

Title: **Adapting to Unknown Sparsity by Controlling the False Discovery Rate**

Author(s): **Felix Abramovich, Yoav Benjamini, David Donoho and Iain Johnstone**

Technical Report number (Dept. of Statistics, Stanford Univ.): **2000-19**

Date: **May 2000**

Abstract:

We attempt to recover a high-dimensional vector observed in white noise, where the vector is known to be sparse, but the degree of sparsity is unknown. We consider three different ways of defining sparsity of a vector: using the fraction of nonzero terms; imposing power-law decay bounds on the ordered entries; and controlling the ℓ^p norm for p small. We obtain a procedure which is asymptotically minimax for ℓ^r loss, simultaneously throughout a range of such sparsity classes.

The optimal procedure is a data-adaptive thresholding scheme, driven by control of the *False Discovery Rate* (FDR). FDR control is a recent innovation in simultaneous testing, in which one seeks to ensure that at most a certain fraction of the rejected null hypotheses will correspond to false rejections.

In our treatment, the FDR control parameter q also plays a controlling role in asymptotic minimaxity. Our results say that letting $q = q_n \rightarrow 0$ with problem size n is sufficient for asymptotic minimaxity, while keeping fixed $q > 1/2$ prevents asymptotic minimaxity.

To our knowledge, this relation between ideas in simultaneous inference and asymptotic decision theory is new.

Our work provides a new perspective on a class of model selection rules which has been introduced recently by several authors. These new rules impose completeness penalization of the form $2 \cdot \log(\text{potential model size} / \text{actual model size})$. We exhibit a close connection with FDR-controlling procedures having q tending to 0; this connection strongly supports a conjecture of simultaneous asymptotic minimaxity for such model selection rules.