# Local Bayesian False Discovery Rate Wavelet Shrinkage with $\Gamma$-minimax prior

HAESONG CHOI AND BRANI VIDAKOVIC

**Abstract.** This study developed Local developed Bayesian False Discovery Rate (LFDR) wavelet shrinkage with Gamma minimax prior, which is well-suited to noisy signals with low signal-to-noise ratio (SNR) (Angelini et al. (2004)). To solve the multiple hypothesis problem in the wavelet domain, this study linked LFDR in Efron (2004) with a developed level-adaptive Bayesian model. For estimating the performance of this wavelet shrinkage, this study compared it with global wavelet shrinkages and showed the superiority of the proposed shrinkage.

**KEY WORDS:** Wavelet Shrinkage; Multiple Hypotheses Testing, False Discovery Rate, Local False Discovery Rate, $\Gamma$-minimax family

## 1 Introduction

Since the early 1990's, researchers have devoted considerable attention to wavelet shrinkage procedures as an effective data analytic tool in a range of areas of theoretical and applied science. Donoho and Johnstone (1994) introduced simple and powerful procedures for estimating the unknown function $f$ by using thresholding in the wavelet domain. Especially, for the thresholding part, they proposed universal and SURE threshold. Nason (1996) proposed the cross-validation method, and Ogden and Parzen (1996) suggested recursive likelihood ratio tests for the thresholding method. Moreover, other researchers approached the wavelet shrinkage problems from a Bayesian standpoint. Since Bayes rules are shrinkers, the Bayes rules can be established to mimic the thresholding rules: to slightly shrink the large coefficients and heavily shrink the small coefficients. In addition, Bayes rules result from realistic statistical models on wavelet coefficients and such models allow for incorporation of prior information about the true signal. Furthermore, most Bayes rules can be easily either computed by simulation or expressed in a closed form. Vidakovic (1999) proposed a bayesian shrinkage of the wavelet procedure in the base of the fact that Bayes rules are shrinkers. Vidakovic and Ruggeri (2001) introduced a level-adaptive Bayesian model, which results in simple optimal shrinkage in fast wavelet denoising. Our paper particularly focused on a leve-adaptive Bayesian model,

and Bayesian model, which particularly applicable to noisy signals with a low signal-to-noise ratio through Angelini et al. (2004). They proposed a level dependent shrinkage rule that turns out to be the $\Gamma$-minimax rule for incorporating the prior information on the wavelet coefficients,which is suited to noisy signal data.

However, these studies have drawn attention to the problem of the multiple hypothesis. Multiple hypothesis is another significant current discussion in a number of sicentific fields, which require the analysis of high-dimensional data sets. This anlysis often involves performing simultaneous hypothesis tests on a variety of variables. Classical multiple hypothesis testing used to utilize the probability of making Type 1 errors among the family of hypothesis tests (family-wise error rate) as a longstanding statistical significance. However, since this family-wise error rate is not adaptive for large multiple hypothesis, researchers have proposed the well established formulations of statistical significance, introducing the false discovery error rate (FDR) (Benjamini and Hochberg, 1995). Efron et al. (2001), Storey (2003) adjusted a bayesian framework in this approach. This framework yields the probability that each null hypothesis is true given the observed data. Therefore, this empirical Bayes appraoch allows the frd analysis to proceed with a minimum of frequentist or Bayesian modeling assumptions. Moreover, Efron (2004) introduced the local false discovery rate methods, which is developed from empirical bayes analysis of a microarray experiment (Efron et al., 2001). Our paper introduced the model Angelini et al. (2004), and developed a level-adaptive Bayesian model with the prior from Angelini et al. (2004). Then, we link local FDR in Efron (2004) with a developed level-adaptive Bayeisan model.

The article is organized as follows. Section 2 introduces $\Gamma$-minimax optimality of Uniform prior, while Sction 3 introduces methodology, reviewing the previous literature. Section 4 presents simulational results for four standard signals and an application of the BaFDR to an atomic force microscopy signal. Conclusions are provided in Section 5.

## 2 $\Gamma$-minimax Optimality of Uniform Prior

In the present paper, we propose a $\Gamma$-minimax approach to discuss the prior on $\theta$ in the local false discovery rate. Let $\mathbf{X}$ be a random variable whose follows $f(\mathbf{x} \mid \theta), \theta \in \Theta$. To make an inference about the parameter $\theta$, $\delta(x)$, which is a **decision procedure** regarding $\theta$ is made. The set of allowable decisions is the class of all $actions$, which all possible realizations of $\delta(x)$. The **loss function** $L(\theta, a)$ maps $\Theta \times A$ into the set of real numbers. The loss function defines the cost to the statistician, reflecting that the closer $a$ is to $\theta$, the less loss is incurred. In a decision theoretic analysis, a **risk function** $R(\theta, \delta)$ characterizes the quality of an estimater $\delta(\mathbf{x})$ of $\theta \in \Theta$; that is, the risk function, a function of $\theta$, is defined as

$$R(\theta, \delta) = E^{X|\theta} L(\theta, \delta(X)), \tag{1}$$

Let $D$ be the set of all measurable decision rules and $\Gamma$ be a family of prior distributions on the parameter

space $\Theta$. Under the minimax principle, the rule $\delta^0$ is $\Gamma$-minimax if it minimizes $\sup_{\pi\in\Gamma} r(\phi,\delta)$; specifically,

$$\delta^0 = arg \inf_{\delta\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\pi,\delta). \tag{2}$$

where $r(\phi,\delta) = E^\theta R(\theta,\delta)$ is the Bayes risk and when the decision problem has a value, then the $\Gamma$-minimax solution coincides with the Bayes rule with respect to the least favorable prior (Angelini and Vidakovic, 2004). Therefore, A rule $\delta^*$ is characterized by

$$\inf_{\delta\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\pi,\delta) = \sup_{\pi\in\Gamma} \inf_{\delta\in\mathcal{D}} r(\pi,\delta^*). \tag{3}$$

In a realistic Bayesian model in the wavelet domain, a prior distribution should be taken into account, which incorporate prior information. A prior distribution on the wavelet coefficient $\theta$ is given by

$$\pi(\theta) = \epsilon_0\delta_0 + (1-\epsilon_0)\xi(\theta), \tag{4}$$

where $\delta_0$ is a point mass at zero, $\xi$ is a symmetric and unimodal distribution on the parameter space $\Theta$ and $\epsilon_0$ is a fixed parameter in [0,1] (Abramovic et. al, 1998; Vidakovic, 1998; Vidakovic and Ruggeri, 2001). The point mass at zero, $\delta_0$, is the shrinkage inducer whereas $\xi$ is a spread distribution that models wavelet coefficients with large magnitudes. However, a single prior distribution $pi$ onthe parameter space $\Theta$ always contains incomplete specification of the prior, which several distribution can match the prior information. On the partial knowledge about the signal, $\Gamma$ family distribution was proposed as one of the plausible priors. Assuming that the parameter space $\Theta$ is bounded, Angelini and Vidakovic (2004) showed for $m \in \Theta$, the least favorable distribution is the uniform on $[-m,m]$ contaminated by a prior mass at zero through the following model

$$\begin{cases} d|\theta & \sim & \mathcal{N}(\theta,1) \\ \theta & \sim & \pi(\theta)\in\Gamma \\ \mathcal{L}(\theta,\delta) & = & (\theta-\delta)^2 \text{ Squared Error Loss.} \end{cases} \tag{5}$$

Under (5) where $\Gamma$ is defined in (3), we have

$$\inf_{\delta\in\mathcal{D}} \sup_{\pi\in\Gamma} r(\pi,\delta) = \sup_{\pi\in\Gamma} \inf_{\delta\in\mathcal{D}} r(\pi,\delta).$$

The associated $\Gamma$-minimax rule is the Bayes rule with respect to the least favorable prior

$$\pi(\theta) = (\epsilon_0 + (1-\epsilon_0)\alpha_0)\delta_0 + (1-\epsilon_0)\sum_{k=1}^{p}\alpha_k\mathcal{U}[-m_k,m_k], \tag{6}$$

where $\alpha_k = \alpha_k(\epsilon_0) \geq 0$, $\sum_{k=0}^{p}\alpha_k = 1$, $m_k = m_k(\epsilon_0)$ s.t. $0 < m_1 < m2 < \ldots < m_p = m$. The corresponding $\Gamma$-minimax rule is given by

$$\delta_\pi(d) = d - \frac{(\epsilon_0 + (1-\epsilon_0)\alpha_0)d\phi(d) - (1-\epsilon_0)\sum_{k=1}^{p}\frac{\alpha_k}{2m_k}\left(\phi(d+m_k)-\phi(d-m_k)\right)}{(\epsilon_0 + (1-\epsilon_0)\alpha_0)\phi(d) + (1-\epsilon_0)\sum_{k=1}^{p}\frac{\alpha_k}{2m_k}\left(\Phi(d+m_k)-\Phi(d-m_k)\right)}, \tag{7}$$

where $\phi$ and $\Phi$ denote the density and the cumulative distribution function of the standard normal random variable and $\mathcal{U}$ denotes the uniform distribution.

Moreover, for any $\epsilon_0$ there exists $m^* = m^*(\epsilon_0)$ such that, for any $m \leq m^*$, the least favorable prior is of the form

$$\pi(\theta) = \epsilon_0 \delta_0 + (1 - \epsilon_0)\mathcal{U}[-m, m] \tag{8}$$

and the $\Gamma$-minimax rule is given by

$$\delta_\pi(d) = d - \frac{\epsilon_0 d\phi(d) - \frac{1-\epsilon_0}{2m}\left(\phi(d+m) - \phi(d-m)\right)}{\epsilon_0 \phi(d) + \frac{1-\epsilon_0}{2m}\left(\Phi(d+m) - \Phi(d-m)\right)}. \tag{9}$$

The value of $m^*(\epsilon_0)$ such that (8) holds is the largest value of m for which the maximum of $\frac{1}{z}\int_0^z R(v, \delta_\pi)dv$ is achieved at $z = m$, where $R(\cdot, \delta_\pi)$ represents the frequentist risk of the rule $\delta_\pi$.

# 3 Methodology

## 3.1 Local False Discovery Rate (lfdr) in the Wavelet Domain

Next, we describe the statistical model. Suppose the observed data is $y_i$, which consists of an unknown signal $f_i$, and random noise $\epsilon_i$.

$$y_i = f_i + \epsilon_i' \quad i = 1, \ldots, n. \tag{10}$$

or, in vector notation, $\mathbf{y} = \mathbf{f} + \mathbf{e}$. It is assumed that $x_i,\ i = 1, \ldots, n$ are equispaced, and the random noise $e_i$ are independent and identically distributed normal, with zero mean and variance $\sigma^2$. Our interest relies in the estimation of the signal function $\mathbf{f}$ using the observations $\mathbf{y}$. After decomposing a linear and orthogonal wavelet transform $\mathbf{d} = \mathbf{W}\mathbf{y}$ of the noise data $\mathbf{y}$, expression (10) becomes

$$d_{jk} = \theta_{jk} + \epsilon_{jk}, \tag{11}$$

where $d_{jk}, \theta_{jk}, \epsilon_{jk}$ are the $j, k$-th coordinates in the traditional scale/shift wavelet-enumeration of vectors $W\mathbf{y}, W\mathbf{f}$ and $W\mathbf{e}$, respectively ($j = 0, \ldots, J; k = 0, \ldots, 2^j - 1$). The independence of the coefficients $d_{jk}$ is assumed in this wavelet transformations. The orthogonality of $\mathbf{W}$ and Gaussianity of $\mathbf{e}$ implies Gaussianity of $\epsilon$. Hence, every discrete wavelet coefficient $d_{jk}$ contributes noise of variance $\sigma^2$. Shrinkage is performed term by term according to diagonal Bayesian approach to wavelet regression. Meanwhile, as a result of the parsimonious representation of the wavelet, yielded by the frequency domain, only a few $d_{jk}$ contain significant information about the real signal. Thus, extracting these significant coefficients is extremely important by thresholding the $d_{jk}$.

From the statistical viewpoint, Donoho and Johnstone (1994), Ogden and Parzen (1996) suggested that thresholding selection in the wavelet domain is closely related to hypothese testing, where all coefficients, at most 20, are tested whether it is significantly different from zero, i.e., $H_0 : \theta_{jk} = 0$. If $H_0$ is accepted, then the coefficient $d_{jk}$ is replaced with 0 in the model. In the contrary, $H_0$ is significantly rejected, $d_{jk}$ is survived. After all $n$ tests are performed, the only survived coefficients are recovered to the domain of the original data. But, as the number of hypotheses is large, people suffered from the loss of power unlike the only one hypothesis tested. For example, when we test 1023 hyptheses at level $\alpha$=0.05 (family-wise error rate), about 51 (1023·0.05) hypotheses are expected to have $p$-value less than $\alpha$=0.05, which are spuriously significant. Therefore, the stronger control of the error rate was required, which went beyond this framework: Bonferroni correction. This approach solved the problem, replacing $p_i \leq \alpha$ with $p_i \leq \alpha/N$. However, the larger $N$ we have, the smaller $\alpha/N$ becomes, which means that many observed coefficients $d_{jk}$ for the null hypotheses $H_0$ are eliminated, and the model goes to be oversmoothing.

Benjamini and Hochberg(1995) introduced the **False Discovery error Rate (FDR)** approach as a useful new approach to simultaneous testing after the discussion of Folks (1984). The review of Benjamini and Hochberg (1995) is as follows.

Let $R$ be the number of wavelet coefficients eliminated in the model. Of the total $R = V + S$ rejections, $S$ of them are true discoveries, and $V$ of them are false discoveries (Type-1 errors). Then, the FDR is defined as

$$\mathbf{FDR} = \mathbb{E}[\frac{V}{R}],$$

which is the expected proportion of errors among the rejected hypotheses. Starting with the null hypotheses, $H_{(1)}, H_{(2)}, \ldots, H_{(N)}$ corresponding to the ordered $p$-values, $p_{(1)}, p_{(2)}, \ldots, p_{(N)}$, we assume that p-value follow a uniform distribution ($p_i \sim \mathbb{U}(0,1)$), and are independent. Now, for any given level of $q^* \in (0,1)$, let $i_{max}$ be the largest index for which

$$p_{(i)} \leq \frac{i_{max}}{N} q^*,$$

and we reject the null hypothesis if $i \leq i_{max}$. Then, $\mathbb{E}[\mathbf{FDR}] = \frac{N_0}{N} q^*$.

Moreover, Efron (2007) suggested a local FDR to examine the inference issue in the real situation such as scientific techniques in genomics and image processing. He found that the empirical null hypotheses are extremely more dispersed than the usual theoretical ones. So, he adjust empirical Bayes techniques to a local FDR to provide both size and power calculations for large-scale studies. We review the local false discovery rate in Efron (2007) and connect this with Bayes factor shrinkage. The following produre shows FDR approach in Efron (2007).

Let $N$ null hypotheses are considered simultaneously ($H_1, H_2, \ldots, H_N$) with large number $N \geq 100$. Then the corresponding test statistics are $z_1, z_2, \ldots, z_N$, where they are not necesarily independent. If $H_i$ is significantly true, then $z_i$ would follows a standard normal distribution; $z_i \mid H_i \sim$N(0,1) $n = 1, 2, \ldots, n$,

which is called the theroetical null hypothesis in Efron (2004). These N tests are divided into two classes, null or non-null with prior probabilities $p_0$ or $p_1 = 1 - p_0$. If one test finally falls into the null case, the coefficient $d_{jk}$ would be eliminated from our interesting group. Were it otherwise, that coefficient would be kept. $z_i$ value has distribution $f_0(z)$ or $f_1(z)$, depending on its class. In other words,

$$p_0 = \mathbf{Pr}\{null \mid z\} \sim f_0(z) = \mathrm{N}(0,1) \quad \text{if } null \tag{12}$$

$$p_1 = \mathbf{Pr}\{non-null \mid z\} \sim f_1(z) \quad \text{if } non\text{-}null. \tag{13}$$

Then, we can define the **mixture density**; $f(z) = \pi_0 f_0(z) + \pi_1 f_1(z)$, and the Bayes posterior probability for the null case is,

$$\mathbf{Ifdr}(z) \equiv \mathbf{Pr}\{null \mid z\} = \frac{\pi_0 f_0(z)}{f(z)} \simeq \frac{f_0(z)}{f(z)}, \tag{14}$$

which is referred to as **the local false discovery rate**. For the large-scale testing in the practical application, $\pi_0 \geq 0.9$, since only a few coefficeints with an important information are retained. Efron(2007) dropped the factor $\pi_0$ for the sake of convenience. Therefore, $\mathbf{Ifdr}(z)$ is an upper bound on $\mathbf{Pr}\{null \mid z\}$.

### 3.2 Bayesian Adaptive Multiscale Shrinkage (BAMS)

Next, we introduce the local fdr for a bayesian shrinkage model first discussed in the wavelet domain by Vidakovic and Ruggeri (2001).

Suppose that $[d|\theta, \sigma^2] \sim N(\theta, \sigma^2)$ and the prior $\sigma^2 \sim \mathcal{E}(\mu), \quad \mu > 0$, with distribution $f(\sigma^2|\mu) = \mu e^{-\mu\sigma^2}$. The marginal likelihood distribution is,

$$d|\theta \sim \mathcal{DE}\left(\theta, \frac{1}{\sqrt{2\mu}}\right), \quad \text{with density } f(d|\theta) = \frac{1}{2}\sqrt{2\mu} e^{-\sqrt{2\mu}|d-\theta|}. \tag{15}$$

If the prior on $\theta$ suggested in Section.2 is

$$[\theta] \sim \pi_0 \delta_0 + \pi_1 \mathcal{U}(-m, m), \tag{16}$$

then the predictive distribution of $d$ is

$$[d] \sim m(d) = \pi_0 \mathcal{DE}\left(0, \frac{1}{\sqrt{2\mu}}\right) + \pi_1 m_1(d), \tag{17}$$

where

$$m_1(d) = \begin{cases} \frac{1}{2m} e^{\sqrt{2\mu}d} \sinh\sqrt{2\mu}m & \text{with} \quad d < -m, \\ \frac{1}{2m}(1 - e^{-\sqrt{2\mu}m}\cosh\sqrt{2\mu}d) & \text{with} \quad -m \leq d \leq m, \\ \frac{1}{2m} e^{-\sqrt{2\mu}d} \sinh\sqrt{2\mu}m & \text{with} \quad m < d, \end{cases} \tag{18}$$

is the marginal distribution with respect to uniform distribution $\mathcal{U}(-m, m)$. The Bayes factor in favor of $H_0 : \theta = 0$ to $H_1 : \theta \neq 0$, with the wavelet coefficient $d_{jk}$ is,

$$B_{01}(d) = \frac{m(\mathbf{d}|H_0)}{m(\mathbf{d}|H_1)} = \frac{f(d\,|0)}{\int f(d\,|\theta \neq 0)\,\pi(\theta \neq 0)\,\mathrm{d}\theta} = \frac{\frac{1}{2}\sqrt{2\mu}e^{-\sqrt{2\mu}|d|}}{m_1(d)}. \tag{19}$$

Then, we can obtain **lfdr**, estimated by an empirical Bayes factor for the empirical null, instead of the theoretical null hypothesis.

$$\mathbf{lfdr}(\mathbf{z}) = \frac{f(d|0)}{\pi_0 f(d|0) + \pi_1 m_1(d)} = \frac{B_{01}(d)}{\pi_1 + \pi_0 B_{01}(d)}. \tag{20}$$

Efron (2003, 2007) discussed the choice of the null hypothesis density with HIV infection data. Comparing the empirical- and the theoretical null, they indicated that the empirical null hypothesis would be adjustable the numerous hypothesis testing situations, because the empirical null is considerably wider than $N(0, 1)$.

### 3.3 Estimation of the Local Bayesian FDR

Rosner and Vidakovic (2000) proposed the bayesian false discovery rate (BFDR) as a way of Bayeisan methods. Here, the rejection rule is denoted in terms of posterior probabilities of the null hypotheses. After testing $n$ hypotheses $H_0$, we acquired the ordered posterior probabilities of the null hypotheses, $p_{(1)}, p_{(2)}, \ldots, p_{(n)}$, and of the alternative hypotheses, $q_{(k)} = 1 - p_{(k)},\ k = 1, \ldots, n$. Then, the maximum posterior probability for rejection is $p_k$, where

$$p_k = max\{p_j : \frac{1}{j}\sum_{i=0}^{j} i \cdot P_j(i) < \alpha\} \tag{21}$$

The probabilities $P_R(i)$ can be calculated efficiently as the coefficient with $z^i$ in the generating polynomial

$$\varphi_j(z) = \prod_{i=0}^{j}(q_{(i)} + p_{(i)}z) = \sum_{i=0}^{j} P_j(i)z^i. \tag{22}$$

Therefore, the following is the algorithm to estimate the bayesian false discovery rate:

**1.** Compute the posterior probabilities of the null, $p_{jk}$ for each test, $j = 0, \ldots, J; k = 0, \ldots, 2^j - 1$, and order them.

**2.** Fix $\alpha$ small and set $j = 1$.

**3.** Calculate $\varphi_j(z)$ using $p_{(1)}, \ldots, p_{(j)}$, as increase $j$ by 1.

**4.** Obtain $P_j(i)$. If $\frac{1}{j}\sum_{i=0}^{j} i \cdot P_j(i) \geq \alpha$, then the maximum posterior probability for the rejection is $p_{(j-1)}$. Otherwise, return to **3.**

Now, we can link **lfdr** mentioned in (20) with the Bayesian FDR, since it retains all the coefficients $d$, s.t. $p_0(d) < p_j$

Note that the bayes posterior probability $p_0(d)$ for the null hypothesis $H_0$ is:

$$\textbf{lfdr} = p_0(d) = \frac{B_{01}(d)}{\frac{\pi_1}{\pi_0} + B_{01}(d)}, \tag{23}$$

where $d$ is observed wavelet coefficient and $\frac{\pi_1}{\pi_0}$ are prior odds in favor of $H_1$.

If the hypotheses $H_0$ is rejected, by (23),

$$p_0(d) \leq \beta \quad \text{is equivalent to} \quad B_{01}(d) \leq \frac{\beta}{1-\beta} \times \frac{\pi_1}{\pi_0}.$$

# 4 Applications

## 4.1 Simulations for Benchmark Signals

We evaluated the performance of our procedure in wavelet domain using four standard test functions (Blocks, Bumps, Doppler and HeaviSine) used by Vannucci and Corradi (1999). These functions are exploited as the benchmark signals in the wavelet shrinkage because they represent patterns that are shown in several scientific fields. These signals are rescaled, and corrupted by a standard normal noise $N(0, \sigma^2)$, so that a signal-to-noise rate (SNR) was obtained. Then we compared the mean square error (MSE) with those of VisuShrink and SureShrink methods. Let $N$ denotes the number of simulational runs (we tried 1000 simulations), the MSE of the estimator can be calculated by,

$$\hat{f}_i = \frac{1}{N}\sum_{j=1}^{N}\sum_{i=0}^{n}(\hat{f}_{i,j} - f_i)^2, \text{with i = 1,. . . , n}$$

where $f_i$ are the estimated components and $\hat{f}_{i,j}$ are corresponding estimates in the $j$-th simulation run. The data was transformed into the Wavelet domain using Symmlet 8 (for Doppler, and HeaviSine), Harr 4 (for Blocks),and Daubechies 6 (for Bumps).

We focus on two scenarios which specifies the hyperparameters. In the Case 1, the parameters are specified in a global way. We obtain the scale of spread part, $\mathbf{m}$ in the prior distribution from the variance of the observations. We compare the performance of this case to that of other popular global methods (VisuShrink and SureShrink (Donoho and Johnstone, 1994; Donoho, 1995; Johnstone and Donoho, 1995).

By contrast, Case 2 shows the model parameters which are dependent on levels. The level dependent shrinkage is compared to ABWS of Chipman, et al. (1997) and BAMS of Vidakovic and Ruggeri (2001) and implemented by Antoniadis, et al. (2001). Here are more detailed description of this hyperparameter selection.

#### 4.1.1 Tuning the Model Parameters: Case 1.

1. $\mu$ is the reciprocal of the mean for the prior on $\sigma^2$, or, equivalently, the square root of the precision for $\sigma^2$. We first estimate $\sigma$ by a robust Tukey's $\mathbf{pseudos} = (Q_1 - Q_3)/C$, where $Q_1$ and $Q_3$ are the first and the third quartiles of the finest level of details in the decomposition and $1.3 \leq C \leq 1.5$. We propose $\frac{1}{\mathbf{pseudos}^2}$ as a default value for $\mu$; according to the Law of Large Numbers, this ratio should be close to the "true" $\mu$.

2. $\pi_0$ is the weight of the point mass at zero in the prior on $\theta$ and taken to be independent of level $j$.

3. $m$ is the scale of the "spread part" in the prior (16). In the case of a uniforml prior, the variance of the signal part is $m^2/3$. Because of the independence between the error and the signal parts, we have $\sigma_d^2 = (1-\pi_0)^2 m^2/3 + 1/\mu$, where $\sigma_d^2$ is the variance of the observation $d$. This yields

$$m = \sqrt{\max\left\{\frac{3(\sigma_d^2 - \frac{1}{\mu})}{(1-\pi_0)^2}, 0\right\}}.$$

   Note when $m = 0$, the prior (also the posterior) put all their mass at 0, which results in $\delta(d) = 0$.

#### 4.1.2 Tuning the Model Parameters: Case 2.

1. $\mu$ is specified as in the Case 1.

2. $\pi_0$ is the weight of the point mass at zero in the prior on $\theta$ and should depend on level $j$. Depending on our prior information about smoothness, $\pi_0$ should be close to 1 at the finest level of detail and close to 0 at the coarsest levels. We propose a hyperbolic decay in $j$,

$$\pi_0(j) = 1 - \frac{1}{(j - \mathbf{coarsest} + 1)^\gamma}, \quad \mathbf{coarsest} \leq j \leq \log_2 n,$$

   where $\mathbf{coarsest}$ is the coarsest level subjected to shrinkage.

3. Specification of $m$ coincides with that in Case 1 but with $\pi_0$ replaced by $\pi_0(j)$.

   In this case, $\sigma_d^2 = (1-\pi_0(j))^2 m^2/3 + 1/\mu$, and

$$m = \sqrt{\max\left\{\frac{3(\sigma_d^2 - \frac{1}{\mu})}{(1-\pi_0(j))^2}, 0\right\}}.$$

Table 1 illustrates the mean-square error, MSE for VisuShrink, SureShrink, BaFDR, and BLFDR-fixed with a global method case. ABWS, BAMS, and BLFDR are also shown in Table 1 with level-dependent case. Signal-to-noise ratio is 7 and sample size is 1,024. With respect to the global shrinkers in our observations, this indicates how comparable the performance of BaFDR, BLFDR-fixed is with other shrinkers. MSE, and

bias of BLFDR-fixed are significantly lower than VisuShrink, SureShrink. However, BaFDR has a slightly inferior performance than SureShrink, except for Blocks. For the level-wise srhinkers, the MSE of BLFDR is little higher than ABWS, BAMS, while Bias of BLFDR is lower than that of the others.

| | blocks | bumps |
|---|---|---|
| VisuShrink | 0.6840 (0.0719 + 0.6122) | 1.5707 (0.1165 + 1.4543) |
| SureShrink | 0.2225 (0.1369 + 0.0856) | 0.6827 (0.2660 + 0.4167) |
| BaFDR | 0.1528 (0.1067 + 0.0462) | 0.6948 (0.2954 + 0.3993) |
| BLFDR-fixed | 0.1217 (0.1087 + 0.013) | 0.3916 (0.2599 + 0.1322) |
| ABWS | 0.0995 (0.0874 + 0.0121) | 0.3495 (0.2228 + 0.1267) |
| BAMS | 0.1107 (0.0965 + 0.0142) | 0.3404 (0.1976 + 0.1428) |
| BLFDR-ld | 0.1293 (0.1245+0.0048) | 0.4049 (0.2700+0.1349) |
| | doppler | heavisine |
| VisuShrink | 0.4850 (0.0523 + 0.4327) | 0.1204 (0.0339 + 0.0864) |
| SureShrink | 0.2285 (0.0946 + 0.1340) | 0.0949 (0.0416 + 0.0534) |
| BaFDR | 0.2906 (0.2906 + 0.1867) | 0.1780 (0.0476 + 0.1304) |
| BLFDR-fixed | 0.1857 (0.1244 + 0.0613) | 0.0828 (0.0604 + 0.0224) |
| ABWS | 0.1646 (0.1006 + 0.0640) | 0.0874 (0.0442 + 0.0433) |
| BAMS | 0.1482 (0.0899 + 0.0584) | 0.0815 (0.0511 + 0.0304) |
| BLFDR-ld | 0.1931 (0.1354 + 0.0577) | 0.1001 (0.0710 + 0.0291) |

Table 1: MSE (Variance+Bias$^2$) for VisuShrink, SureShrink, BaFDR ($\alpha = 0.05$) and BLFDR (as global methods) and ABWS, BAMS, BLFDR (as level-wise methods). The standard test signals are rescaled so that the noise variance $\sigma^2$ equals 1. SNR is 7, and sample size is 1024.

We also examine the performance of global and levelwise BLFDR according to the specific value of SNR (SNR=3, 5, 7, 10). These results are presented in Table 2 and Table 3. For the specification of parameters, $\pi_0 = 0.95$ for global BLFDR, and $\gamma$ = 2 for levelwise BLFDR. In Table 2 and Table3, BLFDR has a superior performance when SNR is low (SNR = 3, 5) except for Blocks. This finding provides evidence that the uniform prior distribution which is derived from $\gamma$-minimax approach is adaptive when the data is noisy signal with low SNR. Table 4 represents the performance of the BaFDR ($\alpha$ = 0.05, $\phi_0$ = 0.9, coarsest = 5). Similar to BLFDR, BaFDR also shows higher performance when the data has a significant noise except for Blocks.

| Function | $n$ | SNR=3 | SNR=5 | SNR=7 | SNR=10 |
|---|---|---|---|---|---|
| Blocks | 512 | 0.2386 | 0.2148 | 0.2035 | 0.1830 |
| | 1024 | 0.1820 | 0.1411 | 0.1217 | 0.1059 |
| | 2048 | 0.1188 | 0.0866 | 0.0683 | 0.0558 |
| Bumps | 512 | 0.5109 | 0.5939 | 0.6785 | 0.7654 |
| | 1024 | 0.3124 | 0.3635 | 0.3916 | 0.4085 |
| | 2048 | 0.2071 | 0.2241 | 0.2311 | 0.2475 |
| Doppler | 512 | 0.2389 | 0.2522 | 0.2741 | 0.2967 |
| | 1024 | 0.1617 | 0.1674 | 0.1857 | 0.1964 |
| | 2048 | 0.1071 | 0.0997 | 0.1045 | 0.1086 |
| Heavisine | 512 | 0.1326 | 0.1304 | 0.1368 | 0.1626 |
| | 1024 | 0.0981 | 0.0819 | 0.0828 | 0.0912 |
| | 2048 | 0.0769 | 0.0582 | 0.0580 | 0.0601 |

Table 2: Performance of Local False Discovery Rate in Wavelet Domain. The table shows average MSE for 1000 simulations, with parameters $\tau$ and $\pi_0 = 0.95$ fixed for all levels.

| Function | $n$ | SNR=3 | SNR=5 | SNR=7 | SNR=10 |
|---|---|---|---|---|---|
| Blocks | 512 | 0.2846 | 0.2620 | 0.2307 | 0.1936 |
| | 1024 | 0.2015 | 0.1559 | 0.1293 | 0.1112 |
| | 2048 | 0.1091 | 0.0904 | 0.0796 | 0.0665 |
| Bumps | 512 | 0.4978 | 0.5791 | 0.6521 | 0.7175 |
| | 1024 | 0.3366 | 0.3836 | 0.4049 | 0.4110 |
| | 2048 | 0.2003 | 0.2314 | 0.2486 | 0.2726 |
| Doppler | 512 | 0.2740 | 0.3027 | 0.3209 | 0.3340 |
| | 1024 | 0.1604 | 0.1775 | 0.1931 | 0.2061 |
| | 2048 | 0.0843 | 0.0965 | 0.1125 | 0.1247 |
| Heavisine | 512 | 0.1727 | 0.1663 | 0.1779 | 0.1960 |
| | 1024 | 0.0925 | 0.1663 | 0.1001 | 0.1170 |
| | 2048 | 0.0538 | 0.0521 | 0.0587 | 0.0698 |

Table 3: Performance of Local False Discovery Rate in Wavelet Domain. The table shows average MSE for 1000 simulations, with level-dependent parameters $m$ and $\pi_0$, $\gamma = 2$.

| Function | $n$ | SNR=3 | SNR=5 | SNR=7 | SNR=10 |
|---|---|---|---|---|---|
| Blocks | 512 | 0.4456 | 0.3770 | 0.2912 | 0.2085 |
| | 1024 | 0.3108 | 0.2429 | 0.1528 | 0.1117 |
| | 2048 | 0.1523 | 0.1107 | 0.0843 | 0.0520 |
| Bumps | 512 | 1.2402 | 1.1828 | 1.1862 | 1.3039 |
| | 1024 | 0.5755 | 0.6607 | 0.6948 | 0.6814 |
| | 2048 | 0.3227 | 0.3671 | 0.3756 | 0.3558 |
| Doppler | 512 | 0.3976 | 0.4156 | 0.4479 | 0.4593 |
| | 1024 | 0.1952 | 0.2519 | 0.2906 | 0.3372 |
| | 2048 | 0.0884 | 0.1067 | 0.1286 | 0.1569 |
| Heavisine | 512 | 0.0962 | 0.1983 | 0.3278 | 0.4453 |
| | 1024 | 0.0758 | 0.1507 | 0.1780 | 0.2329 |
| | 2048 | 0.0547 | 0.0537 | 0.0616 | 0.0867 |

Table 4: Performance of the BaFDR. The average MSE for 1,000 simulations with $\alpha = 0.05$ and $\pi_0 = 0.90$ coarsest=5 for all

## 4.2 An example in AFM spectral data

To illustrate features of the LBFDR shrinkage method, we used 1024 measurements in atomic force microscopy (**AFM**). The **AFM** is a type of scanned proximity probe microscopy (SPM) that can measure the adhesion strength between two materials at the nanonewton scale (Binnig et al., 1986). In AFM, a cantilever beam is adjusted until it bonds with the surface of a sample, and then the force required to separate the beam and sample is measured from the beam deflection. Beam vibration can be caused by factors such as thermal energy of the surrounding air or the footsteps of someone outside the laboratory. The vibration of a beam acts as noise on the deflection signal; in order for the data to be useful this noise must be removed. (Gabriel, 2005) The AFM data from the adhesion measurements between carbohydrate and the cell adhesion molecule (CAM) E-Selectin was collected by Bryan Marshal from the BME Department at Georgia Institute of Technology. The technical description is provided in Marshall et al. (2001). Figure 1 illustrates AFM signal smoothed by BaFDR, BLFDR when $\phi_0$ = 0.999 fixed for all levels, and BLFDR with level-dependent when $\phi_0$ but fixed $\gamma = 5$. As Figure 1 indicates, three shrinkers show considerable a smooth behavior, especially in the long-middle part from about 500 to 1900. In addition, The figure exhibits a clear trend that a signal goes to be smoother from BaFDR to levelwise BLFDR. This result reveals the fact that BLFDR with level dependent $\phi_0$ is the most adaptive for noisy signals, which is consistent with Vidakovic and Ruggeri (2001).

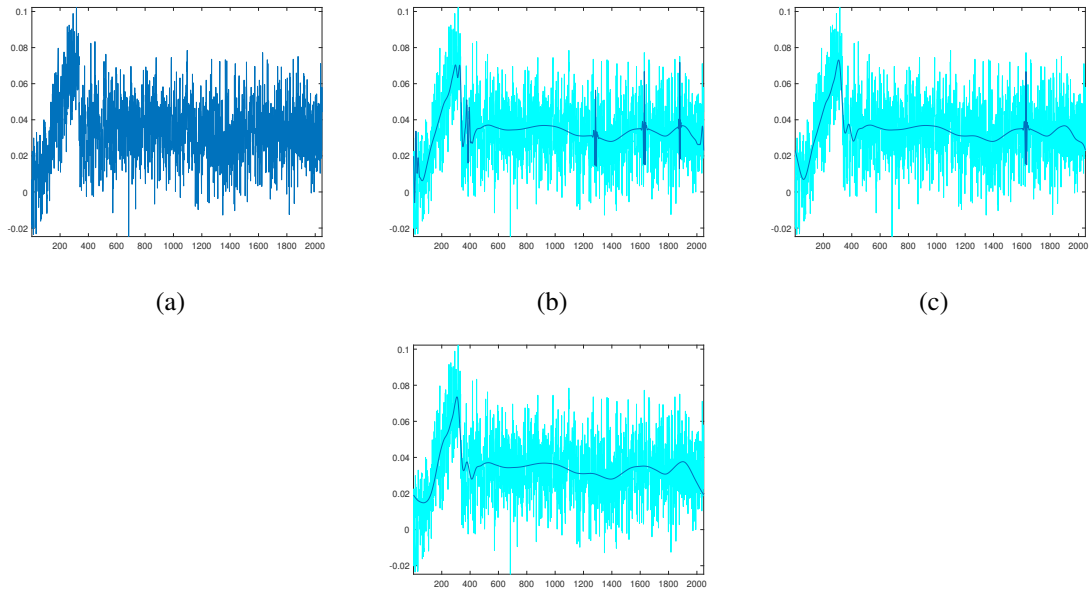


Figure 1: (a) Original AFM signal; (b) Smoothing with BaFDR; (c) Smoothing with BLFDR with $\pi_0 = 0.999$ fixed for all levels; and (d) Smoothing with BLFDR with level-dependent $\pi_0$ but fixed $\gamma = 5$.

# 5 REFERENCES


Abramovich, F. and Benjamini, Y. (1995). Thresholding of wavelet coefficients as a multiple hypotheses testing procedure, In: *Wavelets and Statistics,* Editors A. Antoniadis and G. Oppenheim. Lecture Notes in Statistics, **103**, 5–14. Springer-Verlag, New York.

Abramovich, F. and Benjamini, Y. (1996). Adaptive thresholding of wavelet coefficients *Computational Statistics and Data Analysis*,**22** 351–361.

Angelini, C. and Vidakovic, B. (2003). Some Novel Methods in Wavelet Data Analysis: Wavelet Anova, F-test Shrinkage, and Γ-Minimax Wavelet Shrinkage, In: *Wavelets and their Applications*, Editors Krishna, M., Radha, R., and Thangavely, S. Allied Publishers Ltd, New Delhi, ISBN 81-7764-493-9, pp. 31–45.

Antoniadis, A., Bigot, J. and Sapatinas, T. (2001). Wavelet estimators in nonparametric regression: a comparative simulation study. *Journal of Statistical Software,* **6**, 1–83.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J.R. Stat. Soc. Ser. B Stat. Methodol. 57 289–300.

Binnig, G., Quate, C.F., and Gerber, Ch. (1986). Atomic force microscope. *Phys. Rev. Lett.* **56**, 930–933.

Chipman, H., Kolaczyk, E., and McCulloch, R. (1997). Adaptive Bayesian Wavelet Shrinkage, *Journal of the American Statistical Association*,**92**, 1413–1421.

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

Donoho, D. L. (1995). De-noising by soft thresholding. *IEEE Transactions on Information Theory*,**41**, 613–627.

Efron, B. (2004). Large-Scale Simultaneous Hypothesis Testing: The Choice of a Null Hypothesis *Journal of the American Statistical Association,* **99**, 96–104.

Efron, B. and Tibshirani, R. (2002). Empirical Bayes methods and false discovery rates for microarrays. Genetic Epidemiology 23 70–86.

Efron, B., Tibshirani, R., Storey, J. and Tusher, V. (2001). Empirical Bayes analysis of a microarray experiment. J. Amer. Statist. Assoc. 96 1151–1160.

Folks, J. L. (1984). Combination of Independent Tests, In *Handbook of Statistics, Vol 4.*, P. R. Krishniah and P. K. Sen Editors, Elsevier Scince Publishers, 113–121.

Johnstone, I. M. and Donoho, D. L. (1995). Adapting to smoothness via wavelet shrinkage. *Journal of the Statistical Association*, **90**, 1200–1224.

Marshall, B., McEver, R. and Zhu, C. (2001). Kinetic rates and their force dependence on the P-Selectin/PSGL-1 interaction measured by atomic force microscopy. Proceedings of ASME 2001, Bioengineering Conference, BED - Vol. 50.

Ogden, T. and Parzen, E. (1996). Data dependent wavelet thresholding in nonparametric regression with change-point applications. *Computational Statistics and Data Analysis*, **22**, 53–70.

Rosner, G. and Vidakovic, B. (2000). Wavelet Functional ANOVA, Bayesian False Discovery Rate, and Longitudinal Measurements of Oxygen Pressure in Rats, Technical Report 1/2000, ISyE, Georgia Institute of Technology.

Vidakovic, B. and Ruggeri, F. (2001). BAMS Method: Theory and Simulations. *Sankhyā, Series B,* **63***,2* (Special Issue on Wavelets), 234–249.