# ON OPTIMAL THRESHOLD SELECTION FOR MULTIWAVELET SHRINKAGE

*Tai-Chiu Hsung and Daniel Pak-Kong Lun*

Centre for Multimedia Signal Processing
Department of Electronic and Information Engineering
The Hong Kong Polytechnic University

## ABSTRACT

Recent researches found that multivariate shrinkage on multiwavelet transform coefficients further improves the traditional wavelet methods. It is because multiwavelet transform, with appropriate initialization, provides better representation of signals so that their difference from noise can be clearly identified. In this paper, we consider the optimal threshold selection for multiwavelet denoising by using multivariate shrinkage function. Firstly, we study the threshold selection using the Stein's unbiased risk estimator (SURE) for each resolution level when the noise structure is given. Then, we consider the method of generalized cross validation (GCV) when the noise structure is not known a priori. Simulation results show that the higher multiplicity (>2) wavelets usually give better denoising results. Besides, the proposed threshold estimators often suggest better thresholds as compared with the traditional estimators.

## 1. INTRODUCTION

Consider estimating an unknown deterministic discrete signal $f_i$ from noisy observation $g_i$,

$$g_i = f_i + \varepsilon_i, \qquad i=1,\cdots,n \tag{1}$$

where $\boldsymbol{\varepsilon}=(\ldots,\varepsilon_i,\ldots)$ is independent and identically distributed (*iid*) $N_n(\mathbf{0},\sigma^2\mathbf{I})$ noise. The goal of denoising is to minimize the mean square error (MSE),

$$\frac{1}{n}\left\|\hat{\boldsymbol{f}}-\boldsymbol{f}\right\|^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_i - f_i\right)^2 \tag{2}$$

subject to the condition that the estimated signal $\hat{\boldsymbol{f}}$ is at least as smooth as $\boldsymbol{f}$. It is found that multiwavelet denoising using multivariate shrinkage [1] gives consistently better results than the traditional wavelet shrinkage denoising. The improvement is contributed by the multiwavelet transform, which gives better signal representation so that noise and signal can be separated much easier. Besides, the multivariate shrinkage operator effectively exploits the statistical information of the transform coefficient vectors of noise that improves the denoising performance. Denote the vector filter bank for $J$ levels discrete multiwavelet transform of multiplicity $L$ as $\boldsymbol{W}$ and the prefilter as $\boldsymbol{Q}$, then the discrete multiwavelet transform for scalar signals is $\boldsymbol{M}=\boldsymbol{WQ}$. Hence $\boldsymbol{w}=\boldsymbol{Mg}=\boldsymbol{v}+\boldsymbol{\omega}$, where $\boldsymbol{v}=\boldsymbol{Mf}$ and $\boldsymbol{\omega}=\boldsymbol{M\varepsilon}$. The matrix $\boldsymbol{M}$ is designed such that the $J$ levels output transform coefficient vectors are arranged into an $n\times 1$ vector, i.e. $\boldsymbol{w}=(\ldots,\ \boldsymbol{w}_{j,k}{}^T,\ldots)^T$ where $\boldsymbol{w}_{j,k}$ is $k^{th}$ transform coefficient vector of $\boldsymbol{g}$ at scale $j$, each vector contains $L$ elements. It is shown in [2] that, by adaptively applying different multivariate shrinkage operation to coefficient vectors at different resolution levels, improved denoising performance can be achieved. Let us express the level-dependent multivariate shrinkage as $\boldsymbol{w}_{j,\delta}=\boldsymbol{D}_{j,\delta}\cdot\boldsymbol{w}_j$, where $\boldsymbol{D}_{j,\delta}=diag[\eta_{\delta_j,k}]$ for $k=1,\ldots,n_j$ where $\boldsymbol{w}_j$ is the level $j$ transform coefficient vectors of the observation, $\boldsymbol{w}_{j,\delta}$ is the shrunk coefficient vectors, and $n_j$ is the number of transform coefficient vectors at level $j$. Define $\eta_{\delta_j,k}$ [2] as,

$$\eta_{\delta_j,k}=\boldsymbol{0} \quad \text{for} \quad \vartheta_j<\lambda_j^2, \qquad \eta_{\delta_j,k}=1-\lambda_j^2/\vartheta_j \quad \text{for} \quad \vartheta_j\geq\lambda_j^2 \tag{3}$$

where $$\vartheta_j=\boldsymbol{w}_{j,k}^T\boldsymbol{V}_j^{-1}\boldsymbol{w}_{j,k} \tag{4}$$

and $\delta_j$ denotes the shrinkage parameters $\{\lambda_j,\boldsymbol{V}_j\}$. $\lambda_j$ is the threshold for the shrinkage operation performed at level $j$, and $\boldsymbol{V}_j$ is the covariance matrix of noise at level $j$, where $j\in Z_J$. If we are given covariance of $\boldsymbol{\varepsilon}$, say $N_n(\boldsymbol{\mu},\sigma^2\boldsymbol{\Sigma})$, we can compute the distribution parameters of the transform noise coefficient vector: $\boldsymbol{\mu}_j=\boldsymbol{M}_j\boldsymbol{\mu}$ and $\boldsymbol{V}_j=\boldsymbol{M}_j\boldsymbol{\Sigma}\boldsymbol{M}_j^T$.

The selection of the threshold values is critical to the performance of multiwavelet shrinkage. Since the multivariate shrinkage function is different from that used in wavelet shrinkage, we cannot borrow the risk estimators suggested for scalar wavelet shrinkage to the selection of the parameter set $\delta$. In this paper, we study two risk estimators for finding optimal threshold. We first study the approach based on Stein's unbiased risk estimator (SURE) for each resolution levels, namely LSURE [2]. Then, we further consider the generalized cross validation method [4][5] when the noise structure is not known (LGCV). Simulation results verify that the resulted risk estimators give better indication on threshold selection as compared to the traditional SURE and GCV estimators. Improved denoising performance is then achieved particularly for higher multiplicity multiwavelet shrinkage.

## 2. SURE FOR EACH RESOLUTION LEVEL

Let us consider the residual error [5],

$$T(\delta)=\tfrac{1}{n}\left(\|\boldsymbol{g}_\delta-\boldsymbol{g}\|^2\right)=\tfrac{1}{n}\|(\boldsymbol{w}_\delta-\boldsymbol{v})+(\boldsymbol{v}-\boldsymbol{w})\|^2 \quad (5)$$

where $\boldsymbol{g}_\delta$ be the denoised observation $\boldsymbol{g}$ with parameter $\delta$, $\boldsymbol{w}_\delta$ be the transform coefficient vector $\boldsymbol{w}$ of $\boldsymbol{g}$ after multivariate shrinkage. Note that $\boldsymbol{g}$, $\boldsymbol{w}$, and $\boldsymbol{v}$ are vectors such that $\boldsymbol{g}=(\ldots,g_i,\ldots)^T$; $\boldsymbol{w}=(\ldots,\boldsymbol{w}_{j,k}{}^T,\ldots)^T$ and $\boldsymbol{v}=(\ldots,\boldsymbol{v}_{j,k}{}^T,\ldots)^T$. The risk $E(R(\delta))$ can be obtained as follows:

$$E\big(R(\delta)\big)=E\big(T(\delta)\big)-Tr(\boldsymbol{V})+\tfrac{2}{n}E\big(\langle\boldsymbol{\omega},\boldsymbol{z}_\delta\rangle\big) \quad (6)$$

where $E(\boldsymbol{A})$ stands for the expectation of $\boldsymbol{A}$, and $Tr(\boldsymbol{A})$ denotes the trace of matrix $\boldsymbol{A}$. $\boldsymbol{z}_\delta=\boldsymbol{w}_\delta-\boldsymbol{v}_\delta$ is the difference between the shrunk observations and the shrunk true values. Eqn.(6) formulates the evaluation of the risk function using multivariate shrinkage with parameter set $\delta$. The first term on the right hand side of eqn.(6) can be obtained from observations. Assume that the noise covariance structure is given, $\boldsymbol{V}$ can be obtained using the approach as described above hence the second term is known. The remaining unknown is the last term. Consider,

$$E(\langle\boldsymbol{\omega},\boldsymbol{z}_\delta\rangle)=E\left(\sum_{j,k}\langle\boldsymbol{\omega}_{j,k},\eta_\delta(\boldsymbol{z}_{j,k})\rangle\right)=\sum_{j,k}E\big(\langle\boldsymbol{\omega}_{j,k},\eta_\delta(\boldsymbol{z}_{j,k})\rangle\big) \quad (7)$$

It is equivalent to summing up the expectation of the inner products of $\boldsymbol{\omega}_{j,k}$ and $\eta_\delta(\boldsymbol{z}_{j,k})$ at different scales. The noise vectors are distributed as multivariate normal $\boldsymbol{\omega}_{j,k}\sim N_L(\boldsymbol{0},\sigma^2\boldsymbol{V}_j)$. With an abuse of notation, these expectations of inner products without the index $k$ can be shown to be given by the following lemma.

***Lemma 1:*** *For a multivariate normal distributed $\boldsymbol{\omega}$, $\boldsymbol{\omega}\sim N_L(\boldsymbol{0},\boldsymbol{V})$, and $\boldsymbol{z}_\delta=\boldsymbol{w}_\delta-\boldsymbol{v}_\delta=(\boldsymbol{\omega}+\boldsymbol{v})_\delta-\boldsymbol{v}_\delta$*

$$E(\langle\boldsymbol{\omega},\eta_\delta(\boldsymbol{z})\rangle)=\int_{-\infty}^{\infty}\boldsymbol{\omega}^T\eta_\delta(\boldsymbol{z})h(\boldsymbol{\omega})d\boldsymbol{\omega}$$

$$=Tr(\boldsymbol{V})P\big(\vartheta\ge\lambda^2\big)-Tr(\boldsymbol{V})E\big(\lambda^2/\vartheta\,\big|\,\vartheta\ge\lambda^2\big)+2E\big(\lambda^2/\vartheta^2\,\boldsymbol{w}^T\boldsymbol{w}\,\big|\,\vartheta\ge\lambda^2\big) \quad (8)$$

*where $\vartheta=\boldsymbol{w}^T\boldsymbol{V}^{-1}\boldsymbol{w}$, $h(\boldsymbol{\omega})=K\exp\left(\tfrac{-1}{2}\boldsymbol{\omega}^T\boldsymbol{V}^{-1}\boldsymbol{\omega}\right)$, $K$ is normalization constant and P denotes the probability.* □

It can be shown that if the covariance of noise is $\boldsymbol{I}$, the last two terms of eqn.(8) will reduce to $(2-L)E\big((\lambda^2/\vartheta)\big|\vartheta\ge\lambda^2\big)$, and becomes zero for the case of multiplicity $L$ equals to 2. Then the effect of reducing noise by multivariate shrinkage is equivalent to that of traditional shrinkage on each components of coefficient vector independently. By substituting eqn.(8) to eqn.(6), we can obtain the risk function for a particular parameter set $\delta$. For practical implementation, eqn.(7) and eqn.(8) can be approximated as in eqn.(9). For a particular level $j$, with the knowledge of the noise covariance in the form $\sigma^2\boldsymbol{\Sigma}$,

$$LSURE_j(\delta)=T_j(\delta)-\sigma^2Tr(\boldsymbol{V}_j)+\frac{2\sigma^2}{n_j}\left(Tr(\boldsymbol{V}_j)\#\{k|\vartheta_{j,k}\ge\lambda^2\}-\sum_{k=1}^{n_j}\left(Tr(\boldsymbol{V}_j)\lambda^2/\vartheta_{j,k}-(2\lambda^2/\vartheta_{j,k}^2\,\boldsymbol{w}_{j,k}^T\boldsymbol{w}_{j,k})\right)\right) \quad (9)$$

where $\vartheta_{j,k}=\boldsymbol{w}_{j,k}^T\boldsymbol{V}_j^{-1}\boldsymbol{w}_{j,k}$, $n_j$ is the number of coefficient vectors at level $j$. The operator $\#\{k|\vartheta_{j,k}\ge\lambda^2\}$ counts the number of non-zero coefficient vectors after shrinkage, which approximates the first term of eqn.(8). The optimal threshold $\delta$ can then be estimated by minimizing eqn.(9).

## 3. GENERALIZED CROSS VALIDATION

The above approach required a priori knowledge of noise structure that may not be obtained in some practical situation. To solve the problem, the Generalized Cross Validation (GCV) method is considered. The method of GCV has been applied in the derivation of a risk estimator for wavelet shrinkage [5] as follows:

$$GCV(\delta)=\frac{\frac{1}{n}\|\boldsymbol{w}-\boldsymbol{w}_\delta\|^2}{\left[\frac{Tr(\boldsymbol{I}-\boldsymbol{A}'_\delta)}{n}\right]^2} \quad (10)$$

where $n$ is the total number of signal samples. $\boldsymbol{A}'_\delta$ is the so-called derivative influence matrix such that $\boldsymbol{g}_\delta=\boldsymbol{A}_\delta\cdot\boldsymbol{g}$. It was proved that, by minimizing the GCV score, the threshold $\delta$ thus obtained approaches optimum asymptotically [5]. The same idea can be applied to the level-dependent multivariate shrinkage. Let us define the GCV function for multivariate shrinkage at level $j$ as,

$$LGCV_j(\delta)=\frac{\frac{1}{n_j}\|\boldsymbol{w}_j-\boldsymbol{w}_{j,\delta}\|^2}{\left[\frac{Tr(\boldsymbol{I}-\boldsymbol{A}'_{j,\delta})}{n_j}\right]^2} \quad (11)$$

The numerator in eqn.(11) is the mean square norm of the difference between the shrunk and the original transform coefficient vectors. To evaluate the denominator of eqn.(11), we need to compute the trace of Jacobian of the influence matrix $\boldsymbol{A}_{j,\delta}$ for each resolution level $j$. As indicated in [5], it can be shown that $Tr(\boldsymbol{A}'_{j,\delta})=Tr(\boldsymbol{D}'_{j,\delta})$. So let us consider the computation of $Tr(\boldsymbol{D}'_{j,\delta})$ at a particular level $j$. To simplify the elaboration, we skip the index $j$ in the following formulations. First, we compute the partial derivatives of the shrinkage function eqn.(4) on $\mathbf{y}$, where $\boldsymbol{y}=\boldsymbol{w}_k$, i.e.

$$\frac{\partial\vartheta}{\partial y_i}=\boldsymbol{J}_i^T\boldsymbol{R}\boldsymbol{y}+\boldsymbol{y}^T\boldsymbol{R}\boldsymbol{J}_i\cdot \quad (12)$$

where $\boldsymbol{R}=\boldsymbol{V}^{-1}$, $\boldsymbol{J}_i$ is an $(L\times 1)$ vector with all elements zero except the $i^{th}$ element: $\boldsymbol{J}_i=(0,\ldots,\underset{i}{1},\ldots,0)^T$. For $\vartheta<\lambda^2$, the shrink vector $\boldsymbol{y}_\delta=\boldsymbol{0}$, then the partial derivative of the $i^{th}$ element of $\boldsymbol{y}_\delta$ w.r.t. $y_i$ equal zero for $i=1,\ldots,L$. On the other hand for $\vartheta\ge\lambda^2$, It can be shown that,

$$\sum_{i=1}^{L}\frac{\partial[\boldsymbol{y}_\delta]_i}{\partial y_i}=L-\frac{(L-2)\lambda^2}{\vartheta}\cdot \quad (13)$$

Therefore, $Tr(\boldsymbol{D}'_{j,\delta})$, the sum of all elements of $\boldsymbol{D}'_{j,\delta}$ that are above the threshold, is given by,

$$Tr(\mathbf{A}'_{j,\delta}) = Tr(\mathbf{D}'_{j,\delta}) = L\cdot\#\{k|\vartheta_{j,k} > \lambda_j^2\} - (L-2)\sum_k \left(\lambda_j^2/\vartheta_{j,k} | \vartheta_{j,k} > \lambda_j^2\right). \quad (14)$$

where $k = 1, \ldots, n_j$. It is seen in the formulation above that the quantity $Tr(\mathbf{A}'_{j,\delta})$ counts the number of coefficient vectors that are not shrunk to zero and minus a term that is proportional to the average shrinking fraction $\lambda_j^2/\vartheta_{j,k}$, which must be obtained in performing the shrinkage. It should be noted that the shrinkage fraction is only required for multiwavelet of higher multiplicity. For bi-variate shrinkage ($L=2$), the GCV cost function eqn.(11) is equivalent to that used in wavelet shrinkage. Or in other words, the traditional GCV function can be used for bi-variate shrinkage but not for multivariate shrinkage with $L>2$. It will be verified experimentally in the next section.

## 4. SIMULATIONS

In this section, the performance of the proposed risk estimators for finding optimal threshold is investigated. To verify our findings, we use DGH wavelet with multiplicity 4 and 5 [6][7]. To keep the simulations simple, the prefiltering is orthogonal and non-decimating whereas discrete multiwavelet filter bank is decimating. We use symmetrical extensions on the signal, if necessary, to handle the boundary problem. Simulation results were obtained by averaging the results from 100 trails of multiwavelet denoising on the test signals contaminated with *iid* noise with RNR=7 (RNR=$\sqrt{\text{var}(f)/\sigma^2}$). The test signals include "Blocks" and "HypChirps" which are used in [9]. We measure the performance of different denoising algorithms for each resolution level $j$ in terms of mean square error $SE_j = \sum_{j,k}\left(w_{j,k} - \hat{w}_{j,k}\right)^2$. In the simulations, we test the following risk estimators:

1. SURE estimator borrowed from wavelet shrinkage (*bSURE*) [5]
$$bSURE_j(\delta) = T_j(\delta) - \sigma^2 Tr(\mathbf{V}_j) + \frac{2\sigma^2}{n_j}\left(Tr(\mathbf{V}_j)\#\{k|\vartheta_{j,k} \ge \lambda^2\}\right);$$
2. Proposed $LSURE_j$, in eqn.(9).
3. GCV estimator borrowed from wavelet shrinkage,
$$Tr(\mathbf{A}'_j) = L\cdot\#\{k \mid \vartheta_{j,k} > \lambda_j^2\}; \quad (15)$$
$$bGCV_j(\delta) = \frac{\frac{1}{n_j}\left\|\mathbf{w}_j - \mathbf{w}_{j,\delta}\right\|^2}{\left[\frac{Tr(\mathbf{I}-\mathbf{A}'_j)}{n_j}\right]^2}; \quad (16)$$
4. Proposed $LGCV_j$, in eqn.(11).

Denote the settings for the multiwavelet denoising using DGH wavelet of multiplicity 4 and 5 and the corresponding second order orthogonal prefilters as DGHO4 and DGHO5, respectively. The setting for the traditional bi-variate shrinkage using DGH wavelet of multiplicity 2 and the corresponding prefilter [7] is denoted as GHMXIA. We show the optimal denoising result for "HypChirps" by using several multiwavelet settings in Table 1. We can see that higher multiplicity multiwavelets usually give better performance for "HypChirps". In Table 2 and Table 3, we show the performance of various risk estimators for "HypChirps" and "Blocks", with the filter setting DGHO4, RNR=7, signal sample length n=$2^{13}$ respectively. LSURE and LGCV give the most accurate estimation of the optimal threshold whereas the traditional estimators do not perform satisfactorily in the case of higher multiplicities. We do not show the figures for bSURE because optimal threshold can hardly be derived from the estimator function. Figure 1 and Figure 2 show the square error function versus various risk estimators (LSURE, LGCV, bSURE, bGCV) for multivariate shrinkage on level 1 and 2 transform coefficients of "HypChirps", with the filter setting DGHO4, RNR=7, signal sample length n=$2^{13}$ respectively. We can see that LSURE closely resembles the square error function. LGCV also gives good estimation to the square error function but with a bias. However, it also provides good indication on optimal threshold. On the other hand, the risk estimators borrowed from traditional wavelet shrinkage do not give good estimation to the square error functions and hence the optimal threshold.

## 5. SUMMARY

In this paper, we have studied two level dependent risk estimators for finding optimal threshold using in multiwavelet shrinkage of any multiplicity. Simulations show that they closely resemble the square error functions and lead to good indication to the optimal threshold.


## ACKNOWLEDGEMENT

The work described in this paper was fully supported by a grant from the Research Grant Council of the Hong Kong Special Administrative Region, China and a grant from the Hong Kong Polytechnic University (Project no.: A418).


## REFERENCES


[1] T.R. Downie and B.W. Silverman, "The Discrete Multiple Wavelet Transform and Thresholding Methods", *IEEE Trans. Sig. Proc.*, **46**(9), Sept. 1998, pp.2558-2561.

[2] T.C. Hsung and D.P.K. Lun, "Optimal Thresholds for Multiwavelet Shrinkage", *Electronics Letters*, **39**(5), 6th Mar. 2003, pp.473-474.

[3] E. Bala and A. Ertuzun, "Applications of Multiwavelet Techniques to Image Denoising", Proceedings, *IEEE International conference on Image Processing*, New York, USA, vol.3, pp.581-584, Sept. 22-25, 2002.

[4] G. Wahba, Spline Models for Observational Data, Philadelphia, Pa.: SIAM, 1990.

[5] M. Jansen, M. Malfait and A. Bultheel, "Generalized Cross Validation for wavelet thresholding", *Signal Processing*, **56**(1), Jan. 1997, pp.33-44.

[6] G. C. Donovan, J.S. Geronimo and Douglas P. Hardin, "Orthogonal polynomials and the construction of piecewise polynomial smooth wavelets", *SIAM Journal on Mathematical Analysis*, **30**(5), 1998, pp.1029-1056.
[7] X.G. Xia, "A new prefilter design for discrete multiwavelet transforms", *IEEE Trans. Sig. Proc.*, **46**(6), Jun. 1998, pp.1558-1570.
[8] C. Stein, "Estimation of the mean of a multivariate normal distribution", *Annals of Stat.*, **9**(6), 1981, pp.1135-1151.
[9] D.L. Donoho and I.M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage", *Biometrika*, vol.81, 1994, pp.425-455.

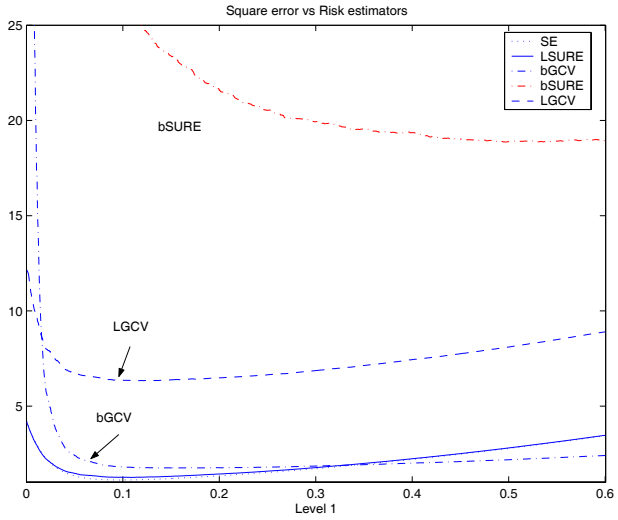


Figure 1: Square error function vs various risk estimators (LSURE, LGCV, bSURE, bGCV) for multivariate shrinkage on level 1 transform coefficients of "HypChirps", with DGHO4, RNR=7, signal sample length n=$2^{13}$.

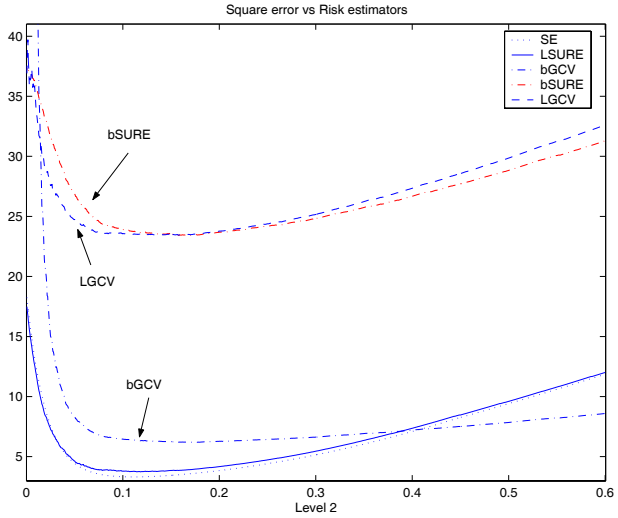


Figure 2: Square error function vs various risk estimators (LSURE, LGCV, bSURE, bGCV) for multivariate shrinkage on level 2 transform coefficients of "HypChirps", with DGHO4, RNR=7, signal sample length n=$2^{13}$.

| X-R | GHMXIA | DGHO4 | DGHO5 |
|---|---|---|---|
| 14-10 | 30.762130 | 21.402257 | **18.899297** |
| 14-7 | 58.585198 | 41.005348 | **35.790809** |
| 14-5 | 107.177545 | 75.335013 | **65.582380** |
| 13-10 | 15.392905 | 10.769218 | **9.433921** |
| 13-7 | 29.314460 | 20.574400 | **17.882119** |
| 13-5 | 53.729335 | 37.766502 | **32.778420** |
| 12-10 | 7.704642 | 5.425909 | **4.710848** |
| 12-7 | 14.683652 | 10.375956 | **8.982693** |
| 12-5 | 26.800919 | 19.009133 | **16.474120** |
| 11-10 | 3.827422 | 2.625961 | **2.383254** |
| 11-7 | 7.290475 | 5.019629 | **4.512822** |
| 11-5 | 13.514009 | 9.242358 | **8.226527** |

Table 1 - Square error with optimal thresholds when applying to the denoising of "HypChirps". X and R denote the signal length $2^X$ and the RNR value respectively.

| Level | Optimal Threshold | $E((\lambda_{estimated}-\lambda_{Optimal})^2)$ | | |
|---|---|---|---|---|
| | | LSURE | bGCV | LGCV |
| 1 | 0.102291 | **0.000111** | 0.001834 | 0.000277 |
| 2 | 0.112112 | **0.000251** | 0.001090 | 0.000310 |
| 3 | 0.061978 | **0.000054** | 0.001183 | 0.000086 |
| 4 | 0.057050 | **0.000078** | 0.001535 | 0.000153 |
| 5 | 0.054908 | **0.000110** | 0.000759 | 0.000277 |
| Level | Optimal SE | $E(SE_j)$ | | |
| | | LSURE | bGCV | LGCV |
| 1 | 1.129816 | **1.134188** | 1.175138 | 1.138291 |
| 2 | 3.161966 | 3.191734 | 3.248009 | **3.197312** |
| 3 | 5.400037 | **5.450049** | 6.060444 | 5.474825 |
| 4 | 6.049695 | **6.110799** | 6.752903 | 6.182111 |
| 5 | 4.832886 | **4.887008** | 5.081518 | 4.998154 |

Table 2: (DGHO4) Performance of the estimated parameter set and the corresponding average square error for LSURE, bGCV and LGCV on test signal "Hypchirps" with RNR = 7 and signal sample length n = $2^{13}$.

| Level | Optimal Threshold | $E((\lambda_{estimated}-\lambda_{Optimal})^2)$ | | |
|---|---|---|---|---|
| | | LSURE | bGCV | LGCV |
| 1 | 1.071360 | **0.034770** | 0.069890 | 0.038834 |
| 2 | 1.078440 | **0.050962** | 0.079986 | 0.057569 |
| 3 | 0.813360 | **0.016893** | 0.084217 | 0.018688 |
| 4 | 0.696030 | **0.010946** | 0.073336 | 0.015144 |
| 5 | 0.674610 | **0.021895** | 0.056818 | 0.024383 |
| Level | Optimal SE | $E(SE_j)$ | | |
| | | LSURE | bGCV | LGCV |
| 1 | 0.352259 | 0.392532 | **0.383042** | 0.401758 |
| 2 | 1.588368 | 1.849667 | **1.732944** | 1.927505 |
| 3 | 5.113884 | **5.463612** | 5.925474 | 5.525445 |
| 4 | 7.818871 | **8.169128** | 9.144206 | 8.437193 |
| 5 | 8.190111 | **8.726356** | 8.971381 | 8.857710 |

Table 3: (DGHO4) Performance of the estimated parameter set and the corresponding average square error for LSURE, bGCV and LGCV on test signal "Blocks" with RNR = 7 and signal sample length n = $2^{13}$.