Generalized Cross-Validation for Wavelet Shrinkage in Nonparametric Mixed-Effects Models

Henry Horng-Shing Lu

Institute of Statistics, National Chiao Tung University

Su-Yun Huang

Institute of Statistical Science, Academia Sinica, Taipei

Fang-Jiun Lin

Institute of Statistics, National Chiao Tung University

August 19, 2002

Abstract. A nonlinear wavelet shrinkage estimator was proposed in Huang and Lu (2000). Such an estimator combined the asymptotic equivalence to the best linear unbiased prediction and the Bayesian estimation in nonparametric mixed-effects models. In this article a data-driven GCV method is proposed to select hyperparameters. The proposed GCV method has low computational cost and can be applied to one or higher dimensional data. It can be used for selecting hyperparameters for either level independent or level dependent shrinkage. It can also be used for selecting the primary resolution level and the number of vanishing moments of wavelet basis. The strong consistency of the GCV method is proved.

Key words and phrases. Asymptotic BLUP, Bayesian wavelet shrinkage, generalized cross-validation, nonparametric mixed-effects models, soft thresholding, wavelet shrinkage.

1. Introduction

In the last decade wavelets have been an important and successful tool in signal and image processing, especially for denoising and compression. In denoising and compression, the wavelet coefficients are truncated or shrunk towards zero. There are different approaches for truncating and shrinking wavelet coefficients. Hard and soft thresholding schemes were proposed and studied in a series of papers by Donoho and Johnstone (see, for instance, Donoho and Johnstone, 1994; Donoho, 1995). Later, various Bayesian wavelet shrinkage methods were studied by several authors (Chipman, Kolaczyk and McCulloch, 1997; Abramovich, Sapatinas and Silverman, 1998; Vidakovic, 1998a and 1998b; Huang and Lu, 2000 and 2001). The works of Huang and Lu proposed an adaptive nonlinear shrinkage method, BLUPWAVE, based on perspectives of the Bayesian estimation and the Gauss-Markov estimation (i.e., the BLUP). The shrinkage curve of BLUPWAVE falls between those of hard and soft thresholdings, as seen in Figure 1. These three different thresholding approaches

can be summarized below.

$$\text{Hard thresholding: } \Delta^H(d,\delta) = d\ I(|d| \geq \delta),$$

$$\text{Soft thresholding: } \Delta^S(d,\delta) = sign(d)\ (|d| - \delta)\ I(|d| \geq \delta),$$

$$\text{BLUPWAVE: } \Delta^D(d,\delta) = \left(1 - \frac{\delta^2}{d^2}\right) d\ I(|d| \geq \delta),$$

where d is a wavelet coefficient, δ is a threshold parameter, $I(\cdot)$ is the indicator function, and $sign(\cdot)$ is the signum function. The BLUPWAVE shrinkage rule compromises between hard and soft thresholding by drawing positive aspects of both strategies.

The generalized cross-validation method for parameter selection was proposed and its related consistency was studied in the literature (Craven and Wahba, 1979; Golub, Heath and Wahba, 1979; Li, 1985, 1986 and 1987; Wahba, 1990). In wavelet shrinkage estimation, data-driven procedures based on cross-validation or generalized cross-validation for selecting parameters of soft and hard thresholdings were studied by Weyrich and Warhola (1995 and 1998), Nason (1996 and 1999), Jansen, Malfait and Bultheel (1997), Jansen and Bultheel (1999, 2001), and Jansen (2001). In this article we propose a generalized cross-validation method for selecting parameters encountered in the BLUPWAVE scheme, including selection of threshold parameters at different resolution levels, the primary resolution level and the number of vanishing moments in the wavelet basis. Extending from consistency results in Li (1985, 1986 and 1987) for linear estimation, we present the consistency of GCV for the nonlinear BLUPWAVE. Simulation studies are performed to explore the finite sample behavior in practice.

2. BLUPWAVE

We consider the following model of a discrete noisy signal:

$$y_i = f(t_i) + \epsilon_i, \quad i = 1, \ldots, n, \tag{1}$$

where t_i's are equally spaced design points over an interval $[a, b]$ and the errors ϵ_i's are iid normal random variables with zero mean and variance σ^2. In vector form, we use the notation

$$y = f + \epsilon, \tag{2}$$

where $y = (y_1, \ldots, y_n)'$, $f = (f(t_1), \ldots, f(t_n))'$ and $\epsilon = (\epsilon_1, \ldots, \epsilon_n)'$. The sample size is assumed $n = 2^{m+1}$, $m \in N$. Expand the mean function $f(t)$ in terms of wavelet basis as

$$f(t) = \sum_k \beta_{j,k}\phi_{j,k}(t) \; + \; \sum_{\ell=j}^{\infty}\sum_k \gamma_{\ell,k}\psi_{\ell,k}(t), \tag{3}$$

where j is the primary resolution level, $\{\phi(\cdot), \psi(\cdot)\}$ are a pair of orthogonal scaling function and mother wavelet that generate a multiresolution analysis, and $\phi_{j,k}(t) = 2^{j/2}\phi(2^j t - k)$, $\psi_{\ell,k}(t) = 2^{j/2}\psi(2^j t - k)$.

We adopt a Bayesian approach, with coefficients $\{\beta_{j,k}\}$ being treated as unknown scalars (i.e., fixed effects) and $\{\gamma_{\ell,k}\}$ being treated as random variables (i.e., random effects) having prior distribution $N(0, \eta_\ell)$, where η_ℓ is a variance. An adaptive reconstruction scheme, called BLUPWAVE, has been proposed in Huang and Lu (2000) for this model (also known as a nonparametric mixed-effects model). The reconstruction therein is a thresholding rule that combines keep-or-kill and shrinkage with the following form

$$\hat{f}(t) = \sum_{k=1}^{2^j}\hat{\beta}_{j,k}\phi_{j,k}(t) + \sum_{\ell=j}^{m}\sum_{k=1}^{2^\ell}\left(1 - \frac{\lambda}{\hat{\gamma}^2_{\ell,k}}\right)_+ \hat{\gamma}_{\ell,k}\psi_{\ell,k}(t) \tag{4}$$

where $\hat{\beta}_{j,k}$'s and $\hat{\gamma}_{\ell,k}$'s are the empirical scaling and wavelet coefficients respectively, λ is a certain parameter involving σ, and $(\cdot)_+$ means $\max(\cdot, 0)$. The Bayesian estimation for nonparametric mixed effects is also the best linear unbiased prediction (BLUP). Such a predictor has the following asymptotic form

$$\hat{f}_{BLUP}(t) = \sum_{k=1}^{2^j}\hat{\beta}_{j,k}\phi_{j,k}(t) + \sum_{\ell=j}^{m}\sum_{k=1}^{2^\ell}\left(1 - \frac{\lambda}{\eta_\ell + \lambda}\right)\hat{\gamma}_{\ell,k}\psi_{\ell,k}(t) + O\left(\frac{1}{n\lambda}\right) \quad a.s.$$

Every $\hat{\gamma}^2_{\ell,k}$, $k = 1, \ldots, 2^\ell$, is an asymptotically unbiased estimator for $\eta_\ell + \lambda$. Hence, it leads to the nonlinear BLUPWAVE scheme (4). The details are studied in Huang and Lu (2000). The empirical coefficients, $\hat{\beta}_{j,k}$ and $\hat{\gamma}_{\ell,k}$, can be obtained in linear complexity by a discrete wavelet transform (DWT) of y

$$\begin{pmatrix} \hat{\beta} \\ \hat{\gamma} \end{pmatrix} = \frac{Wy}{\sqrt{n}},$$

where W is the orthogonal matrix associated with the DWT (Mallat 1989). Hereafter, we work on the coefficients w which comes from $w = Wy$, instead of the coefficients $(\hat{\beta}, \hat{\gamma})$. Model (2) can then

be represented as

$$w = Wf + W\epsilon = u + \epsilon^*, \tag{5}$$

where the errors ϵ^* are still iid normal with zero mean and variance σ^2. For notational simplicity, we make no distinction between ϵ and ϵ^* henceforth, and use the notation ϵ for both of the errors ϵ and ϵ^*. The BLUPWAVE thresholding scheme becomes

$$\hat{u}_{\lambda,i} = \begin{cases} w_i & \text{for } i = 1, \ldots, 2^j, \\ \left(1 - \frac{n\lambda}{w_i^2}\right)_+ w_i & \text{for } i = 2^j + 1, \ldots, n. \end{cases} \tag{6}$$

The focus of this article is on GCV selection of the parameter λ. The GCV method can be also extended to select level dependent (i.e., scale dependent) thresholds for one or higher dimensional data.

3. The generalized cross-validation method

3.1. GCV for level independent thresholding. Define the mean square error $R_n(\lambda) = \frac{1}{n}\|\hat{u}_\lambda - u\|_2^2 = \frac{1}{n}\sum_{i=1}^n (\hat{u}_{\lambda,i} - u_i)^2$. The GCV score is defined by

$$\begin{aligned} GCV_n(\lambda) &= \frac{\frac{1}{n}\|\hat{u}_\lambda - w\|_2^2}{\left(1 - \frac{1}{n}\sum_{i=1}^n \frac{\partial \hat{u}_{\lambda,i}}{\partial w_i}\right)^2} \\ &= \frac{\frac{1}{n}\|\hat{u}_\lambda - w\|_2^2}{\left(1 - \frac{2^j}{n} - \frac{1}{n}\sum_{i=2^j+1}^n (1 + \frac{n\lambda}{w_i^2}) I(w_i^2 > n\lambda)\right)^2}. \end{aligned}$$

The GCV estimate of λ, given by $\hat{\lambda}_n = \arg\min_{\lambda \geq 0} GCV_n(\lambda)$, serves as an estimate of the argument minimizer for $R_n(\lambda)$. The strong consistency of GCV theorem below assures that the GCV estimate $\hat{\lambda}_n$ achieves the minimum value for $R_n(\lambda)$ asymptotically under Assumptions 1-3.

Assumption 1 *Assume that the underlying function $f(t)$ is in the Sobolev space $W_2^s[a, b]$, where $s > 0$ is the degree of regularity (or smoothness). Also assume that the wavelet basis used has the number of vanishing moments v with $v > s$.*

Assumption 2 *Assume that $\lambda \to 0$ and $n\lambda \to \infty$, as $n \to \infty$; also assume that the primary resolution level $j \to \infty$, as $n \to \infty$.*

Assumption 3 *Assume that the threshold parameter λ and the primary resolution level j satisfy the constraint $\lim_{n\to\infty} 2^{2js}\lambda = \infty$.*

By Assumption 1 f is in a certain Sobolev space W_2^s. There are also some related Besov spaces $B_{p,q}^s$ imbedded in W_p^s. As the error criterion adopted in this article is the mean square error, we take $p = 2$. Let C^α denote the class of functions that are uniformly Lipschitz with exponent $\alpha > 0$. The following imbeddings are useful: $W_2^s = B_{2,2}^s$, $B_{2,q^-}^{s^+} \subset B_{2,q}^s$, and $C^s \subset W_2^{s^-}$, where $0 < s^- < s \le s^+ < \infty$, and $0 < q^- \le q$. We know that piecewise Lipschitz functions with exponent s are not in W_2^s. They are not in $B_{2,q}^s$, either. The discontinuity at jump point(s) will lower the global regularity. However, such functions are in some other Sobolev spaces with a smaller regularity depending on the type(s) of singularity at the jump point(s). For examples, we discuss below four test functions, Blocks, Bumps, HeaviSine and Doppler, which are used later in the simulation studies. For the Blocks, every block is C^∞ in the interior of its block interval. However, at jump points, the regularity is less than $1/2$. The entire Blocks signal belongs to the space $W_2^s[0,1]$ for any $0 < s < 1/2$, but the regularity will never reach $1/2$. By the above imbedding results, the Blocks signal is also in $B_{2,q}^s$ for any $0 < s < 1/2$ and $2 \le q \le \infty$. The HeaviSine signal has the same phenomenon. It is C^∞ in the interiors of intervals $(0, 0.3)$, $(0.3, 0.72)$ and $(0.72, 1)$. At jump points, 0.3 and 0.72, the regularity is less than $1/2$. The HeaviSine signal belongs to the same Sobolev space as the Blocks does. As for the Bumps, the signal is in $W_2^s[0,1]$ for any $0 < s < 3/2$, and hence in $B_{2,q}^s[0,1]$ for any $0 < s < 3/2$ and $2 \le q \le \infty$. As for the Doppler, the signal is in $C^\infty[h, 1-h]$ for arbitrarily small $h > 0$, but the regularity near the right boundary is less than $3/2$. Hence, the Doppler signal belongs to $W_2^s[h,1]$ for any $0 < s < 3/2$ and $B_{2,q}^s[h,1]$ for any $0 < s < 3/2$ with $2 \le q \le \infty$.

The following Sobolev characterization in Mallat (1989) will be used. A function f is in $W_2^s[a,b]$ if and only if it satisfies the condition: $\sum_{\ell,k\in Z}(1+2^{2\ell s})|\langle f, \psi_{\ell,k}\rangle|^2 < \infty$. Therefore, for $f \in W_2^s[a,b]$, we have

$$\frac{1}{n}\sum_{i=2^j+1}^{n} u_i^2 = \sum_{\ell=j}^{m}\sum_{k=1}^{2^\ell} |\langle f, \psi_{\ell,k}\rangle|^2 = O\left(2^{-2js}\right).$$

As required by Assumption 2, the primary resolution level j increases to infinity as the data size $n \to \infty$. Usually, j goes to infinity at a somewhat slow rate. An intuitive explanation for this slow rate is as follows. For smoother functions (i.e., larger s), j goes to infinity at a slower rate to allow a wider smoothing bandwidth. Assumption 2 requires that $\lambda \to 0$ and $n\lambda \to \infty$, while Assumption 3 controls the convergence speed so that λ goes to zero at a rate slower than $O(2^{-2js})$. If the primary resolution level is of the optimal order $2^j = O(n^{1/(2s+1)})$, then λ is required to go to zero at a rate slower than $O(n^{-2s/(2s+1)})$ and $n\lambda$ is required to go to infinity at a rate faster than $O(n^{1/(2s+1)})$.

That is, the threshold $n\lambda$ for BLUPWAVE is theoretically approaching infinity at a faster rate than the universal threshold $2\sigma^2 \log n$. From Figure 1, the BLUPWAVE must have a larger threshold value than soft thresholding has in order to have the same shrinkage effect (i.e., have the same resulting shrunk coefficient). If we draw a horizontal line in the upper half plane in Figure 1, we have to shift the curve of soft thresholding to the left, which means that the soft thresholding must have a smaller threshold parameter δ, to meet with the BLUPWAVE curve at the same intersection point with the horizontal line.

Theorem 1 (GCV Theorem) *Under Assumptions 1, 2 and 3, we have*

$$\lim_{n\to\infty} \frac{R_n(\hat{\lambda}_n)}{R_n(\lambda_n^*)} = 1 \ \ a.s.,$$

where $\hat{\lambda}_n = \arg\min_{\lambda\geq 0} GCV_n(\lambda)$ *and* $\lambda_n^* = \arg\min_{\lambda\geq 0} R_n(\lambda)$.

Theorem 2 *Under Assumptions 1, 2 and 3, we have*

$$\lim_{n\to\infty} GCV_n(\lambda) = \lim_{n\to\infty} R_n(\lambda) + \sigma^2 \ \ a.s.$$

3.2. Proofs for Theorem 1 and Theorem 2. Define the following notation:

$$\mu_n(\lambda) = \frac{1}{n} \sum_{i=2^j+1}^{n} \left(1 + \frac{n\lambda}{w_i^2}\right) I(w_i^2 > n\lambda),$$

$$p_n(\lambda) = \frac{1}{n} \sum_{i=2^j+1}^{n} (\hat{u}_{\lambda,i} - u_i)(w_i - u_i),$$

$$h_n(\lambda) = R_n(\lambda) - GCV_n(\lambda) + \sigma^2 - 2p_n(\lambda).$$

The next lemmas are established for Theorem 1 and Theorem 2.

Lemma 1 *Under Assumptions 1, 2 and 3, we have*

$$\frac{1}{n} \sum_{i=2^j+1}^{n} Prob\{w_i^2 > n\lambda\} = O\left(\frac{e^{-n\lambda/(8\sigma^2)}}{\sqrt{n\lambda}}\right).$$

Proof: Recall that $n^{-1}\sum_{i=2^j+1}^{n} u_i^2 = O\left(2^{-2js}\right)$. By Assumption 3, $\lambda = o(2^{-2js})$, then we have $|u_i| < \sqrt{n\lambda}/2$ for $i = 2^j + 1, \ldots, n$, for sufficiently large n. Then

$$\frac{1}{n} \sum_{i=2^j+1}^{n} Prob\{w_i^2 > n\lambda\} = \frac{1}{n} \sum_{i=2^j+1}^{n} Prob\{(\epsilon_i + u_i)^2 > n\lambda\}$$

$$\leq \frac{2}{n}\sum_{i=2^j+1}^{n}\int_{\sqrt{n\lambda}/2}^{\infty}\frac{e^{-y^2/(2\sigma^2)}}{\sqrt{2\pi}\ \sigma}dy \leq \frac{1}{\sqrt{\pi}}\int_{n\lambda/(8\sigma^2)}^{\infty}\frac{e^{-t}}{\sqrt{t}}\,dt$$
$$\leq \frac{2\sqrt{2}\ \sigma e^{-n\lambda/(8\sigma^2)}}{\sqrt{\pi n\lambda}} = O\left(\frac{e^{-n\lambda/(8\sigma^2)}}{\sqrt{n\lambda}}\right).$$

The last inequality follows from the inequality $1/\sqrt{t} \leq 2\sqrt{2}\,\sigma/\sqrt{n\lambda}$, valid for t in the limits of integration. □

Lemma 2 *Assume Assumptions 1, 2 and 3. For any arbitrary $\alpha > 0$, we have*

$$\frac{1}{n}\sum_{i=2^j+1}^{n} I(w_i^2 > n\lambda) = o\left(\frac{1}{(n\lambda)^\alpha}\right) \quad a.s.$$

Proof: For arbitrary $\alpha > 0$ and $q > 2\alpha$, we have

$$\max_{\substack{2^j+1\leq i\leq n\\ |u_i|<\sqrt{n\lambda}/2}} (n\lambda)^{2\alpha} E\, I(w_i^2 > n\lambda) = \max_{\substack{2^j+1\leq i\leq n\\ |u_i|<\sqrt{n\lambda}/2}} (n\lambda)^{2\alpha} Prob\{w_i^2 > n\lambda\}$$
$$\leq \max_{2^j+1\leq i\leq n} (n\lambda)^{2\alpha} Prob\{|\epsilon_i| > \sqrt{n\lambda}/2\} \leq \max_{2^j+1\leq i\leq n} \frac{4^q E|\epsilon_i|^{2q}}{(n\lambda)^{q-2\alpha}} < \infty.$$

The second inequality above is a Markov inequality. We then apply the strong law of large numbers for uncorrelated random variables having a common upper bound to their second moments (Theorem 5.1.2. in Chung (1974)), and get

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=2^j+1}^{n}(n\lambda)^\alpha I(w_i^2 > n\lambda) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=2^j+1}^{n}(n\lambda)^\alpha E\, I(w_i^2 > n\lambda) \quad a.s. \tag{7}$$

By Lemma 1, the limit in (7) is zero. Therefore, we can conclude Lemma 2. □

Lemma 3 *Assume Assumptions 1, 2 and 3. For any arbitrary $\alpha > 0$, we have*

$$\mu_n(\lambda) = o\left(\frac{1}{(n\lambda)^\alpha}\right) \quad a.s.$$

Proof: Note that $(1 + \frac{n\lambda}{w_i^2})I(w_i^2 > n\lambda) \leq 2I(w_i^2 > n\lambda)$. Thus $\mu_n(\lambda) \leq \frac{2}{n}\sum_{i=2^j+1}^{n} I(w_i^2 > n\lambda)$. We can conclude Lemma 3 from Lemma 2. □

Lemma 4 *Assume Assumptions 1, 2 and 3. For any arbitrary $\lambda_1 > 0$, $\lambda_2 > 0$ and $\alpha > 0$, we have* $|p_n(\lambda_1) - p_n(\lambda_2)| = o\left((n\lambda_2)^{1/2}/(n\lambda_1)^\alpha\right)$ *a.s.*

Proof: Without loss of generality, we may assume that $\lambda_1 \leq \lambda_2$. For $\kappa = 1, 2$, let $A_{\kappa,n} = \{i : 2^j + 1 \leq i \leq n \text{ and } w_i^2 > n\lambda_\kappa\}$ and $A_{\kappa,n}^c = \{i : 2^j + 1 \leq i \leq n\} \setminus A_{\kappa,n}$.

$$
\begin{aligned}
|p_n(\lambda_1) - p_n(\lambda_2)| &= \frac{1}{n} \left| \sum_{i=2^j+1}^{n} (\hat{u}_{\lambda_1,i} - \hat{u}_{\lambda_2,i})(w_i - u_i) \right| \\
&= \frac{1}{n} \left| \sum_{A_{2,n}} \frac{(n\lambda_1 - n\lambda_2)\epsilon_i}{w_i} + \sum_{A_{1,n} \cap A_{2,n}^c} \left(w_i - \frac{n\lambda_1}{w_i} \right) \epsilon_i \right|. \qquad (8)
\end{aligned}
$$

Similar to the proof for Lemma 2, we have

$$
\frac{1}{n} \sum_{A_{2,n}} \left| \frac{(n\lambda_1 - n\lambda_2)\epsilon_i}{w_i} \right| \leq \sqrt{n\lambda_2} \left(\frac{\sum_{A_{2,n}} |\epsilon_i|}{n} \right) = o\left((n\lambda_2)^{-\alpha+1/2} \right) \quad a.s.
$$

for arbitrary $\alpha > 0$, and

$$
\frac{1}{n} \sum_{A_{1,n} \cap A_{2,n}^c} w_i \left(1 - \frac{n\lambda_1}{w_i^2} \right) \epsilon_i = o\left(\frac{(n\lambda_2)^{1/2}}{(n\lambda_1)^\alpha} \right) \quad a.s. \qquad (9)
$$

Therefore, we can conclude Lemma 4. □

Lemma 5 *Assume Assumptions 1, 2 and 3. For any arbitrary $\lambda_1 > 0$ and $\lambda_2 > 0$, we have*

$$
|h_n(\lambda_1) - h_n(\lambda_2)| = O\left(\frac{4^j}{n^2} \right) + O\left(\frac{1}{n\ 2^{j(s-1)}} \right) \quad a.s.
$$

Proof: Without loss of generality, we may assume $\lambda_1 \leq \lambda_2$. Let $\nu_n(\lambda) = 2^j/n + \mu_n(\lambda)$. Observe that $\frac{1}{n}\|\hat{u}_\lambda - w\|_2^2 = R_n(\lambda) - 2p_n(\lambda) - \frac{1}{n}\sum_{i=1}^{2^j} \epsilon_i^2 + \frac{1}{n}\sum_{i=2^j+1}^{n} \epsilon_i^2$. Then $h_n(\lambda)$ and $h_n(\lambda_1) - h_n(\lambda_2)$ can be expressed as follows.

$$
\begin{aligned}
h_n(\lambda) &= \frac{(R_n(\lambda) + \sigma^2)(1 - \nu_n(\lambda))^2 - n^{-1}\|\hat{u}_\lambda - w\|_2^2 - 2p_n(\lambda)(1 - \nu_n(\lambda))^2}{(1 - \nu_n(\lambda))^2}, \\
h_n(\lambda_1) - h_n(\lambda_2) &= \frac{\{R_n(\lambda_1) + \sigma^2 - 2p_n(\lambda_1)\}\{-2\nu_n(\lambda_1) + \nu_n^2(\lambda_1)\}}{\{1 - \nu_n(\lambda_1)\}^2} \\
&\quad - \frac{\{R_n(\lambda_2) + \sigma^2 - 2p_n(\lambda_2)\}\{-2\nu_n(\lambda_2) + \nu_n^2(\lambda_2)\}}{\{1 - \nu_n(\lambda_2)\}^2} \\
&= \sigma^2 \left(\frac{\{-2\nu_n(\lambda_1) + \nu_n^2(\lambda_1)\}}{\{1 - \nu_n(\lambda_1)\}^2} - \frac{\{-2\nu_n(\lambda_2) + \nu_n^2(\lambda_2)\}}{\{1 - \nu_n(\lambda_2)\}^2} \right) \\
&\quad + \frac{\{R_n(\lambda_1) - 2p_n(\lambda_1)\}\{-2\nu_n(\lambda_1) + \nu_n^2(\lambda_1)\}}{\{1 - \nu_n(\lambda_1)\}^2} \\
&\quad - \frac{\{R_n(\lambda_2) - 2p_n(\lambda_2)\}\{-2\nu_n(\lambda_2) + \nu_n^2(\lambda_2)\}}{\{1 - \nu_n(\lambda_2)\}^2} \stackrel{\text{def}}{=} I + II - III. \qquad (10)
\end{aligned}
$$

As $R_n(\lambda) = O\left(\frac{2^j}{n}\right) + O\left(2^{-2js}\right)$, $p_n(\lambda) = O\left(2^{-js}\right)$ and $\nu_n(\lambda) = O\left(\frac{2^j}{n}\right)$, *a.s.*, we have

$$|I| = O(\nu_n(\lambda_1)) + O(\nu_n(\lambda_2)) = O\left(\frac{2^j}{n}\right), \text{ and}$$

$$\begin{aligned} |II| &= O\left(R_n(\lambda_1) - 2p_n(\lambda_1)\right) O\left(\nu_n(\lambda_1)\right) \\ &= \left(O\left(\frac{2^j}{n}\right) + O\left(2^{-js}\right)\right) O\left(\nu_n(\lambda_1)\right) \\ &= O\left(\frac{4^j}{n^2}\right) + O\left(\frac{1}{n2^{j(s-1)}}\right). \end{aligned} \tag{11}$$

Similarly, III has the same order as II. Therefore, we can conclude Lemma 5. □

Proof for the GCV Theorem: Notice that

$$\begin{aligned} 1 &\leq \frac{R_n(\hat{\lambda}_n)}{R_n(\lambda_n^*)} = \frac{GCV_n(\hat{\lambda}_n) - \sigma^2 + 2p_n(\hat{\lambda}_n) + h_n(\hat{\lambda}_n)}{R_n(\lambda_n^*)} \\ &\leq \frac{GCV_n(\lambda_n^*) - \sigma^2 + 2p_n(\hat{\lambda}_n) + h_n(\hat{\lambda}_n)}{R_n(\lambda_n^*)} \\ &= \frac{R_n(\lambda_n^*) - 2p_n(\lambda_n^*) + 2p_n(\hat{\lambda}_n) + h_n(\hat{\lambda}_n) - h_n(\lambda_n^*)}{R_n(\lambda_n^*)} \\ &= 1 + \frac{-2p_n(\lambda_n^*) + 2p_n(\hat{\lambda}_n) + h_n(\hat{\lambda}_n) - h_n(\lambda_n^*)}{R_n(\lambda_n^*)}. \end{aligned} \tag{12}$$

Let $A_n = \{i : 2^j + 1 \leq i \leq n \text{ and } w_i^2 > n\lambda\}$ and $A_n^c = \{i : 2^j + 1 \leq i \leq n\} \setminus A_n$. Notice that

$$R_n(\lambda) = \frac{1}{n}\sum_{i=1}^{2^j} \epsilon_i^2 + \frac{1}{n}\sum_{i \in A_n} (\hat{u}_i - u_i)^2 + \frac{1}{n}\sum_{i \in A_n^c} u_i^2 \geq \frac{1}{n}\sum_{i=1}^{2^j} \epsilon_i^2 \sim \frac{2^j\sigma^2}{n}, \tag{13}$$

where $a_n \sim b_n$ means that $\lim_{n\to\infty} a_n/b_n = 1$. By Lemmas 4 and 5 and inequalities (12) and (13), we have

$$\lim_{n\to\infty} \frac{-2p_n(\lambda_n^*) + 2p_n(\hat{\lambda}_n) + h_n(\hat{\lambda}_n) - h_n(\lambda_n^*)}{R_n(\lambda_n^*)} = 0 \quad a.s.$$

Therefore, $\lim_{n\to\infty} R_n(\hat{\lambda}_n)/R_n(\lambda_n^*) = 1$, *a.s.* □

Proof for Theorem 2: This follows immediately from the proof for Theorem 1 by observing that $\lim_{n\to\infty} h(\lambda) = 0$ *a.s.* and that $\lim_{n\to\infty} p_n(\lambda) = 0$ *a.s.* □

3.3. Level dependent thresholding. A more flexible approach, introduced by Johnstone and Silverman (1997), is to allow the threshold parameter of wavelet shrinkage to be level dependent. We will consider the following level-dependent BLUPWAVE scheme

$$\hat{u}_{\lambda,i} = \begin{cases} w_i & \text{for } i = 1, \ldots, 2^j, \\ \left(1 - \frac{n\lambda_\ell}{w_i^2}\right)_+ w_i & \text{for } 2^\ell + 1 \leq i \leq 2^{\ell+1},\ \ell = j, \ldots, m. \end{cases} \tag{14}$$

The thresholds, $\Lambda = (\lambda_j, \ldots, \lambda_m)$, form an array of nonnegative parameters.

Since the GCV score function for level dependent thresholding is multivariate, it is difficult to minimize. To reduce the multivariate minimization problem to a univariate problem iteratively, we minimize the GCV score function by coordinatewise descent. This turns the multivariate minimization problem into a sequence of easily solved one dimensional problems. The initial threshold parameters are set equal to the level independent threshold, i.e., $\hat{\Lambda}_0 = (\hat{\lambda}, \ldots, \hat{\lambda})$ and $k = 0$. The next iteration, $\hat{\Lambda}_{k+1}$ is computed as follows.

- $\hat{\lambda}_{k+1,m} = \arg\min_{\lambda_m \geq 0} GCV(\hat{\lambda}_{k,j}, \ldots, \hat{\lambda}_{k,m-1}, \lambda_m)$, $\lambda_m \leftarrow \hat{\lambda}_{k+1,m}$.
- $\hat{\lambda}_{k+1,m-1} = \arg\min_{\lambda_{m-1} \geq 0} GCV(\hat{\lambda}_{k,j}, \ldots, \hat{\lambda}_{k,m-2}, \lambda_{m-1}, \hat{\lambda}_{k+1,m})$, $\lambda_{m-1} \leftarrow \hat{\lambda}_{k+1,m-1}$.
  $\vdots$
- $\hat{\lambda}_{k+1,j} = \arg\min_{\lambda_j \geq 0} GCV(\lambda_j, \hat{\lambda}_{k+1,j+1}, \ldots, \hat{\lambda}_{k+1,m})$, $\lambda_j \leftarrow \hat{\lambda}_{k+1,j}$.

For each level, the golden section search method is used to solve the one dimensional minimization problem. We found that GCV scores are close to convergence after one iteration of coordinatewise descent. Furthermore, the average square errors (ASEs) are found to be reduced by this method. Hence, in principle, one iteration of coordinatewise descent is used in our simulation studies reported later. The iteration order can go from m to j or from j to m. Our simulation results shows that different iteration orders are not significantly different.

3.4. GCV selection for the primary resolution level and for the number of vanishing moments of wavelet basis. Nason (1999) discussed choosing the number of vanishing moments in the wavelet basis, primary resolution j and the threshold in wavelet shrinkage by cross-validation. He considered wavelet shrinkage using the universal threshold $\delta = \sqrt{2 \log n}\ \sigma$ (with $\sigma = 1$ in his simulation study) and applied the leave-one-out cross-validation method to select the level j as well as the number of vanishing moments in the basis v. Conditioned on the selected values of (j, v), he then minimized the leave-one-out cross-validation score function to select the level independent threshold λ.

To reduce the computational complexity, we use the GCV method to select level dependent thresholds $\Lambda = (\lambda_j, \ldots, \lambda_m)$ and parameters (j, v) for the BLUPWAVE. We first compute the minimums of the GCV score functions for the BLUPWAVE with a level independent threshold with respect to all possible values of the primary resolution level ranging from 0 to m and the number of

vanishing moments of Symmlet basis ranging from 4 to 10, where Symmlet refers to the least asymmetric wavelet described in Daubechies (1992). The GCV scores are denoted as $GCV(\hat{\lambda}(j,v);j,v)$, $j=0,\ldots,m$, and $v=4,\ldots,10$. Then, we select the primary resolution level $\hat{j}$ and the number of vanishing moments $\hat{v}$ which achieves the minimum among $GCV(\hat{\lambda}(j,v);j,v)$. Conditioned on the selected $(\hat{j},\hat{v})$, we then consider the level dependent thresholds and apply the multivariate GCV method described in the previous section to select the threshold for every level.

3.5. High dimensional data. The above discussion for 1D signals can be extended naturally to 2D images or higher p-dimensional data. Assume the sample size is $n=2^{p(m+1)}$. The discrete wavelet transform and its inverse can be proceeded by taking the tensor product of multiresolution analysis for one dimensional space. Since the above wavelet shrinkage is performed componentwisely, the GCV method can be easily applied to select the level independent threshold λ as well as level dependent thresholds as in 1D cases. For 2D images, the level dependent thresholds are $\lambda_{\ell,k}$, where $k=1,2,3$ are indices for horizontal, vertical and diagonal wavelet components respectively, and $\ell=j,\ldots,m$. For high dimensional cases, the procedures are similar. Moreover, the primary resolution level j and the number of vanishing moments of wavelet basis v can be adjusted to improve the performance of reconstruction similarly as in the 1D cases.

4. Simulation results and discussions

4.1. One dimensional signals. Four test signals, Blocks, Bumps, HeaviSine and Doppler, from Donoho and Johnstone (1994), are used in the simulation study. The periodic Symmlet wavelet basis over $[0,1]$ is applied, since these four examples are periodic. The computation is based on the WaveLab package developed by Donoho, Duncan, Huo, and Levi-Tsabari (1999) for MATLAB.

4.1.1. Level independent thresholding for 1D data. Define the average square error by $ASE=\sum_{i=1}^{n}\{\hat{f}(t_i)-f(t_i)\}^2/n$ and the standardized ASE by

$$\text{standardized } ASE=\frac{1}{n\sigma^2}\sum_{i=1}^{n}\{\hat{f}(t_i)-f(t_i)\}^2.$$

Notice that the ASE is the same as the mean square error $R_n(\cdot)$, since $f=W'u$ and $\hat{f}=W'\hat{u}$, where W is the orthogonal matrix associated with the DWT. The GCV/σ^2 and ASE/σ^2 curves by BLUPWAVE for four test noisy signals are plotted in Figure 2, where $n=1024$ and the signal to noise ratio is $SNR=7$. The SNR is defined by $(\sqrt{n}\,\sigma)^{-1}\|f-\bar{f}\|_{\ell_2}$ with $\bar{f}=\sum_{i=1}^{n}f(t_i)/n$.

Two reference statistics are given for each plot to check the performance of the GCV procedure in finite samples. They are the difference, diff $= \min_\lambda GCV(\lambda)/\sigma^2 - \min_\lambda ASE(\lambda)/\sigma^2$, and the relative efficiency

$$\mathcal{R} = \frac{\min_{\lambda \geq 0} ASE(\lambda)}{ASE(\hat{\lambda})},$$

where $\hat{\lambda}$ is the GCV selection of λ. Both quantities, diff and $\mathcal{R}$, should approach 1 as n goes to infinity. The search domain of the threshold parameter, $\delta^2 = n\lambda$, is set to $[0, 2\sigma^2 \log n]$. For the purpose of presentation, the horizontal axis in Figure 2 is taken to be $x = n\lambda/(\sigma^2 \log n)$, $0 \leq x \leq 2$.

These four test signals are further studied with sample sizes varying from $n = 256$ to $n = 8192$, SNRs equal to 3, 5, 7 and 10. The Symmlets with 8 vanishing moments are used. The primary resolution level is $j = 5$ and the remaining fine scales are all included up to the finest possible resolution. The averages and standard errors of standardized average square errors based on 100 replication runs with various sample sizes and SNRs are reported in Table 1 for BLUPWAVE and in Table 2 for soft thresholding. Note that the reported numbers are based on standardized ASEs, which are ASEs scaled by error variance σ^2. In other words, the larger SNR cases have the ASEs divided by smaller σ^2 and result in inflated standardized ASEs. Therefore, the reader may find larger standardized ASEs for higher SNR cases. The purpose of standardization is to facilitate comparison of the GCV and ASE curves. Plots of standardized ASEs along with ± 1.96 times standard errors are given in Figure 3. In the case of Blocks, Bumps and Doppler, the BLUPWAVE performs better than the soft thresholding does, while in the case of HeaviSine both methods perform with the same quality.

4.1.2. Level dependent thresholding for 1D data. The averages and standard errors of standardized ASEs based on 100 replications of test signals with various sample sizes and SNRs are reported in Table 3 for BLUPWAVE and in Table 4 for soft thresholding.

The comparison of standardized ASEs of BLUPWAVE with or without level dependent (LD) thresholding based on 100 replications when SNR=7 is shown in Figure 4. We observe that the BLUPWAVE has similar quality of performance for level independent and level dependent thresholding. The primary resolution level here is $j = 5$. The comparison plots for BLUPWAVE and soft thresholding are in Figure 5. Again, with level dependent thresholding, the BLUPWAVE has smaller standardized ASEs than soft thresholding does in most cases.

4.1.3. Adjustment for the primary resolution level and the number of vanishing

moments of wavelet basis for 1D data. We take the case of Blocks with $n = 1024$ and SNR=7 in this simulation as an example. The minimums of GCV/σ^2 values for the BLUPWAVE with level independent thresholding for various values of (j, v) are reported in Table 5. Among these values the minimum in this table is marked with an asterisk $*$, which corresponds to $\hat{j} = 3$ and $\hat{v} = 8$. Then, conditioned on $\hat{j} = 3$ and $\hat{v} = 8$, the level dependent thresholds $\hat{\lambda}_3, \ldots, \hat{\lambda}_9$ are computed and the standardized ASE is found to be 0.3054. The corresponding value for the case of level dependent thresholding with pre-assigned primary resolution level $j = 5$ and pre-assigned number of vanishing moments $v = 8$ is 0.3344 in Table 3. Thus the further adjustment of (j, v) does indeed reduce the standardized ASE.

The standardized ASE for BLUPWAVE with pre-assigned values $j = 5$ and $v = 8$ using level dependent thresholding (denoted by BLUP/LD) is compared in Figure 6 with the standardized ASE with further adjustment for (j, v) (BLUP/LD/JV) based on 100 replications and with SNR=7. It is observed that the GCV selection for j and v do reduce the standardized ASE, especially when n is small. For the case of SNR=7, the averages and standard errors of standardized ASEs based on 100 replications for various sample sizes are reported in Table 6 for BLUPWAVE and in Table 7 for soft thresholding. Comparisons of BLUP/LD/JV and SOFT/LD/JV are shown in Figure 7. The standardized ASEs of BLUP/LD/JV are smaller than those of SOFT/LD/JV in most cases.

4.2. 2D Images. Eight test images are used to investigate the performance of the GCV selection. They are Lena 512×512, Lena 256×256, couple 512×512, couple 256×256, goldhill, peppers, fishing boat, and Tiffany (the latter four are of size 512×512). These images can be obtained from the web sites at

http://sipi.usc.edu/services/database/database.cgi?volume=misc, and

http://links.uwaterloo.ca/greyset1.base.html.

The original images are shown in Figure 8 and Figure 9. For every image noises, at levels $\sigma = 10$, 20 and 40, are added to the images for simulation study. The Daubechies wavelets with 8 vanishing moments and the smooth padding of DWT extension mode (i.e., first derivative interpolation at the edges) are used, as the test images are not periodic. The computation is based on the Wavelet Toolbox for MATLAB.

4.2.1. Level independent thresholding for 2D data. The primary resolution level is $j = m - 3$; i.e., $j = 5$ for the 512×512 resolution images and $j = 4$ for the 256×256 resolution images. The GCV/σ^2 and ASE/σ^2 curves by BLUPWAVE are plotted in Figure 10 and Figure 11,

with SNR = 3. The averages and standard errors of standardized ASEs, based on 50 replications with three noise levels, are reported in Table 8 for BLUPWAVE and soft thresholding. The two thresholding schemes have similar performance quality.

4.2.2. Level dependent thresholding for 2D data. The averages and standard errors of standardized ASEs based on 50 replications with three noise levels are reported in Table 9 for BLUPWAVE and soft thresholding. These standardized ASEs along with ± 1.96 times standard errors are plotted in Figure 12 and Figure 13. We observe that the performance of BLUPWAVE is better than the soft thresholding.

4.2.3. Adjustment for the primary resolution level and the number of vanishing moments of wavelet basis for 2D data. The GCV method is used to select level dependent thresholds, the primary resolution level and the number of vanishing moments of wavelet basis. The averages and standard errors of standardized ASEs based on 50 replications with three noise levels are reported in Table 10 for BLUPWAVE and soft thresholding. Figure 14 and Figure 15 are the standardized ASEs along with ± 1.96 times standard errors for BLUPWAVE and soft thresholding. Again, we observe that the performance of BLUPWAVE is better than the soft thresholding, and out-performance is especially obvious when the SNRs are small.

5. Concluding discussion

The BLUPWAVE shrinkage scheme is based on the use of best linear unbiased prediction as well as Bayesian modelling and estimation. The GCV selection for hyperparameters is proposed. Based on our simulation experience, we find the GCV selection method fast and reliable, and it can also be applied to the selection of level dependent thresholds, the primary resolution level and the number of vanishing moments of wavelet basis. With the help of the GCV selection for parameters involved in the BLUPWAVE scheme, we find that BLUPWAVE performs well. Compared to other Bayesian shrinkage schemes, the BLUPWAVE has a very simple asymptotic form. Though its derivation and theoretical background may seem complicated, its final form for shrinkage is really straightforward and easy to implement.

For one dimensional test signals, BLUPWAVE has smaller standardized ASEs than soft thresholding does. Moreover, the BLUPWAVE has similar performance quality for both the level independent thresholding and the level dependent thresholding. Furthermore, standardized ASEs can be reduced by using GCV to also select the primary resolution level and the number of vanish-

ing moments in the wavelet basis. Especially when n is small, the reduction is quite significant. BLUPWAVE also offers some advantages for denoising 2D images.

Acknowledgments. This research was partially supported by the National Science Council, Taiwan, R.O.C. The authors thank the associate editor and referees for their valuable comments and editorial suggestions.

References

Abramovich, F., Sapatinas, T. and Silverman, B. W. (1998). Wavelet thresholding via a Bayesian approach. *J. Roy. Statist. Soc.*, B 60, 725–749.

Chipman, H. A., Kolaczyk, E. D. and McCulloch, R. E. (1997). Adaptive Bayesian wavelet shrinkage. *J. Amer. Statist. Assoc.* 92, 1413–1421.

Chung, K. L. (1974). *A Course in Probability Theory.* (2nd Edition). Academic Press, Boston.

Craven, P. and Wahba, G. (1979). Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, 31, 377–403.

Daubechies, I. (1992). *Ten Lectures on Wavelets.* CBMS-NSF Series of Applied Mathematics, SIAM, Philadelphia.

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

Donoho, D. L., Duncan, M., Huo, X., and Levi-Tsabari, O. (1999). About WaveLab. Technical Report, Department of Statistics, Stanford University.

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

Golub, G., Heath, M. and Wahba, G. (1979). Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics*, 21, 215–223.

Huang, S.Y. and Lu, H. H.-S. (2000). Bayesian wavelet shrinkage for nonparametric mixed-effects models. *Statist. Sinica*, 10, 1021–1040.

Huang, S.Y. and Lu, H. H.-S. (2001). Extended Gauss-Markov theorem for nonparametric mixed-effects models. *J. Multivariate Analy.*, 76, 2, 249–266.

Jansen, M. and Bultheel, A. (1999). Multiple wavelet threshold estimation by generalized cross validation for images with correlated noise. *IEEE Trans. Image Proc.*, 8(7), 947–953.

Jansen, M. and Bultheel, A. (2001). Asymptotic behavior of the minimum mean squared error threshold for noisy wavelet coefficients of piecewise smooth signals. *IEEE Trans. Signal Proc.*, 49(6), 1113–1118.

Jansen, M., Malfait, M. and Bultheel, A. (1997). Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1), 33–44.

Jansen, M. (2001). *Noise Reduction by Wavelet Thresholding.* Lecture Notes in Statistics, 161, Springer-Verlag, New York.

Johnstone, I.M. and Silverman, B.W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc.*, B, 59, 319–351.

Li, K.C. (1985). From Stein's unbiased risk estimates to the method of generalized cross validation. *Ann. Statist.*, 13, 1352–1377.

Li, K.C. (1986). Asymptotic optimality of C_L and generalized cross-validation in ridge regression with application to spline smoothing. *Ann. Statist.*, 14, 1101–1112.

Li, K.C. (1987). Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: discrete index set. *Ann. Statist.*, 15, 958–975.

Mallat, S.G. (1989). Multiresolution approximations and wavelet orthonormal bases of $L^2(R)$. *Trans. Amer. Math. Soc.*, 315, 69–87.

Nason, G.P. (1996). Wavelet shrinkage using cross-validation. *J. Roy. Statist. Soc.*, B, 58, 463–479.

Nason, G.P. (1999). Fast cross-validatory choice of wavelet smoothness, primary resolution and threshold in wavelet shrinkage using the Kovac-Silverman algorithm. Technical report, Department of Mathematics, University of Bristol, United Kingdom.

Vidakovic, B. (1998a). Non-linear wavelet shrinkage with Bayes rules and Bayes factors. *J. Amer. Statist. Assoc.* 93, 173–179.

Vidakovic, B. (1998b). Wavelet-based nonparametric Bayes methods. In *Practical Nonparametric and Semiparametric Bayesian Statistics* (eds. D. Dey, P. Müller and D. Sinha), 133–155. Springer-Verlag, New York.

Wahba, G. (1990). *Spline Models for Observational Data.* CBMS-NSF Series of Applied Mathematics, SIAM, Philadelphia.

Weyrich, N. and Warhola, G.T. (1995). De-noising using wavelets and cross validation. *Approximation Theory, Wavelets and Applications*, 523–532.

Weyrich, N. and Warhola, G.T. (1998). Wavelet shrinkage and generalized cross validation for image denoising, *IEEE Tran. Image Processing*, 7(1), 82–90.

Henry Horng-Shing Lu
Fang-Jiun Lin
Institute of Statistics
National Chiao Tung University
Hsinchu, Taiwan, R.O.C.
e-mail : hslu@stat.nctu.edu.tw

Su-Yun Huang
Institute of Statistical Science
Academia Sinica
Taipei, Taiwan, R.O.C.
e-mail : syhuang@stat.sinica.edu.tw

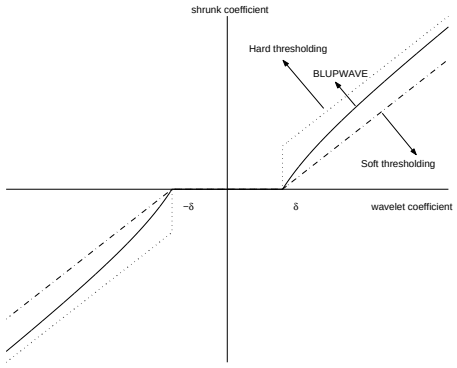

Figure 1: Shrinkage schemes by hard, soft and BLUPWAVE thresholdings.

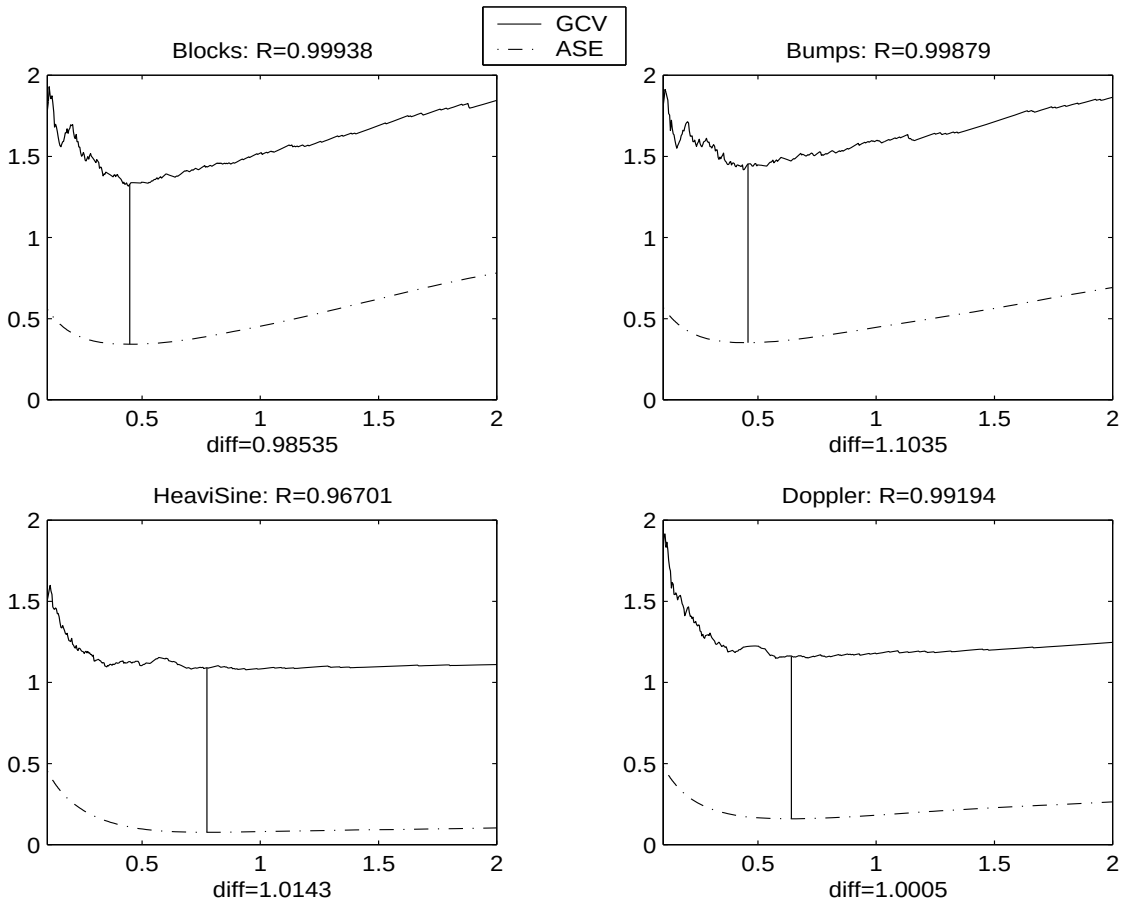

Figure 2: The GCV/σ^2 and ASE/σ^2 curves by BLUPWAVE for four noisy signals when $n = 1024$ and SNR=7.

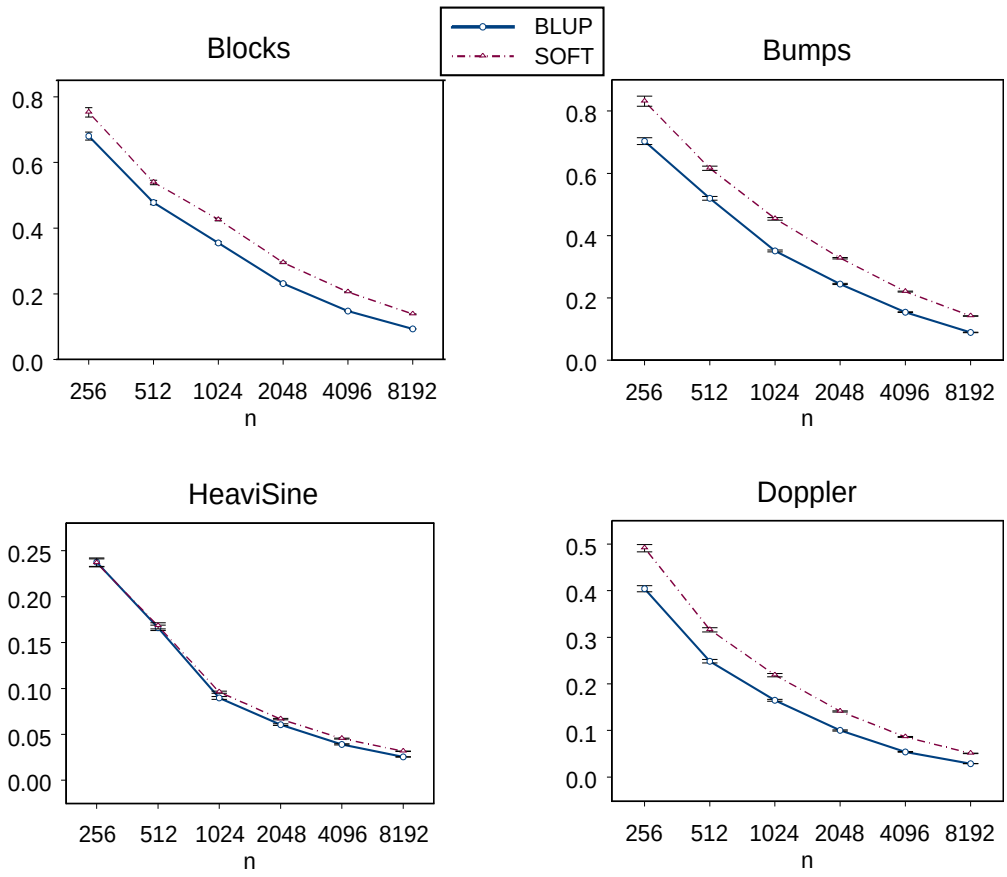

Figure 3: Averages and standard errors of standardized ASEs based on 100 replications of BLUP-WAVE and soft thresholding, with SNR=7.

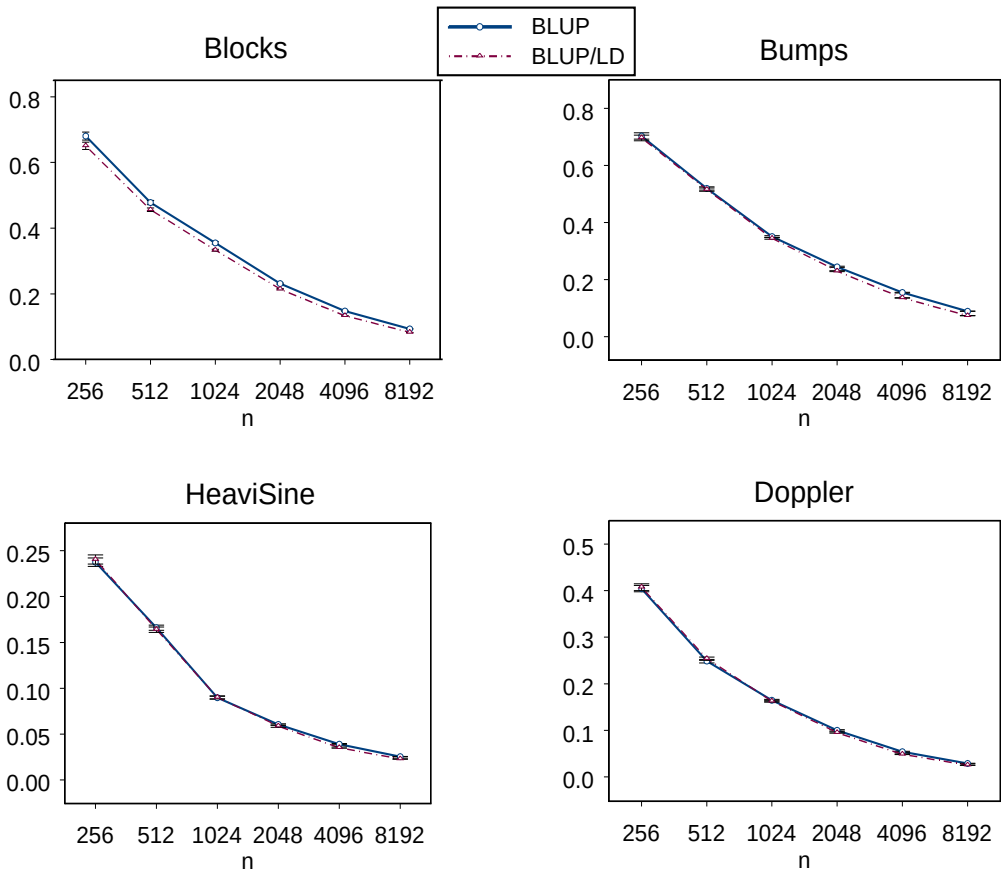

Figure 4: Comparison of standardized ASEs for BLUPWAVE without or with level dependent (LD) thresholding based on 100 replications, with SNR=7.

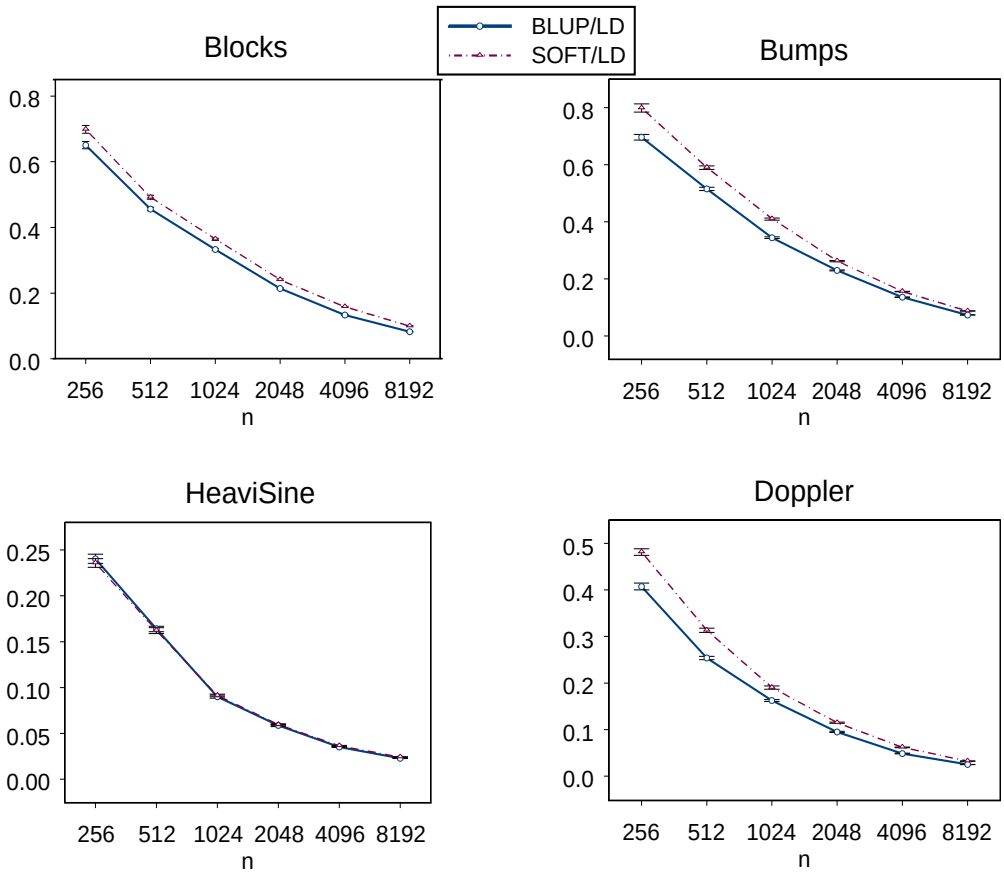

Figure 5: Comparison of the standardized ASEs for BLUPWAVE and soft thresholding with level dependent (LD) thresholds, with SNR=7.

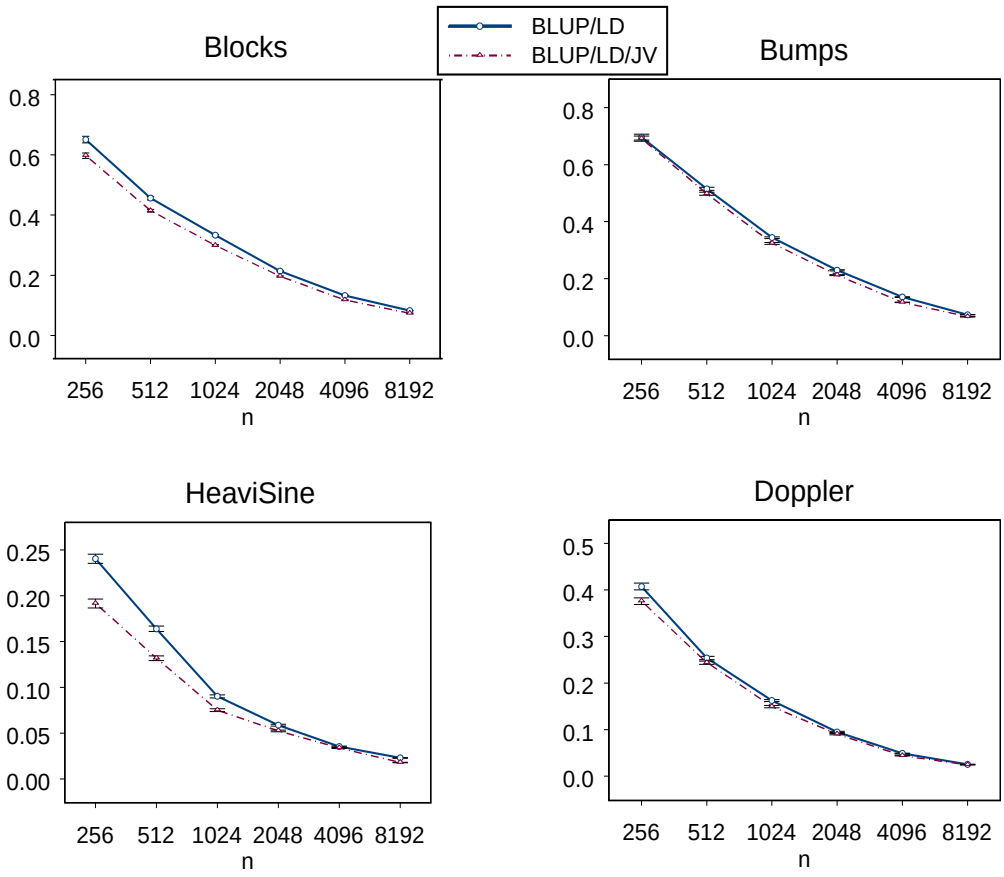

Figure 6: Comparison of the standardized ASEs for BLUP/LD (with pre-assigned $j = 5$ and $v = 8$) *vs.* BLUP/LD/JV (using GCV selection for j and v), with SNR=7.

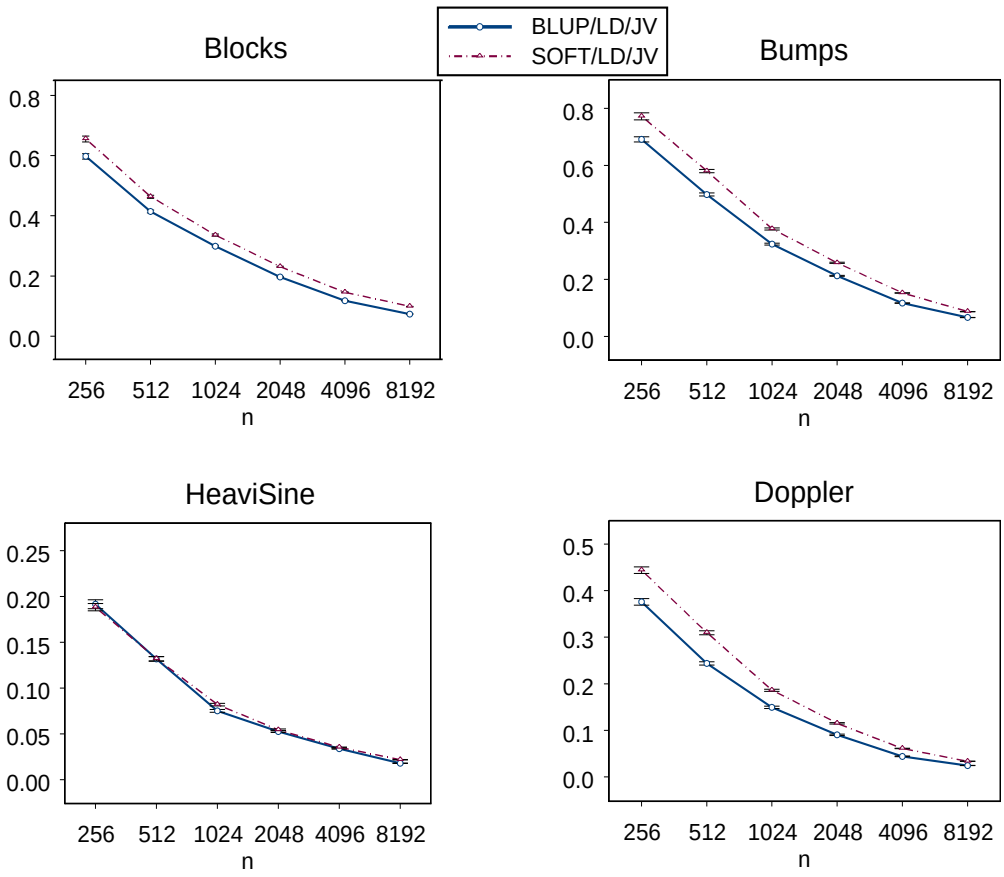

Figure 7: Comparison of the standardized ASEs for BLUP/LD/JV *vs.* SOFT/LD/JV based on 100 replications, with SNR=7.

Figure 8: Test images used in the study. Top left is for Lena: 512×512, top right is for Lena: 256×256, bottom left is for couple: 512×512, and bottom right is for couple: 256×256.

Figure 9: Test images used in the study. Top left is for goldhill: 512×512, top right is for peppers: 512×512, bottom left is for boat: 512×512, and bottom right is for Tiffany: 256×256.

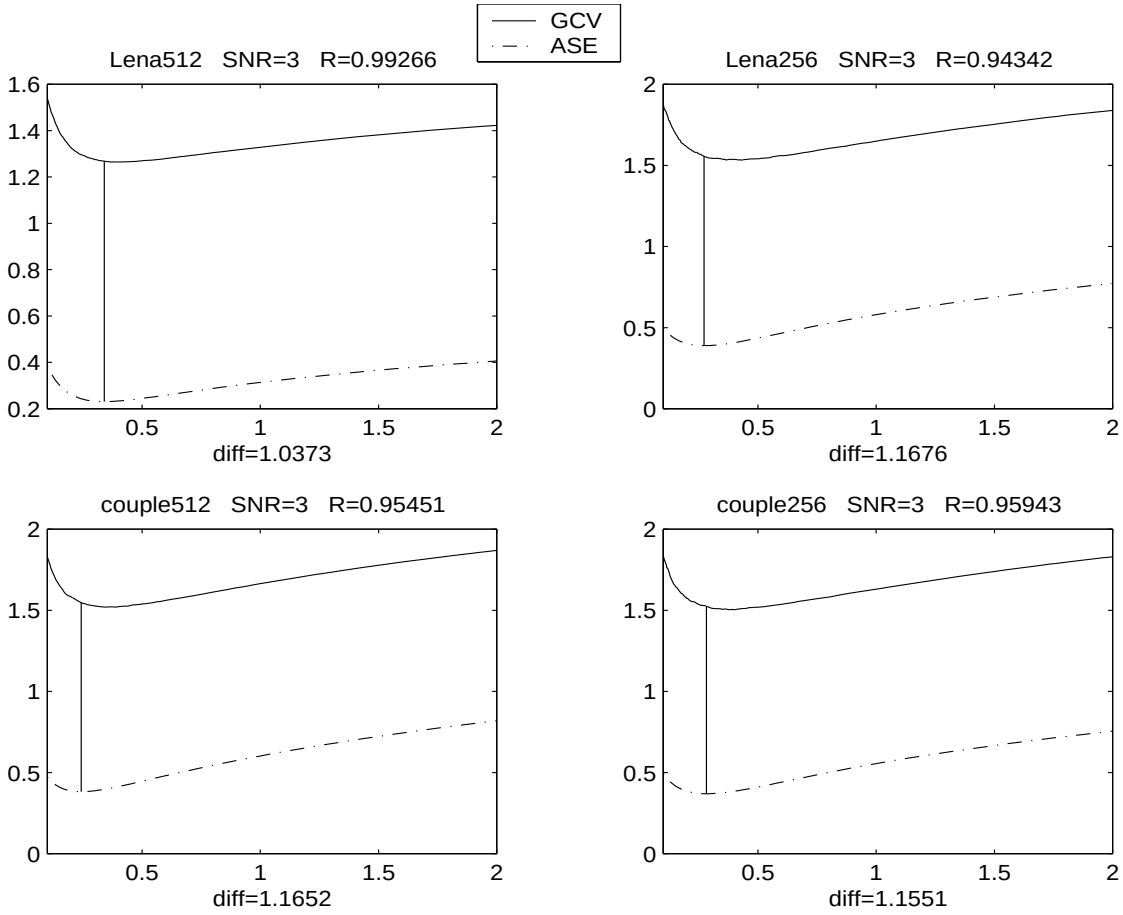

Figure 10: The GCV/σ^2 and ASE/σ^2 curves for BLUPWAVE. Top left is for Lena: 512×512, top right is for Lena: 256×256, bottom left is for couple: 512×512, and bottom right is for couple: 256×256.

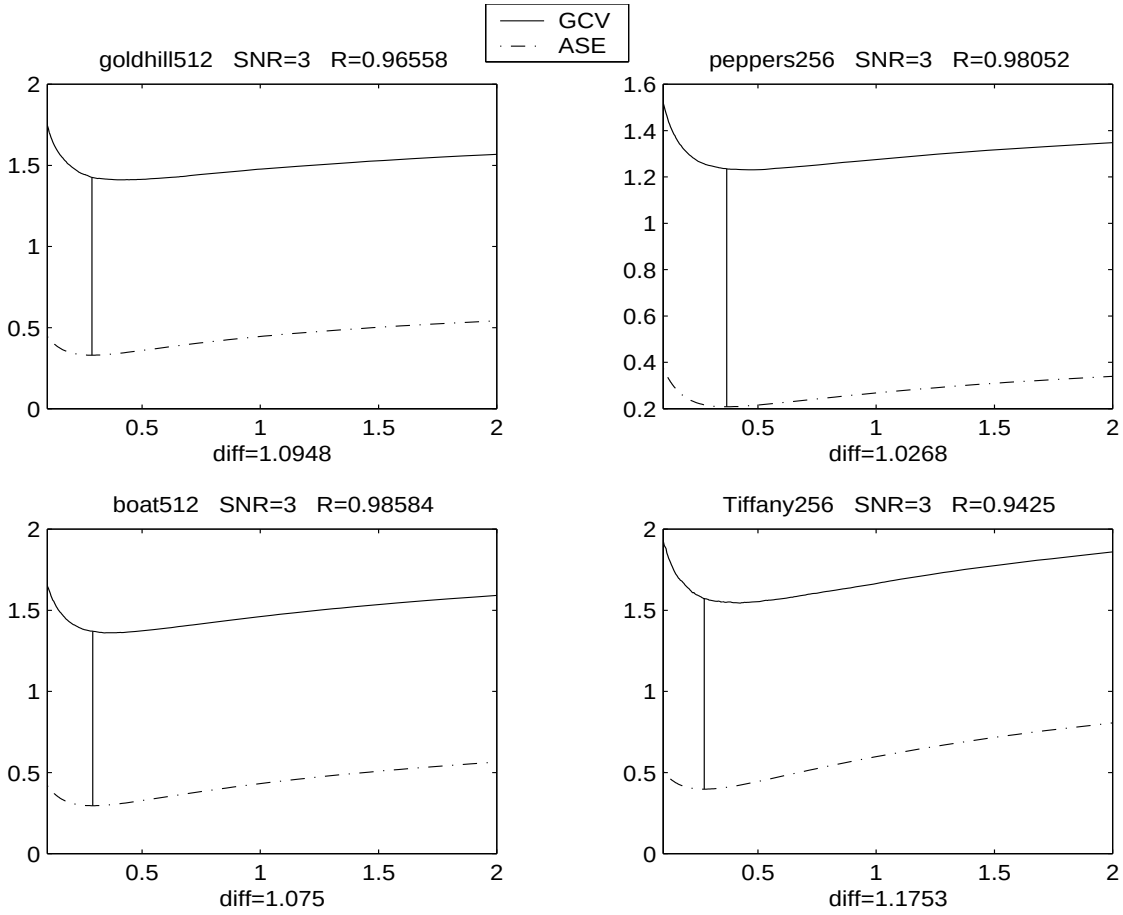

Figure 11: The GCV/σ^2 and ASE/σ^2 curves for BLUPWAVE. Top left is for goldhill: 512×512, top right is for peppers: 512×512, bottom left is for boat: 512×512, and bottom right is for Tiffany: 256×256.

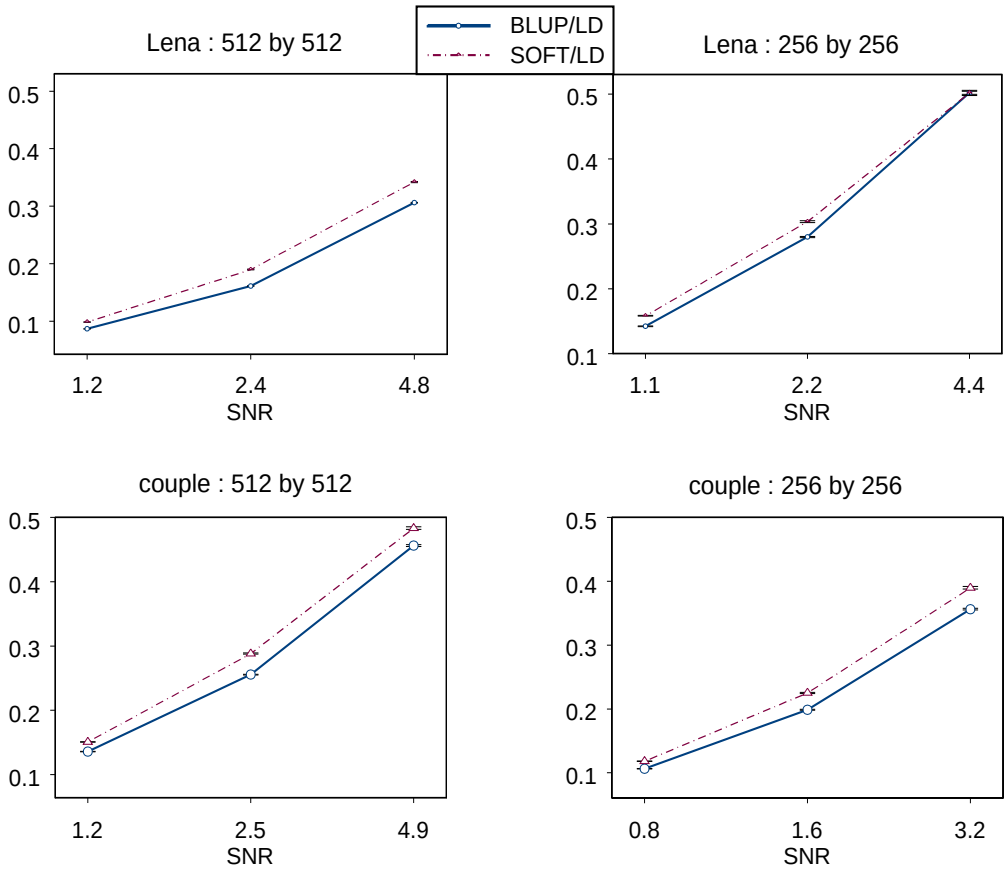

Figure 12: Comparison of the standardized ASEs of BLUPWAVE (with level dependent thresholds) and soft thresholding (with level dependent thresholds) based on 50 replications. Top left is for Lena: 512×512, top right is for Lena: 256×256, bottom left is for couple: 512×512, and bottom right is for couple: 256×256.

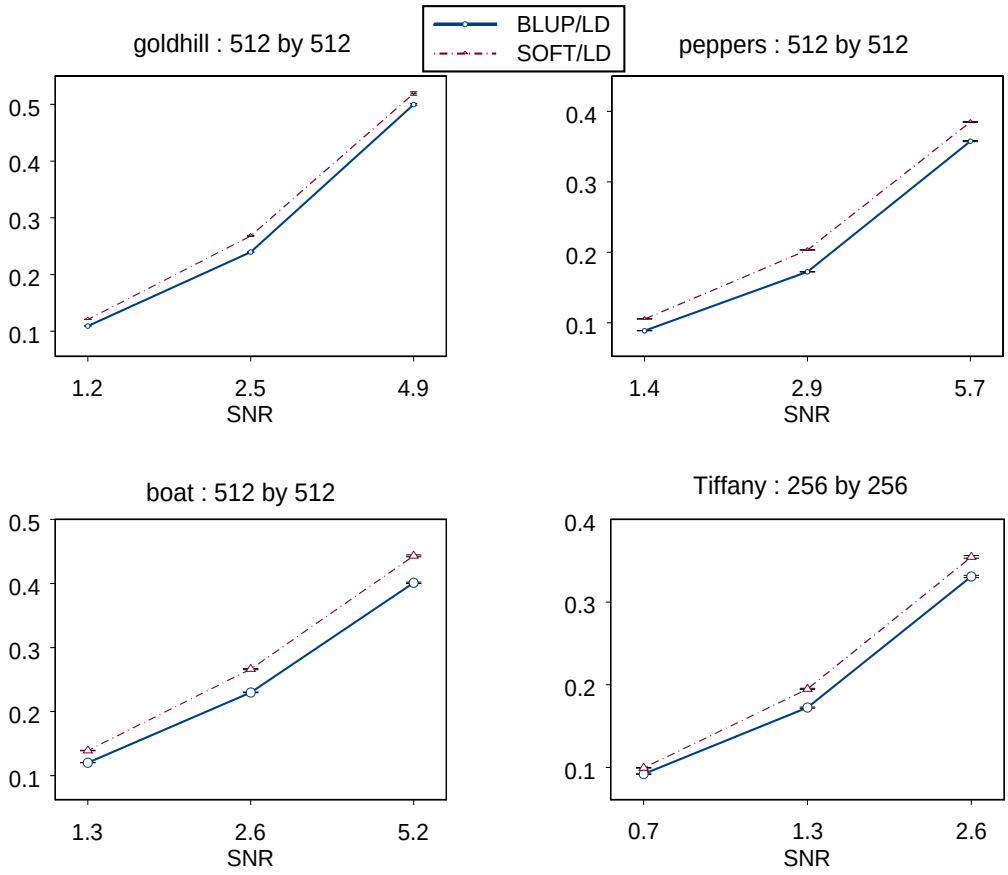

Figure 13: Comparison of the standardized ASEs for BLUPWAVE (with level dependent thresholds) and soft thresholding (with level dependent thresholds) based on 50 replications. Top left is for goldhill: 512×512, top right is for peppers: 512×512, bottom left is for boat: 512×512, and bottom right is for Tiffany: 256×256.

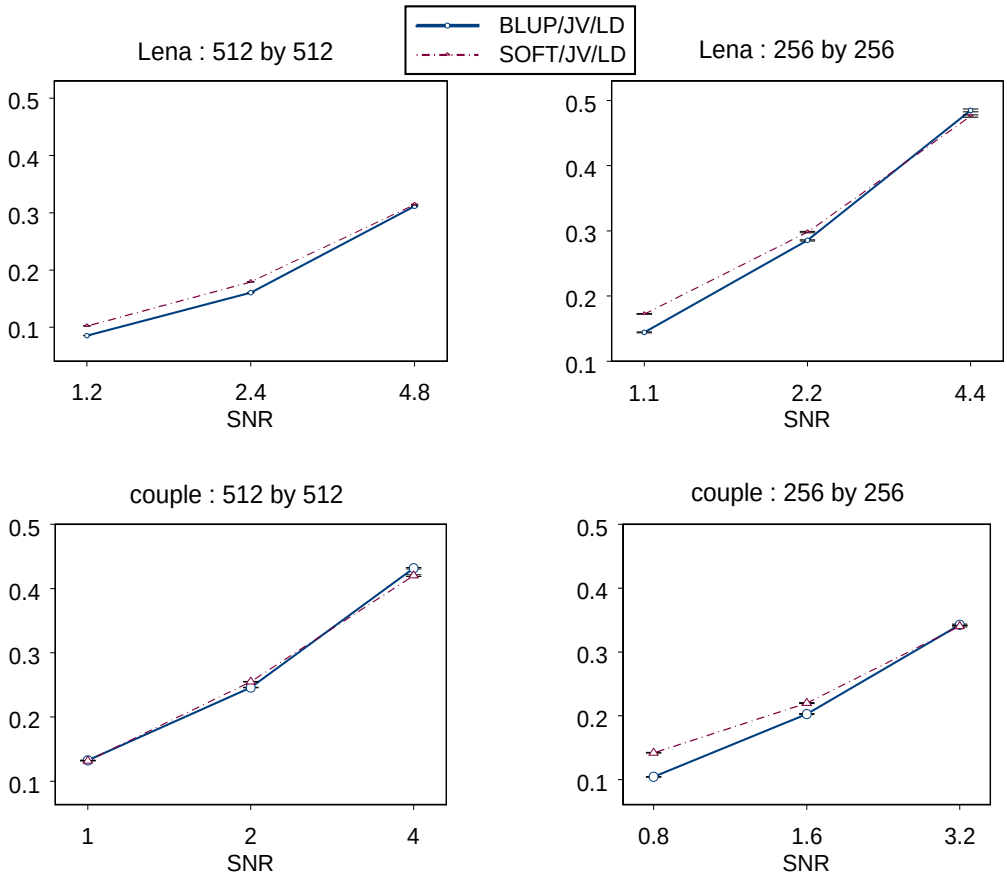

Figure 14: Comparison of the standardized ASEs for BLUP/LD/JV *vs.* SOFT/LD/JV based on 50 replications. Top left is for Lena: 512×512, top right is for Lena: 256×256, bottom left is for couple: 512×512, and bottom right is for couple: 256×256.

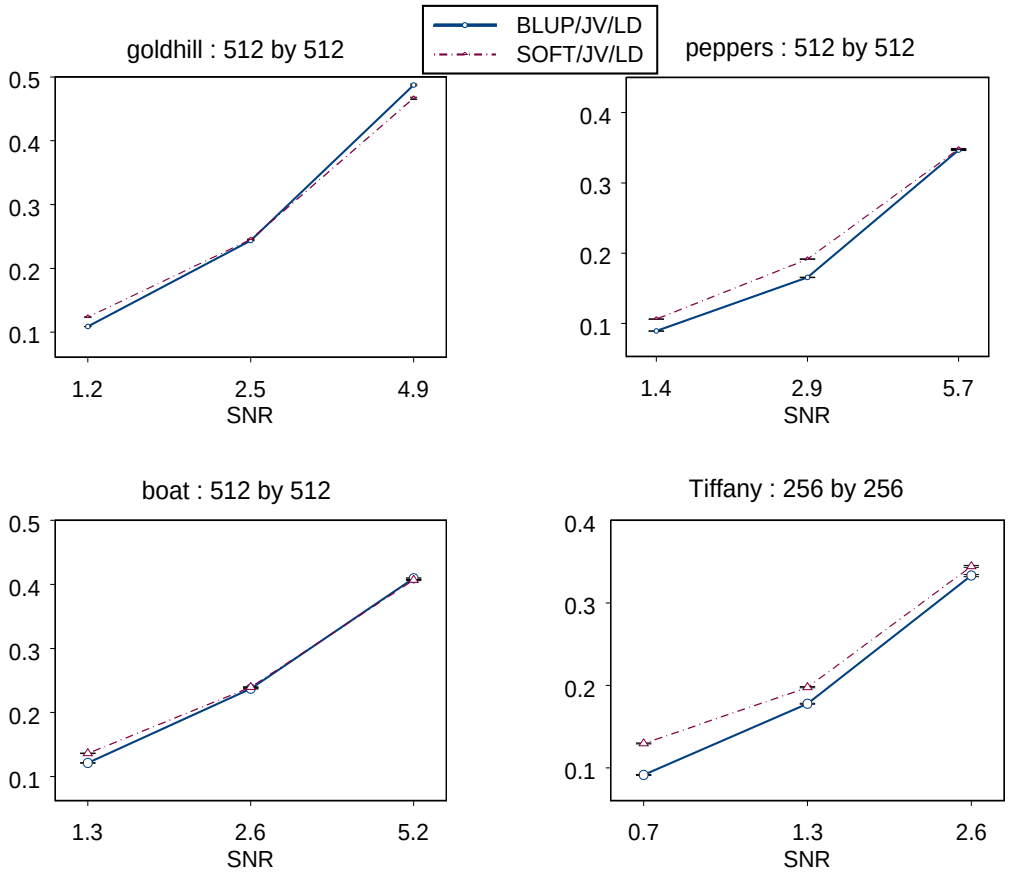

Figure 15: Comparison of the standardized ASEs for BLUP/LD/JV *vs.* SOFT/LD/JV based on 50 replications. Top left is for goldhill: 512×512, top right is for peppers: 512×512, bottom left is for boat: 512×512, and bottom right is for Tiffany: 256×256.

Table 1: Averages and standard errors of standardized ASEs based on 100 replications for BLUP-WAVE.

n	**Blocks**	**Bumps**	**HeaviSine**	**Doppler**
		SNR=3		
256	0.5793 (0.0089)	0.6369 (0.0126)	0.1654 (0.0043)	0.3815 (0.0062)
512	0.4052 (0.0061)	0.4566 (0.0063)	0.0970 (0.0020)	0.2283 (0.0042)
1024	0.2904 (0.0032)	0.3026 (0.0031)	0.0597 (0.0015)	0.1336 (0.0020)
2048	0.1837 (0.0018)	0.1950 (0.0017)	0.0382 (0.0008)	0.0792 (0.0010)
4096	0.1299 (0.0008)	0.1249 (0.0010)	0.0254 (0.0005)	0.0480 (0.0006)
8192	0.0820 (0.0005)	0.0756 (0.0005)	0.0163 (0.0003)	0.0266 (0.0003)
		SNR=5		
256	0.6408 (0.0114)	0.6569 (0.0105)	0.1989 (0.0037)	0.3946 (0.0065)
512	0.4589 (0.0056)	0.5033 (0.0063)	0.1360 (0.0027)	0.2434 (0.0039)
1024	0.3291 (0.0031)	0.3277 (0.0031)	0.0755 (0.0016)	0.1546 (0.0021)
2048	0.2142 (0.0017)	0.2233 (0.0020)	0.0488 (0.0008)	0.0917 (0.0012)
4096	0.1404 (0.0010)	0.1426 (0.0010)	0.0327 (0.0005)	0.0532 (0.0006)
8192	0.0893 (0.0006)	0.0838 (0.0006)	0.0216 (0.0003)	0.0277 (0.0003)
		SNR=7		
256	0.6803 (0.0121)	0.7032 (0.0111)	0.2375 (0.0046)	0.4043 (0.0067)
512	0.4781 (0.0062)	0.5194 (0.0061)	0.1661 (0.0029)	0.2486 (0.0038)
1024	0.3549 (0.0037)	0.3507 (0.0029)	0.0897 (0.0015)	0.1646 (0.0021)
2048	0.2309 (0.0019)	0.2446 (0.0018)	0.0606 (0.0008)	0.0999 (0.0014)
4096	0.1475 (0.0009)	0.1542 (0.0010)	0.0391 (0.0006)	0.0540 (0.0008)
8192	0.0932 (0.0006)	0.0889 (0.0006)	0.0255 (0.0003)	0.0287 (0.0003)
		SNR=10		
256	0.6730 (0.0110)	0.7783 (0.0117)	0.2724 (0.0050)	0.4182 (0.0068)
512	0.4918 (0.0053)	0.5372 (0.0064)	0.1843 (0.0033)	0.2541 (0.0038)
1024	0.3778 (0.0034)	0.3776 (0.0033)	0.1075 (0.0018)	0.1688 (0.0022)
2048	0.2436 (0.0019)	0.2588 (0.0019)	0.0736 (0.0010)	0.1043 (0.0014)
4096	0.1568 (0.0009)	0.1615 (0.0011)	0.0441 (0.0007)	0.0538 (0.0007)
8192	0.0980 (0.0006)	0.0949 (0.0006)	0.0275 (0.0003)	0.0301 (0.0003)

Table 2: Averages and standard errors of standardized ASEs based on 100 replications for soft thresholding.

n	**Blocks**	**Bumps**	**HeaviSine**	**Doppler**
		SNR=3		
256	0.5854 (0.0108)	0.7029 (0.0139)	0.1581 (0.0038)	0.4060 (0.0067)
512	0.4085 (0.0051)	0.5182 (0.0064)	0.0953 (0.0021)	0.2581 (0.0044)
1024	0.3179 (0.0037)	0.3658 (0.0034)	0.0571 (0.0013)	0.1610 (0.0022)
2048	0.2160 (0.0021)	0.2514 (0.0022)	0.0382 (0.0006)	0.1077 (0.0010)
4096	0.1554 (0.0010)	0.1717 (0.0012)	0.0264 (0.0004)	0.0680 (0.0007)
8192	0.1041 (0.0005)	0.1117 (0.0006)	0.0181 (0.0003)	0.0408 (0.0004)
		SNR=5		
256	0.6809 (0.0112)	0.7760 (0.0141)	0.1940 (0.0036)	0.4542 (0.0073)
512	0.5009 (0.0066)	0.5806 (0.0060)	0.1333 (0.0025)	0.2985 (0.0049)
1024	0.3876 (0.0044)	0.4181 (0.0038)	0.0785 (0.0015)	0.1949 (0.0025)
2048	0.2648 (0.0024)	0.2976 (0.0024)	0.0532 (0.0007)	0.1276 (0.0012)
4096	0.1865 (0.0012)	0.2017 (0.0013)	0.0370 (0.0005)	0.0792 (0.0008)
8192	0.1252 (0.0007)	0.1300 (0.0007)	0.0258 (0.0003)	0.0469 (0.0005)
		SNR=7		
256	0.7528 (0.0144)	0.8315 (0.0163)	0.2368 (0.0044)	0.4915 (0.0078)
512	0.5392 (0.0068)	0.6162 (0.0068)	0.1683 (0.0033)	0.3160 (0.0045)
1024	0.4260 (0.0041)	0.4537 (0.0038)	0.0960 (0.0012)	0.2185 (0.0036)
2048	0.2947 (0.0025)	0.3274 (0.0024)	0.0666 (0.0008)	0.1408 (0.0012)
4096	0.2055 (0.0012)	0.2203 (0.0012)	0.0454 (0.0005)	0.0859 (0.0009)
8192	0.1385 (0.0006)	0.1420 (0.0008)	0.0315 (0.0003)	0.0507 (0.0004)
		SNR=10		
256	0.7923 (0.0158)	0.9196 (0.0239)	0.2793 (0.0060)	0.5215 (0.0077)
512	0.5913 (0.0083)	0.6545 (0.0076)	0.1998 (0.0048)	0.3428 (0.0055)
1024	0.4665 (0.0043)	0.4861 (0.0042)	0.1187 (0.0021)	0.2368 (0.0038)
2048	0.3246 (0.0026)	0.3511 (0.0025)	0.0823 (0.0010)	0.1541 (0.0016)
4096	0.2271 (0.0014)	0.2376 (0.0012)	0.0541 (0.0006)	0.0918 (0.0009)
8192	0.1529 (0.0007)	0.1552 (0.0008)	0.0372 (0.0004)	0.0552 (0.0005)

Table 3: Averages and standard errors of standardized ASEs based on 100 replications for BLUP-WAVE with level dependent thresholds.

n	**Blocks**	**Bumps**	**HeaviSine**	**Doppler**
		SNR=3		
256	0.5632 (0.0081)	0.6207 (0.0102)	0.1795 (0.0052)	0.3818 (0.0064)
512	0.3880 (0.0051)	0.4520 (0.0054)	0.1045 (0.0022)	0.2268 (0.0039)
1024	0.2696 (0.0026)	0.2876 (0.0029)	0.0606 (0.0014)	0.1272 (0.0019)
2048	0.1641 (0.0017)	0.1769 (0.0016)	0.0386 (0.0009)	0.0728 (0.0010)
4096	0.1102 (0.0008)	0.1030 (0.0009)	0.0238 (0.0006)	0.0432 (0.0007)
8192	0.0663 (0.0005)	0.0593 (0.0004)	0.0144 (0.0003)	0.0220 (0.0003)
		SNR=5		
256	0.6296 (0.0102)	0.6596 (0.0104)	0.2087 (0.0044)	0.3999 (0.0066)
512	0.4371 (0.0056)	0.4977 (0.0056)	0.1389 (0.0028)	0.2465 (0.0039)
1024	0.3088 (0.0030)	0.3206 (0.0027)	0.0743 (0.0015)	0.1500 (0.0018)
2048	0.1977 (0.0017)	0.2071 (0.0018)	0.0481 (0.0009)	0.0867 (0.0013)
4096	0.1235 (0.0010)	0.1216 (0.0010)	0.0306 (0.0006)	0.0480 (0.0007)
8192	0.0775 (0.0005)	0.0679 (0.0005)	0.0197 (0.0003)	0.0237 (0.0003)
		SNR=7		
256	0.6507 (0.0110)	0.6966 (0.0101)	0.2404 (0.0050)	0.4074 (0.0073)
512	0.4562 (0.0056)	0.5153 (0.0055)	0.1639 (0.0029)	0.2537 (0.0036)
1024	0.3328 (0.0030)	0.3441 (0.0030)	0.0901 (0.0016)	0.1623 (0.0020)
2048	0.2141 (0.0018)	0.2295 (0.0019)	0.0587 (0.0011)	0.0950 (0.0013)
4096	0.1329 (0.0009)	0.1355 (0.0010)	0.0353 (0.0006)	0.0486 (0.0007)
8192	0.0822 (0.0005)	0.0733 (0.0005)	0.0229 (0.0003)	0.0251 (0.0003)
		SNR=10		
256	0.6496 (0.0089)	0.7622 (0.0107)	0.2720 (0.0052)	0.4238 (0.0070)
512	0.4748 (0.0047)	0.5285 (0.0059)	0.1800 (0.0033)	0.2599 (0.0037)
1024	0.3573 (0.0031)	0.3738 (0.0034)	0.1079 (0.0017)	0.1696 (0.0022)
2048	0.2303 (0.0017)	0.2469 (0.0018)	0.0694 (0.0011)	0.0988 (0.0014)
4096	0.1432 (0.0010)	0.1465 (0.0010)	0.0406 (0.0006)	0.0484 (0.0007)
8192	0.0874 (0.0005)	0.0807 (0.0005)	0.0256 (0.0003)	0.0265 (0.0003)

Table 4: Averages and standard errors of standardized ASEs based on 100 replications for soft thresholding with level dependent thresholds.

n	**Blocks**	**Bumps**	**HeaviSine**	**Doppler**
		SNR=3		
256	0.5495 (0.0080)	0.6701 (0.0112)	0.1639 (0.0039)	0.3987 (0.0066)
512	0.3778 (0.0047)	0.4859 (0.0055)	0.0995 (0.0024)	0.2452 (0.0043)
1024	0.2698 (0.0034)	0.3146 (0.0036)	0.0566 (0.0014)	0.1388 (0.0018)
2048	0.1672 (0.0018)	0.1946 (0.0018)	0.0363 (0.0007)	0.0838 (0.0011)
4096	0.1136 (0.0007)	0.1122 (0.0008)	0.0228 (0.0005)	0.0493 (0.0006)
8192	0.0706 (0.0004)	0.0657 (0.0004)	0.0144 (0.0002)	0.0261 (0.0003)
		SNR=5		
256	0.6438 (0.0100)	0.7436 (0.0127)	0.1985 (0.0040)	0.4508 (0.0072)
512	0.4555 (0.0053)	0.5559 (0.0055)	0.1326 (0.0025)	0.2918 (0.0045)
1024	0.3293 (0.0034)	0.3707 (0.0036)	0.0732 (0.0013)	0.1688 (0.0023)
2048	0.2145 (0.0023)	0.2346 (0.0019)	0.0486 (0.0009)	0.1028 (0.0011)
4096	0.1406 (0.0010)	0.1371 (0.0009)	0.0305 (0.0005)	0.0577 (0.0007)
8192	0.0890 (0.0005)	0.0782 (0.0005)	0.0201 (0.0003)	0.0296 (0.0003)
		SNR=7		
256	0.6986 (0.0117)	0.7988 (0.0143)	0.2358 (0.0047)	0.4815 (0.0072)
512	0.4919 (0.0059)	0.5898 (0.0063)	0.1621 (0.0033)	0.3134 (0.0044)
1024	0.3642 (0.0032)	0.4095 (0.0038)	0.0914 (0.0013)	0.1902 (0.0037)
2048	0.2408 (0.0024)	0.2621 (0.0019)	0.0596 (0.0010)	0.1147 (0.0013)
4096	0.1584 (0.0010)	0.1555 (0.0009)	0.0366 (0.0005)	0.0618 (0.0007)
8192	0.0999 (0.0005)	0.0873 (0.0005)	0.0243 (0.0003)	0.0323 (0.0003)
		SNR=10		
256	0.7396 (0.0119)	0.8758 (0.0178)	0.2777 (0.0067)	0.5139 (0.0076)
512	0.5406 (0.0058)	0.6289 (0.0075)	0.1903 (0.0047)	0.3361 (0.0051)
1024	0.3995 (0.0034)	0.4416 (0.0043)	0.1115 (0.0017)	0.2125 (0.0036)
2048	0.2653 (0.0020)	0.2887 (0.0019)	0.0715 (0.0010)	0.1262 (0.0015)
4096	0.1763 (0.0011)	0.1743 (0.0009)	0.0442 (0.0006)	0.0653 (0.0007)
8192	0.1119 (0.0005)	0.0980 (0.0005)	0.0290 (0.0004)	0.0348 (0.0003)

Table 5: The minimums of GCV/σ^2 values for various primary resolution levels (j), and the number of vanishing moments of Symmlets (v). Test data is the noisy Blocks with $n = 1024$ and SNR=7.

	the number of vanishing moments, v						
j	**4**	**5**	**6**	**7**	**8**	**9**	**10**
0	1.34384	1.36370	1.42276	1.35806	1.30824	1.45862	1.50087
1	1.34379	1.36369	1.42273	1.35442	1.30823	1.46090	1.50085
2	1.34378	1.36366	1.42272	1.35430	1.30823	1.46083	1.50084
3	1.34366	1.36357	1.42248	1.35496	*1.30802	1.46008	1.49940
4	1.34263	1.36597	1.41981	1.35107	1.31140	1.45830	1.49811
5	1.34064	1.35137	1.39775	1.34341	1.40833	1.44583	1.49015
6	1.32112	1.34952	1.37714	1.33316	1.36415	1.41922	1.46039
7	1.32823	1.38788	1.37893	1.35368	1.37244	1.47803	1.50491
8	1.50028	1.53409	1.54756	1.43989	1.45617	1.59628	1.60713
9	2.09728	2.17332	2.14222	1.96848	2.00363	2.22018	2.22672

Table 6: Averages and standard errors of standardized ASEs based on 100 replications of BLUP/LD/JV for various sample sizes and SNR=7.

n	**Blocks**	**Bumps**	**HeaviSine**	**Doppler**
256	0.5975 (0.0089)	0.6916 (0.0090)	0.1915 (0.0049)	0.3758 (0.0070)
512	0.4143 (0.0049)	0.4978 (0.0054)	0.1318 (0.0026)	0.2436 (0.0035)
1024	0.2992 (0.0027)	0.3235 (0.0033)	0.0750 (0.0015)	0.1493 (0.0023)
2048	0.1963 (0.0017)	0.2123 (0.0017)	0.0524 (0.0010)	0.0903 (0.0014)
4096	0.1176 (0.0009)	0.1167 (0.0008)	0.0337 (0.0005)	0.0440 (0.0006)
8192	0.0736 (0.0004)	0.0662 (0.0005)	0.0179 (0.0003)	0.0242 (0.0003)

Table 7: Averages and standard errors of standardized ASEs based on 100 replications of SOFT/LD/JV for various sample sizes and SNR=7.

n	**Blocks**	**Bumps**	**HeaviSine**	**Doppler**
256	0.6552 (0.0099)	0.7723 (0.0125)	0.1883 (0.0039)	0.4440 (0.0072)
512	0.4635 (0.0051)	0.5802 (0.0060)	0.1320 (0.0024)	0.3096 (0.0044)
1024	0.3357 (0.0036)	0.3770 (0.0035)	0.0819 (0.0014)	0.1860 (0.0022)
2048	0.2309 (0.0018)	0.2579 (0.0021)	0.0544 (0.0009)	0.1150 (0.0014)
4096	0.1460 (0.0011)	0.1525 (0.0009)	0.0353 (0.0004)	0.0610 (0.0007)
8192	0.0987 (0.0005)	0.0869 (0.0005)	0.0217 (0.0003)	0.0334 (0.0004)

Table 8: Averages and standard errors of standardized ASEs for BLUPWAVE and soft thresholding, based on 50 replications.

SNR	**BLUP**	**SOFT**
	Lena : 512 × 512	
4.8	0.340299 (0.000450)	0.349786 (0.000615)
2.4	0.190690 (0.000177)	0.195083 (0.000294)
1.2	0.103531 (0.000147)	0.100750 (0.000108)
	Lena : 256 × 256	
4.4	0.512849 (0.001975)	0.511180 (0.002163)
2.2	0.312310 (0.000738)	0.310408 (0.000894)
1.1	0.167269 (0.000283)	0.161216 (0.000311)
	couple : 512 × 512	
4.0	0.489354 (0.000943)	0.488529 (0.001050)
2.0	0.284726 (0.000430)	0.290842 (0.000505)
1.0	0.151385 (0.000151)	0.151266 (0.000203)
	couple : 256 × 256	
3.2	0.390904 (0.000878)	0.395889 (0.001103)
1.6	0.227962 (0.000370)	0.228002 (0.000450)
0.8	0.127576 (0.000493)	0.118677 (0.000213)
	goldhill : 512 × 512	
4.9	0.546266 (0.001329)	0.528330 (0.001549)
2.5	0.274588 (0.000271)	0.272021 (0.000403)
1.2	0.125817 (0.000122)	0.121537 (0.000119)
	peppers : 512 × 512	
5.7	0.386394 (0.000681)	0.392051 (0.000695)
2.9	0.199743 (0.000174)	0.206665 (0.000255)
1.4	0.106592 (0.000168)	0.106369 (0.000119)
	boat : 512 × 512	
5.2	0.446919 (0.001080)	0.450541 (0.000817)
2.6	0.267619 (0.000292)	0.270649 (0.000479)
1.3	0.142249 (0.000139)	0.140604 (0.000161)
	Tiffany : 256 × 256	
2.6	0.365157 (0.001003)	0.364561 (0.000942)
1.3	0.202361 (0.000328)	0.200073 (0.000448)
0.7	0.110993 (0.000483)	0.100582 (0.000206)

Table 9: Averages and standard errors of standardized ASEs for BLUPWAVE and soft thresholding with level dependent thresholds, based on 50 replications.

SNR	**BLUP**	**SOFT**
	Lena : 512 × 512	
4.8	0.306159 (0.000320)	0.342088 (0.000532)
2.4	0.161366 (0.000174)	0.190043 (0.000244)
1.2	0.086905 (0.000103)	0.098783 (0.000103)
	Lena : 256 × 256	
4.4	0.502055 (0.001207)	0.501758 (0.001958)
2.2	0.279845 (0.000513)	0.303544 (0.000779)
1.1	0.142318 (0.000256)	0.158422 (0.000283)
	couple : 512 × 512	
4.0	0.456290 (0.000789)	0.483463 (0.000985)
2.0	0.255461 (0.000243)	0.288218 (0.000467)
1.0	0.135795 (0.000148)	0.150638 (0.000195)
	couple : 256 × 256	
3.2	0.356370 (0.000617)	0.389875 (0.000988)
1.6	0.198796 (0.000365)	0.225111 (0.000404)
0.8	0.106527 (0.000225)	0.118094 (0.000218)
	goldhill : 512 × 512	
4.9	0.499870 (0.000789)	0.519145 (0.001434)
2.5	0.239473 (0.000217)	0.268232 (0.000366)
1.2	0.109399 (0.000098)	0.120945 (0.000120)
	peppers : 512 × 512	
5.7	0.357637 (0.000400)	0.384949 (0.000614)
2.9	0.172366 (0.000143)	0.203351 (0.000227)
1.4	0.088847 (0.000098)	0.105482 (0.000117)
	boat : 512 × 512	
5.2	0.400942 (0.000550)	0.443362 (0.000743)
2.6	0.229787 (0.000284)	0.266316 (0.000424)
1.3	0.120199 (0.000120)	0.139164 (0.000144)
	Tiffany : 256 × 256	
2.6	0.330946 (0.000707)	0.354760 (0.000825)
1.3	0.172339 (0.000321)	0.194962 (0.000384)
0.7	0.091885 (0.000215)	0.099395 (0.000216)

Table 10: Averages and standard errors of standardized ASEs for BLUP/LD/JV and SOFT/LD/JV, based on 50 replications.

SNR	**BLUP**	**SOFT**
	Lena : 512 × 512	
4.8	0.306159 (0.000320)	0.342088 (0.000532)
2.4	0.161366 (0.000174)	0.190043 (0.000244)
1.2	0.086905 (0.000103)	0.098783 (0.000103)
	Lena : 256 × 256	
4.4	0.502055 (0.001207)	0.501758 (0.001958)
2.2	0.279845 (0.000513)	0.303544 (0.000779)
1.1	0.142318 (0.000256)	0.158422 (0.000283)
	couple : 512 × 512	
4.0	0.456290 (0.000789)	0.483463 (0.000985)
2.0	0.255461 (0.000243)	0.288218 (0.000467)
1.0	0.135795 (0.000148)	0.150638 (0.000195)
	couple : 256 × 256	
3.2	0.356370 (0.000617)	0.389875 (0.000988)
1.6	0.198796 (0.000365)	0.225111 (0.000404)
0.8	0.106527 (0.000225)	0.118094 (0.000218)
	goldhill : 512 × 512	
4.9	0.499870 (0.000789)	0.519145 (0.001434)
2.5	0.239473 (0.000217)	0.268232 (0.000366)
1.2	0.109399 (0.000098)	0.120945 (0.000120)
	peppers : 512 × 512	
5.7	0.357637 (0.000400)	0.384949 (0.000614)
2.9	0.172366 (0.000143)	0.203351 (0.000227)
1.4	0.088847 (0.000098)	0.105482 (0.000117)
	boat : 512 × 512	
5.2	0.400942 (0.000550)	0.443362 (0.000743)
2.6	0.229787 (0.000284)	0.266316 (0.000424)
1.3	0.120199 (0.000120)	0.139164 (0.000144)
	Tiffany : 256 × 256	
2.6	0.330946 (0.000707)	0.354760 (0.000825)
1.3	0.172339 (0.000321)	0.194962 (0.000384)
0.7	0.091885 (0.000215)	0.099395 (0.000216)