# Wavelet Estimation of a Baseline Signal From Repeated Noisy Measurements by Vertical Block Shrinkage

Woojin Chang[1]
and
Brani Vidakovic[2]

**Abstract.** In this paper a new wavelet shrinkage technique is proposed and investigated. When data consist of a multiplicity of related noisy signals, we propose a wavelet-based shrinkage estimation procedure to summarize all data components into a single regularized and representative signal ("base-line"). This fusion of information from different runs is done via Stein-type shrinkage rule resulting from an empirical Bayes argument. The proposed shrinkage estimators maximize the predictive density under appropriate model assumptions on the wavelet coefficients. Features of this model-induced shrinkage are that it is "block-vertical" and local in time.

The method, called VERTISHRINK, is evaluated on a battery of test signals under various signal-to-noise ratios and various number of vector components. An application in estimating the base-line signal in an experiment in tumor physiology is provided as well.

**KEY WORDS:** Empirical Bayes Method, Stein-type Estimators, Wavelet Shrinkage

## 1 Introduction

The wavelet shrinkage paradigm has been proven to be an efficient and surprisingly simple data analytic tool introduced in statistics by work of Donoho, Johnstone, and their coauthors in the early 1990's. The wavelet shrinkage usually refers to application of shrinkage rules in the wavelet domain, i.e., on the wavelet-transformed observations. Thresholding rules are simple and their calculation has been done from many different data-analytic and statistical points of view, such as asymptotic minimax, exact-risk, penalized mean

[1]Woojin Chang is an Ph.D candidate, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 30332-0265,

[2]Brani Vidakovic is an Associate Professor, School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia, 30332-0265.

square, minimum-description length, Bayes, and empirical Bayes (Bruce and Gao, 1996; Donoho *et al.*, 1995; Leporini and Pesquet, 1998; Saito, 1994; Vidakovic, 1998). Most of applications of wavelet shrinkage deal with a single run of a time series expressed in a regression fashion: data = signal + noise.

Examples of multivariate signals measured by several acquisition channels are ubiquitous in industrial monitoring, geophysics, chemometrics, biometry, etc. It is often of interest to produce an estimator of a "base-line" signal from multiplicity of related noisy measurements. A "solution" that will average individually smoothed signals will not use all available information at the regularization stage. This paper proposes a method, based on statistical model in the wavelet domain to exhibit the common base-line signal by averaging smoothed signals regularized by incorporating all available information.

The original Stein-type phenomenon relates to improvement in the risk of an estimator of a multivariate normal mean by shrinking the standard estimator $\bar{X}$ towards 0 (see Stein, 1981; Gruber, 1998). Stein-type shrinkage was proposed and used in the wavelet setup by several authors (Cai and Silverman 2001; Hall et al., 1999). The novel idea of this paper is that the shrinkage is applied on multiple simultaneous measurements all at a fixed multiresolution index in the wavelet domain. Therefore we define "block-type" thresholding where the blocks are "vertical", i.e., extend over the multivariate observations at a fixed position in the wavelet domain. This thresholding is local, while combining information among the components in a multivariate signal. We arrive to a thresholding rule in a form of non-negative part of a Stein estimator by an empirical Bayes argument. Our estimator maximizes the predictive distribution under appropriate modeling assumptions.

The paper is organized as follows. The Chapter 2 describes the model leading to the thresholding estimate. We consider the models with various prior covariance matrices as well as suitable modifications of exhibited shrinkage rules. The extensive simulational results and a real-life application are presented in Chapter 3. Discussion and conclusions are given in Chapter 4.

# 2 The Model

The wavelet shrinkage is often related to statistical modeling in the wavelet domain. Traditionally, the models in the domain of wavelet coefficients are of location type in which estimating the location corresponds to estimating the signal part. The estimators are shrinkage and their wavelet-inversion yields the estimator of unknown signal in the domain of measured data.

In what follows, we describe the statistical model that underlines our derivations. We assume that the observed data $\mathbf{y}_i = (y_i^1, \ldots, y_i^M)'$ represent the sum of $M$ unknown signals $\mathbf{f}_i = (f_i^1, \ldots, f_i^M)'$ and random noises $\boldsymbol{\eta}_i = (\eta_i^1, \ldots, \eta_i^M)'$. Thus, the data can be thought as a matrix $\mathbf{Y} = (\mathbf{y}_1 | \mathbf{y}_2 | \ldots | \mathbf{y}_n)$ of size $M \times n$. Column-wise,

$$\mathbf{y}_i = \mathbf{f}_i + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \ \mathcal{MVN}_M(0, \sigma^2 \mathbf{I}_{M\times M}), \ i = 1, \ldots, n \tag{1}$$

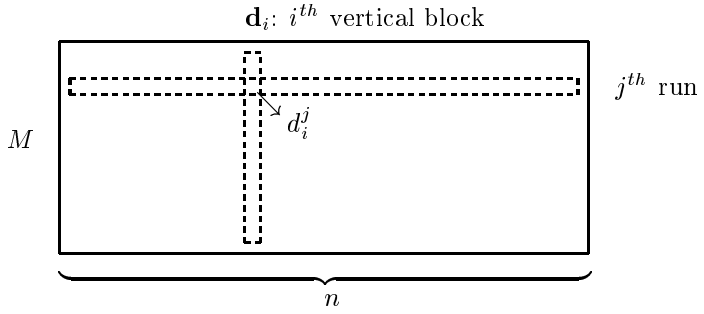


Figure 1: Structure of matrix $\mathbf{D}$

An orthogonal wavelet transformation $\mathbf{W}$ transforms the observed data $\mathbf{Y}$ to

$$\mathbf{D} = \mathbf{YW}, \tag{2}$$

where $\mathbf{D} = (\mathbf{d}_1|\mathbf{d}_2|\ldots|\mathbf{d}_n)$ is a matrix of size equal to that of $\mathbf{Y}$, and $\mathbf{W}$ is the wavelet matrix of size $n \times n$. The structure of matrix $\mathbf{D}$ is shown in Figure 1 where columns $\mathbf{d}_i$ correspond to vertical blocks. Vertical blocks are coefficients from different signals at the same location in their wavelet decompositions.

From this point on, we drop the index $i$ in $\mathbf{d}_i$ for the simplicity of notation. According to the model (1), the likelihood for the column vector $\mathbf{d}$ in the wavelet domain is

$$\mathbf{d}|\boldsymbol{\theta}, \sigma \sim \mathcal{MVN}(\boldsymbol{\theta}, \sigma^2\mathbf{I}), \tag{3}$$

with the density

$$f(\mathbf{d}|\boldsymbol{\theta}, \sigma^2) = (2\pi\sigma^2)^{-M/2}\exp\{-(\mathbf{d}-\boldsymbol{\theta})'(\mathbf{d}-\boldsymbol{\theta})/(2\sigma^2)\}. \tag{4}$$

Note that unlike traditional block-modeling where the blocks involve coefficients in a single run, the proposed model involves wavelet coefficients from multiple runs, but at the same location point (Figure 1).

To complete the model we elicit the prior on $[\boldsymbol{\theta}, \sigma^2]$. A proposal utilized by several researchers (DeCanditiis, 2001; Vannucci and Coradi, 1997; Vidakovic and Müller, 1995) involves the prior from the Normal-Inverse Gamma family. The density of the prior $\pi(\boldsymbol{\theta}, \sigma^2|a, b, \mathbf{m}, \boldsymbol{\Sigma})$ is (O'Hagan, 1994),

$$\begin{aligned}\pi(\boldsymbol{\theta}, \sigma^2|a, b, \mathbf{m}, \boldsymbol{\Sigma}) &= \frac{(a/2)^{b/2}}{(2\pi)^{M/2}|\boldsymbol{\Sigma}|^{1/2}\Gamma(b/2)}(\sigma^2)^{-(b+M+2)/2} \\ &\quad \times\exp\{-\big((\boldsymbol{\theta}-\mathbf{m})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}-\mathbf{m})+a\big)/(2\sigma^2)\},\end{aligned} \tag{5}$$

and is denoted by $\mathcal{NIG}(a, b, \mathbf{m}, \boldsymbol{\Sigma})$.
This selection is motivated by the conjugate property of $\mathcal{NIG}$ distribution, i.e. under the model (4), the posterior is also $\mathcal{NIG}$. Moreover, by selecting $\boldsymbol{\Sigma}$ different than $\mathbf{I}$, one can model dependence between wavelet coefficients. It is

interesting that the conditional prior distribution of $\sigma^2|\boldsymbol{\theta}$ is the inverse gamma distribution with hyperparameters $a$ and $b$ (O'Hagan, 1994),

$$f(\sigma^2) = \frac{(a/2)^{b/2}}{\Gamma(b/2)}(\sigma^2)^{-(b+2)/2}\exp\{-a/(2\sigma^2)\}. \tag{6}$$

while the conditional prior distribution for $\boldsymbol{\theta}|\sigma^2$ is normal. Thus the name Normal-Inverse Gamma is justified. However, the variables $\boldsymbol{\theta}$ and $\sigma^2$ under the model (5) are not independent.

Now suppose that $\mathcal{NIG}(a, b, 0, \boldsymbol{\Sigma})$ is adopted as the prior distribution for $[\boldsymbol{\theta}, \sigma^2]$. Selection of $\mathbf{m} = 0$ reflects the propensity of a wavelet coefficients to center about 0, as it is traditionally modeled. Because of conjugacy the resulting posterior distribution is

$$\begin{aligned}
\pi(\boldsymbol{\theta}, \sigma^2|a^*, b^*, \mathbf{m}^*, \boldsymbol{\Sigma}^*) &= \frac{[a^*/2]^{\frac{b^*}{2}}}{(2\pi)^{\frac{M}{2}}|\boldsymbol{\Sigma}^*|^{\frac{1}{2}}\Gamma(\frac{b^*}{2})}(\sigma^2)^{\frac{-(b^*+M+2)}{2}} \\
&\quad \times \exp\Big\{\frac{-((\boldsymbol{\theta}-\mathbf{m}^*)'(\boldsymbol{\Sigma}^*)^{-1}(\boldsymbol{\theta}-\mathbf{m}^*)+a^*}{2\sigma^2}\Big\} \\
&= \frac{[(a+\mathbf{d}'(\boldsymbol{\Sigma}+\mathbf{I})^{-1}\mathbf{d})/2]^{\frac{b+M}{2}}}{(2\pi)^{\frac{M}{2}}|\boldsymbol{\Sigma}^{-1}+\mathbf{I}|^{\frac{-1}{2}}\Gamma(\frac{b+M}{2})}(\sigma^2)^{\frac{-(b+2M+2)}{2}} \\
&\quad \times \exp\Big\{\frac{-(\boldsymbol{\theta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}+(\mathbf{d}-\boldsymbol{\theta})'(\mathbf{d}-\boldsymbol{\theta})+a)}{2\sigma^2}\Big\},
\end{aligned} \tag{7}$$

with

$$\begin{aligned}
a^* &= a + \mathbf{d}'\mathbf{d} - (\mathbf{m}^*)'(\boldsymbol{\Sigma}^*)^{-1}\mathbf{m}^*, \\
b^* &= b + M, \\
\mathbf{m}^* &= (\boldsymbol{\Sigma}^{-1}+\mathbf{I})^{-1}\mathbf{d}, \text{ and} \\
\boldsymbol{\Sigma}^* &= (\boldsymbol{\Sigma}^{-1}+\mathbf{I})^{-1}.
\end{aligned} \tag{8}$$

The Bayes rule for estimating $\boldsymbol{\theta}$ is the mean of posterior, and in our case will depend on $\boldsymbol{\Sigma}$. The Empirical Bayes paradigm is concerned in specifying hyper-parameters $a, b$, and $\boldsymbol{\Sigma}$ in a data-dependent fashion. This is usually done via maximizing the marginal (predictive) distribution (or their product) with respect to hyper-parameters and is known as ML-II principle, since it resembles the maximum likelihood.

By Bayes formula, the predictive density can be expressed by

$$p(\mathbf{d}|a, b, \boldsymbol{\Sigma}) = \frac{f(\mathbf{d}|\boldsymbol{\theta}, \sigma^2)\pi(\boldsymbol{\theta}, \sigma^2|a, b, 0, \boldsymbol{\Sigma})}{\pi(\boldsymbol{\theta}, \sigma^2|a^*, b^*, \mathbf{m}^*, \boldsymbol{\Sigma}^*)},$$

where $a^*, b^*, \mathbf{m}^*$, and $\boldsymbol{\Sigma}^*$ are given by (8). This yields a multivariate $t$ distribution,

$$p(\mathbf{d}|a, b, \boldsymbol{\Sigma}) = \frac{a^{b/2}\Gamma\left(\frac{b+M}{2}\right)}{\pi^{M/2}|\boldsymbol{\Sigma}+\mathbf{I}|^{1/2}\Gamma(b/2)[a+\mathbf{d}'(\boldsymbol{\Sigma}+\mathbf{I})^{-1}\mathbf{d}]^{\frac{b+M}{2}}}\ , \tag{9}$$

or simpler,

$$p(\mathbf{d}|a,b,\mathbf{\Sigma}) = \text{const.} \times |\mathbf{I}+\mathbf{\Sigma}|^{-1/2}(a+\mathbf{d}'(\mathbf{I}+\mathbf{\Sigma})^{-1}\mathbf{d})^{-\frac{b+M}{2}}. \tag{10}$$

We are concerned in maximizing (10) with respect to costrained $\mathbf{\Sigma}$, by keeping $a$ and $b$ fixed. The hyperparameters $a$ and $b$ will be determined from data by moment matching procedure. The constraint is obtained from the requirement that the Bayes estimator remains in a Stein-type shrinkage form. Since Bayes estimator is $\hat{\boldsymbol{\theta}} = [\mathbf{I}-(\mathbf{I}+\mathbf{\Sigma})^{-1}]\mathbf{d}$, we consider the equation

$$\hat{\boldsymbol{\theta}} = [\mathbf{I}-(\mathbf{I}+\mathbf{\Sigma})^{-1}]\mathbf{d} = \Big(1-\frac{C}{\mathbf{d}'\mathbf{d}}\Big)\mathbf{d} \tag{11}$$

and determine $C$ by an Empirical Bayes argument. A sufficient condition for (11) is

$$(\mathbf{I}+\mathbf{\Sigma})^{-1} = \frac{C}{\mathbf{d}'\mathbf{d}}\mathbf{I} \tag{12}$$

and a solution for $\mathbf{\Sigma}$ is

$$\mathbf{\Sigma} = \Big(\frac{\mathbf{d}'\mathbf{d}}{C}-1\Big)\mathbf{I}, \tag{13}$$

where $0 < C < \mathbf{d}'\mathbf{d}$, because $\mathbf{\Sigma}$ is positive definite. Then (10) can be rewritten as

$$p(\mathbf{d}|a,b,\mathbf{\Sigma}) = \text{const.} \times \Big(\frac{C}{\mathbf{d}'\mathbf{d}}\Big)^{M/2}(a+C)^{-\frac{b+M}{2}}.$$

Denote

$$f(C) = \Big(\frac{C}{\mathbf{d}'\mathbf{d}}\Big)^{M/2}(a+C)^{-\frac{b+M}{2}} \qquad 0 < C < \mathbf{d}'\mathbf{d}. \tag{14}$$

By taking the logarithm of $f(C)$,

$$\ln f = \frac{M}{2}\ln C - \frac{M}{2}\ln\,\mathbf{d}'\mathbf{d} - \frac{b+M}{2}\ln(a+C) \tag{15}$$

and differentiating with respect to $C$, we obtain the solution

$$\frac{\partial \ln f}{\partial C} = \frac{1}{2}\Big(\frac{M}{C}-\frac{b+M}{a+C}\Big) = 0 \Leftrightarrow \quad C = \frac{aM}{b}. \tag{16}$$

When $C = C^* = \min(aM/b, \mathbf{d}'\mathbf{d})$, (14) has a maximum value, since

$$\frac{\partial^2 \ln f}{\partial C^2}\Big|_{C=a\,M/b} = \frac{-b^3}{2a^2M\,(b+M)} < 0.$$

By substituting such $C^*$ in (11), we obtain the thresholding rule

$$\hat{\boldsymbol{\theta}} = \Big(1-\frac{aM/b}{\mathbf{d}'\mathbf{d}}\Big)_+\mathbf{d}\ , \tag{17}$$

and two hyparameters to be specified are $a$ and $b$. Since $\sigma^2 \sim\ \mathcal{IG}(a,b)$ and it is always unimodal with mode at $a/(b+2)$ and the mean $E[\sigma^2] = a/(b-2),\ \ b > 2$

(O'Hagan, 1994), we can "replace" $a/b$ with an estimator of $\sigma^2$. The estimator for $\sigma^2$ can be obtained from the coefficients corresponding to the finest level of detail in the matrix $\mathbf{D}$. In homoscedastic setups there is $M \times n/2$ coefficients available to estimate the variance. The argument for selecting finest levels of detail is quite standard in the wavelet shrinkage theory, see Vidakovic (1999) for discussion and overview of some commonly used estimators.

Then (17) becomes

$$\hat{\boldsymbol{\theta}} = \left(1 - \frac{M\hat{\sigma}^2}{\mathbf{d}'\mathbf{d}}\right)_+ \mathbf{d}. \tag{18}$$

A reasonable strategy to estimate a base-line signal is to average wavelet coefficients at the same position. Because of linearity of wavelet transformation, such an averaging corresponds to averaging the signals in the time domain. We propose a method that averages already shrunk coefficients thus incorporating a regularization.

Thus, a modification of (18), suitable for estimating a base-line signal is

$$\hat{\theta}_b = \left(1 - \frac{M\hat{\sigma}^2}{\mathbf{d}'\mathbf{d}}\right)_+ \overline{\mathbf{d}}, \tag{19}$$

where $\overline{\mathbf{d}}$ is the average of components in $\mathbf{d}$ and $\hat{\theta}_b$ stands for the base-line wavelet coefficient. Of course, time-domain baselines are obtained by wavelet-inversion.

In Application section estimator $\hat{\theta}_b$ as in (19) was used and evaluated.

**Remark.** A simpler model in which the signal part is considered common can be obtained from the discussed model under degenerated covariance structure in which $\boldsymbol{\Sigma}$ is proportional to the matrix consisting exclusively of ones. In such simple model, the shrinkage rule is

$$\hat{\theta}'_b = \left(1 - \frac{\hat{\sigma}^2}{M\bar{\mathbf{d}}^2}\right)_+ \bar{\mathbf{d}}, \tag{20}$$

instead of rule (19). That means that in the simple rule, expression $\frac{\mathbf{d}'d}{M} = \frac{(d^1)^2 + \cdots + (d^M)^2}{M}$ from (19) is replaced by $\frac{(d^1 + \cdots + d^M)^2}{M}$ leading to a different type of shrinkage.

# 3 Applications

In this section we provide an extensive simulational study involving the battery of standard univariate signals. Such studies are important since the "truth" is known and the method can be evaluated and compared to other procedures. In addition, we provide an application in determining the base-line response in the case of multiple, related, and noisy physiological measurements. First, we assess the MSE performance of our threshold estimator on the standard signals (**blocks, bumps, doppler** and **heavisine**). To produce multiple runs, we added an i.i.d. normal noise with unit variance to the re-scaled signals, so

that the signal to noise ratio (SNR) is 3,5,7 and 10. The decomposing wavelets are: Haar for **blocks**, Daubechies 6 for **bumps** and Symmlet 8 for **doppler** and **heavisine**, as standardly selected in previous studies on wavelet denoising. Lengths of signals are $n = 256, 512, 1024, 2048$ and 4096. The base-line is estimated on basis of 20 noisy signals, i.e., $M = 20$. For $r^{th}$ $(r = 1, \ldots, R)$ simulational run, the MSE$(r)$ of the estimator $(\hat{f}_1^r, \ldots, \hat{f}_n^r)$ is calculated as

$$\text{MSE}(r) = \frac{1}{n} \sum_{i=1}^{n} (\hat{f}_i^r - f_i)^2,$$

where $f_i$'s are discretized components of the true, unknown signal and $\hat{f}_i^r$'s are corresponding estimators in $r^{th}$ simulation run. The averaged mean square error, AMSE of the estimators $(\hat{f}_1^r, \ldots, \hat{f}_n^r)$ $r = 1, \ldots, R$ is calculated as,

$$\text{AMSE} = \frac{1}{R} \sum_{r=1}^{R} \text{MSE}(r),$$

where $R$ is the number of replications.

| FUNCTION | $n$ | SNR=3 | SNR=5 | SNR=7 | SNR=10 |
|---|---|---|---|---|---|
| BLOCKS | 256 | 0.0201 | 0.0142 | 0.0131 | 0.0112 |
| | 512 | 0.0141 | 0.0116 | 0.0099 | 0.0085 |
| | 1024 | 0.0095 | 0.0071 | 0.0064 | 0.0056 |
| | 2048 | 0.0059 | 0.0050 | 0.0043 | 0.0040 |
| | 4096 | 0.0037 | 0.0033 | 0.0032 | 0.0030 |
| BUMPS | 256 | 0.0676 | 0.0696 | 0.0670 | 0.0633 |
| | 512 | 0.0511 | 0.0489 | 0.0475 | 0.0463 |
| | 1024 | 0.0311 | 0.0306 | 0.0307 | 0.0312 |
| | 2048 | 0.0189 | 0.0185 | 0.0188 | 0.0193 |
| | 4096 | 0.0111 | 0.0112 | 0.0115 | 0.0121 |
| DOPPLER | 256 | 0.0422 | 0.0409 | 0.0388 | 0.0382 |
| | 512 | 0.0210 | 0.0255 | 0.0230 | 0.0239 |
| | 1024 | 0.0147 | 0.0156 | 0.0153 | 0.0149 |
| | 2048 | 0.0088 | 0.0094 | 0.0093 | 0.0095 |
| | 4096 | 0.0054 | 0.0054 | 0.0054 | 0.0055 |
| HEAVISINE | 256 | 0.0235 | 0.0259 | 0.0279 | 0.0307 |
| | 512 | 0.0158 | 0.0191 | 0.0207 | 0.0204 |
| | 1024 | 0.0103 | 0.0117 | 0.0124 | 0.0127 |
| | 2048 | 0.0072 | 0.0078 | 0.0079 | 0.0083 |
| | 4096 | 0.0045 | 0.0049 | 0.0050 | 0.0051 |

Table 1: AMSE ($M = 20$, $R = 50$) for **blocks, bumps, doppler** and **heavisine**: a variety of sample sizes and SNR's.

We show a denoising example using (19) in Figure .

The simulation results using (19) are summarized in Table 1. Figure 2 shows a noisy and denoised **doppler** signal from one simulational run, for the choice: $n = 1024$, SNR=7, and $M = 20$.

To evaluate our method, the following simulational comparison was made. We averaged regularized coefficients where the regularization was performed independently signal-by-signal, using the universal hard- and soft-thresholding. We found that averaging of such independently shrunk signals was AMSE under-performing compared to our proposed model. Results for universal hard-thresholding are summarized in Table 2.

We also compared rules (19) and (20). For $n$ small-to-moderate the rules are comparable. However, when $n$ is large, say larger than 1024, the rule in (19) outperforms (20).

| FUNCTION | $n$ | SNR=3 | SNR=5 | SNR=7 | SNR=10 |
|---|---|---|---|---|---|
| BLOCKS | 256 | 0.1507 | 0.0603 | 0.0212 | 0.0180 |
| | 512 | 0.1012 | 0.0666 | 0.0469 | 0.0303 |
| | 1024 | 0.0840 | 0.0430 | 0.0290 | 0.0167 |
| | 2048 | 0.0471 | 0.0282 | 0.0143 | 0.0068 |
| | 4096 | 0.0246 | 0.0101 | 0.0052 | 0.0044 |
| BUMPS | 256 | 0.3568 | 0.3287 | 0.3375 | 0.3615 |
| | 512 | 0.3016 | 0.3021 | 0.2978 | 0.2562 |
| | 1024 | 0.2237 | 0.2139 | 0.1856 | 0.1827 |
| | 2048 | 0.1465 | 0.1377 | 0.1273 | 0.1190 |
| | 4096 | 0.0800 | 0.0799 | 0.0500 | 0.0499 |
| DOPPLER | 256 | 0.2363 | 0.2060 | 0.2145 | 0.2143 |
| | 512 | 0.1181 | 0.1200 | 0.1209 | 0.1161 |
| | 1024 | 0.0852 | 0.1003 | 0.0984 | 0.0946 |
| | 2048 | 0.0499 | 0.0530 | 0.0585 | 0.0595 |
| | 4096 | 0.0271 | 0.0311 | 0.0296 | 0.0277 |
| HEAVISINE | 256 | 0.0798 | 0.1311 | 0.1460 | 0.1428 |
| | 512 | 0.0635 | 0.0907 | 0.1133 | 0.1339 |
| | 1024 | 0.0422 | 0.0590 | 0.0761 | 0.0821 |
| | 2048 | 0.0258 | 0.0444 | 0.0551 | 0.0582 |
| | 4096 | 0.0159 | 0.0222 | 0.0267 | 0.0286 |

Table 2: AMSE ( $M = 20$, $R = 50$, $\hat{\theta}_b = \frac{1}{M}\sum_{j=1}^{M} d^j 1(|d^j| > \sqrt{2\log n})$ ) for **blocks, bumps, doppler** and **heavisine** by averaging method: a variety of sample sizes and SNR's.

Next, we provide an example of estimating base-line response in an experiment in tumor physiology performed in Duke Medical Center. Here is a brief description.

Experiments carried out *in vitro* with tumor cell lines have demonstrated that tumor cells respond to radiation and anti-cancer drugs differently, depending on the environment. In particular, available oxygen is important. Efforts

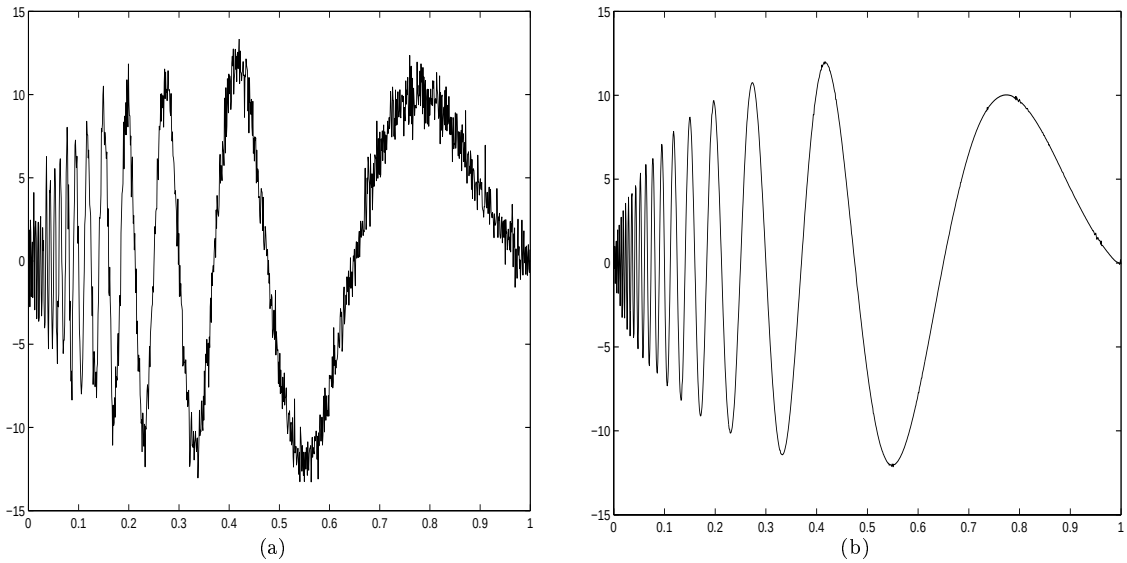


Figure 2: (a) A noisy **doppler** signal [SNR=7, $n = 1024$, noise variance $\sigma^2 = 1$, $M = 20$] (b) Reconstructed Signal

to increase the level of oxygen within tumor cells have included laboratory rats with implanted tumors breathing pure oxygen. Unfortunately, animals breathing pure oxygen may experience large drops in blood pressure, enough to make this intervention risky for clinical use (Dewhirst, Lanzen, and Braun, 1998).

Researchers sought to evaluate carbogen (95% pure oxygen and 5% carbon dioxide) as a breathing mixture that might improve tumor oxygenation without causing a drop in blood pressure. The protocol called for making measurements on each animal over 20 minutes of breathing room air, followed by 40 minutes of carbogen breathing. The responses are longitudinal measurements of oxygen partial pressure ($PO_2$). Microelectrodes, inserted into the tumors (one per animal) measured $PO_2$ at a particular location within the tumor throughout the study period. Each animal wore a face mask for administration of breathing gases (room air or carbogen). (See Lanzen, Braun, Ong, and Dewhirst, 1998, for more information about these experiments.)

Figure 3 show $M = 9$ $PO_2$ measurements. The plots show several features, including an obvious rise in $PO_2$ at the 20-minute mark among the animals. No physiologic model exists that would characterize the shapes of these profiles mathematically. The primary study question concerned evaluating the effect of carbogen breathing on $PO_2$. The analysis is made complicated by the knowledge that there may be acute change in $PO_2$ after carbogen breathing starts.

The analyses concern inference on change in some physiologic measurements after the intervention. The problem for the data analysis is how best to define

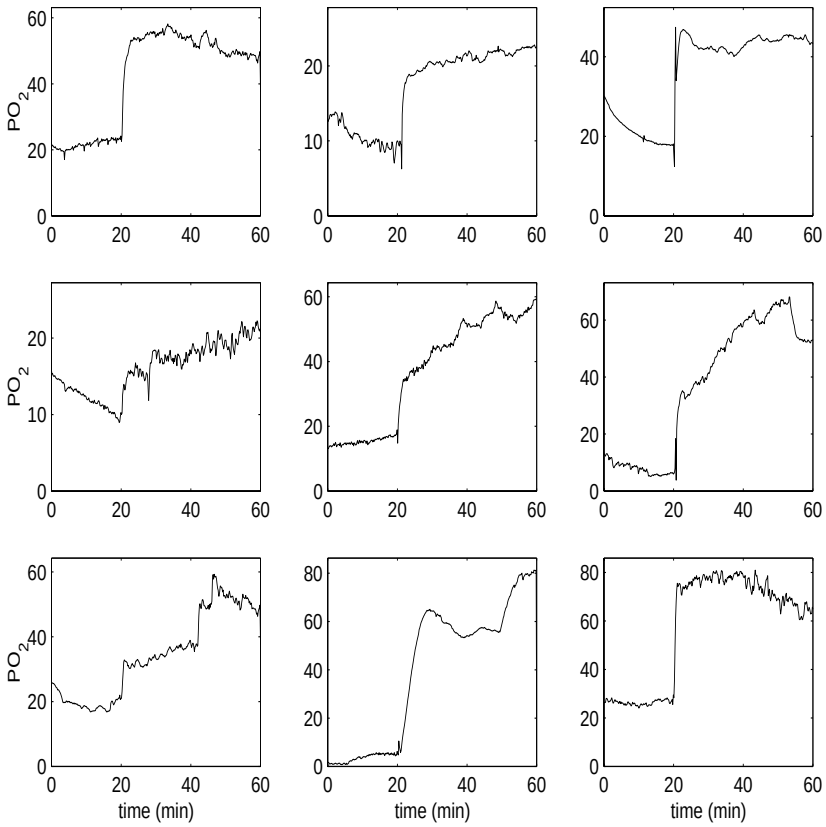


Figure 3: $PO_2$ measurements. Notice that despite a variety of functional responses and a lack of a simple parametric model, at the time $t^* = 20'$ the pattern generally changes.

base-line signal that reflects "change" to allow for various inferences by the investigators.

From panels in Figure 3 it is clear that the main challenge arises from highly irregular behavior of responses. Neither physiologically considerations nor any exploratory data analysis motivate any parsimonious parametric form. Different animals seem to exhibit similar but varying response patterns. However, it is clear from inspection of the data that for all response series a definite change is taking place at time $t^*$.

Researchers are interested in the rates of increase of oxygen after time $t^* = 20'$. The baseline signal captures this dynamics more precisely than the average signal because of regularization (denoising).

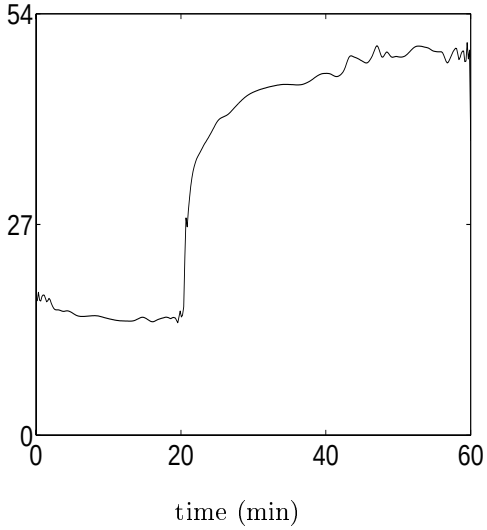


Figure 4: A base-line estimator for $PO_2$ measurements.

# 4 Discussion and Conclusions

In this paper we proposed "vertical block" wavelet shrinkage in which a base-line signal is extracted from a set of multiple measurements. The shrinkage procedure is of Stein-type and automatically regularizes the signal. Since the shrinkage is local (bind to a position in the wavelet domain) applications to $n$-dimensional repeated data are straightforward.

Another version of proposed VERTISHRINK method is so called "supervised" shrinkage. The researcher might be interested not in baseline signal from multiple runs, but in a particular run. VERTISHRINK provides sensible incorporation of information about the run of interest contained in other runs. An example can be the shrinkage estimation of vertical velocity of air in fully developed turbulent environment by taking into account properly normalized related signals (two horizontal velocities, temperature, ozone concentration, etc.) which are measured concurently.

Possible generalizations are numerous. Prior information about the baseline signal can be incorporated in the shrinkage estimator so that the shrinkage is not toward 0 but toward the transformed prior values. In other words, the $\mathcal{NIG}$ prior can be centered at $\mu$, where $\mu$ is wavelet coefficient of a prior signal proposed by experts.

The models can also differ from position to position. One can generalize the method to account for time-varying variances or to dependence in noise. Such generalizations are automatic, but their cost is difficulty in specifying the hyperparameters in the generalized model.

The VERTISHRINK method can also be applied in more general settings, such as wavelet-based ANOVA and wavelet-based experimantal design procedures. In the fixed effect ANOVA setup, the base-line estimator would correspond to the estimator of the grand mean.

# References


[1] Bruce, A. and Gao, H-Y. (1996). Understanding WaveShrink: Variance and bias estimation, Biometrika, 83, 727–745.

[2] Cai, T. and Silverman, B. (2001). Incorporating information on neighboring wavelet coefficients into wavelet estimators, To appear in Sankhya.

[3] DeCanditiis (2001) Wavelet methods for nonparametric regression, Ph. D. Thesis at Consiglio Nazionale Delle Ricerche, IAM, Naples, Italy.

[4] Dewhirst, M. W., Braun, R. D., and Lanzen, J. L. (1998). Temporal changes in $PO_2$ of R3230Ac tumors in Fischer-344 rats. International Journal of Radiation Oncology, Biology, and Physics, 42, 723-726.

[5] Donoho D., Johnstone, I., Kerkyacharian, G., and Picard, D. (1995), Wavelet Shrinkage: Asymptopia? (with discussion), J. Roy. Statist. Soc. Ser. B, 57, 301–369.

[6] Gruber, M. (1998). Improving efficiency by shrinkage: The James-Stein and ridge regression estimators. Marcel Dekker Inc. New York.

[7] Hall, P., Kerkyacharian, G., and Picard, D. (1999). On the minimax optimality of blockthresholded wavelet estimators, Statistica Sinica, 9, 33-50.

[8] Lanzen, J. L., Braun, R. D., Ong, A. L., and Dewhirst, M. W. (1998). Variability in blood flow and $PO_2$ in tumors in response to carbogen breathing. International Journal of Radiation Oncology, Biology, and Physics, 42, 855-859.

[9] Leporini, D. and Pesquet, J-C. (1998). Wavelet Thresholding for some classes of non-Gaussian noise, Preprint at Laboratoire des Signaux et Systèmes, CNRS Université Paris-Sud.

[10] O'Hagan A. (1994). Kendall's Advanced theory of statistics, volume 2B, Bayesian Inference Edward Arnold. London.

[11] Saito, N.(1994) Simultaneous Noise Suppresion and Signal Compression Using a Library of Orthonormal Bases and the Minimum Description Length Criterion Wavelets in Geophysics 299–324.

[12] Stein, C.(1981) Estimation of the Mean of a Multivariate Normal Distribution. Ann. Statist 9 1135–1151

[13] Vannucci, M. and Corradi, F. (1997). Some Findings on the Covariance Structure of Wavelet Coefficients: Theory and Models in a Bayesian Perspective. Technical Report UKC/IMS/97/05, University of Kent at Canterbury, Institute of Mathematics and Statistics.

[14] Vidakovic, B. and Müller, P. (1995). Wavelet shrinkage with affine Bayes rules with applications. ISDS Discussion Paper 95-34, Duke University.

[15] Vidakovic, B. (1998). Nonlinear Wavelet Shrinkage With Bayes Rules and Bayes Factors, Journal of the American Statistical Association, 93, 441, 173–179.

Woojin Chang and Brani Vidakovic
School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332-0205
woojin|brani@isye.gatech.edu