of one scaling term and $p - 1$ difference terms; that is, an observation \mathbf{x} is decomposed according to

$$\mathbf{x} = s\phi + \sum_{i=1}^{p-1} d_i\psi_i,$$

where the first term is a coarse-grained representation of the signal and the d -terms represent “differences” between node representations at two consecutive levels in the tree.

2. Orthogonal versus overcomplete bases. Tibshirani and Bickel/Ritov correctly point out that one need not restrict attention to one treelet level. An overcomplete dictionary of treelets can certainly be used for prediction. The “tree