

## Research Article

# Supervised and Unsupervised Subband Adaptive Denoising Frameworks with Polynomial Threshold Function

Tierui Gong,<sup>1,2</sup> Zhijia Yang,<sup>1</sup> Gengshan Wang,<sup>1,2</sup> and Ping Jiao<sup>1</sup>

<sup>1</sup>Department of Industrial Control Networks and Systems, Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang 110016, China

<sup>2</sup>University of Chinese Academy of Sciences, Beijing 100049, China

Correspondence should be addressed to Tierui Gong; [gongtierui@sia.cn](mailto:gongtierui@sia.cn)

Received 18 September 2016; Revised 28 December 2016; Accepted 22 January 2017; Published 9 March 2017

Academic Editor: Barak Fishbain

Copyright © 2017 Tierui Gong et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Unlike inflexible structure of soft and hard threshold function, a unified linear matrix form with flexible structure for threshold function is proposed. Based on the unified linear flexible structure threshold function, both supervised and unsupervised subband adaptive denoising frameworks are established. To determine flexible coefficients, a direct mean-square error (MSE) minimization is conducted in supervised denoising while Stein's unbiased risk estimate as a MSE estimate is minimized in unsupervised denoising. The SURE rule requires no hypotheses or a priori knowledge about clean signals. Furthermore, we discuss conditions to obtain optimal coefficients for both supervised and unsupervised subband adaptive denoising frameworks. Applying an Odd-Term Reserving Polynomial (OTRP) function as concrete threshold function, simulations for polynomial order, denoising performance, and noise effect are conducted. Proper polynomial order and noise effect are analyzed. Both proposed methods are compared with soft and hard based denoising technologies—VisuShrink, SureShrink, MiniMaxShrink, and BayesShrink—in denoising performance simulation. Results show that the proposed approaches perform better in both MSE and signal-to-noise ratio (SNR) sense.

## 1. Introduction

Signals are usually corrupted by noise in capturing and transmission stages due to environment disturbance and device error. Signal denoising has become an important research topic for a long time and a wide variety of denoising methods have been proposed. Due to their effectiveness and good performance, wavelet threshold methods have become a powerful tool for denoising problems since Donoho and several others' fundamental works. The main purpose of these methods is to estimate a wide class of functions in some smoothness spaces from their corrupted versions [1]. The different energy distribution property between smoothness spaces' functions and noise makes these methods effective: the energy of a function in smoothness is often concentrated on few coefficients while noise is spread on all coefficients.

Of the various wavelet threshold schemes, soft and hard based threshold methods are the most popular technologies and have been theoretically verified by Donoho and

Johnstone [2]. They gave a near optimal threshold value  $T = \sqrt{2\sigma^2 \log N}$  in minimax sense, where the threshold is a function of noise variance  $\sigma$  and the number of samples  $N$ . This universal thresholding method is known as VisuShrink [3]. In MSE sense, better threshold methods have been proposed including SureShrink [4] and BayesShrink [5]. SureShrink gets the threshold via minimizing Stein's unbiased risk estimate [6]. Assuming a generalized Gaussian distribution for wavelet coefficients, BayesShrink obtains the threshold through a Bayesian framework. Different a priori assumptions about the statistical distribution of wavelet coefficients, such as Gaussian scale mixture [7], Mixture of Laplacians [8], and alpha-stable [9], are also proposed, while the dependency on statistical distribution restricts their flexibility.

Soft and hard based threshold schemes suffer their own flaws due to inherent defects of soft and hard threshold functions. For soft threshold schemes, systematical biased

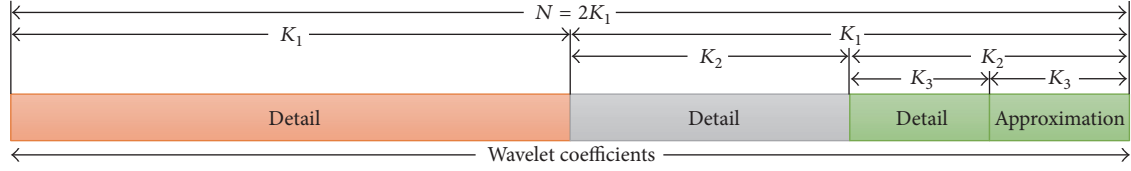


FIGURE 1: A three-level wavelet decomposition (each color represents a level). There are three detail subbands each with a length of  $K_j$  ( $1 \leq j \leq 3$ ) and one approximation subband with a length of  $K_3$ .

estimation could happen, while hard threshold schemes are less biased but less sensitive to small perturbations in the data. In addition, the more important drawback is that soft and hard threshold functions do not have continuous derivatives. Various improvements had been proposed by exploring new threshold functions [1, 10–17], but the nonnegative garrote-like functions [10–14] are still not differentiable. Zhang [1, 15], Nasri and Nezamabadi-pour [16], and Wu et al. [17], respectively, proposed a series of threshold functions with adjustable parameters. The differentiable property makes these threshold functions suitable for gradient based minimization problems.

Most of mentioned methods above depend on a single parameter threshold, which makes those methods very sensitive to threshold value and lack of freedom. More flexible and convenient strategies had been proposed by Luisier et al. [18] and Smith et al. [19]. Both employ an idea of linear combination: the threshold function is formed by a linear combination of a set of parameters. The parameters are determined via minimizing a MSE problem. Compared to gradient based minimization, this minimization is linear and is easy to solve. In this paper, we adopt the linear combination idea and propose a unified matrix form with flexible structure for threshold function. Based on the unified flexible structure threshold function, both supervised and unsupervised subband adaptive denoising frameworks are established. In supervised denoising, a direct MSE minimization is conducted while in unsupervised denoising, we apply Stein's unbiased risk estimate as MSE estimation. The SURE rule requires no hypotheses or a priori knowledge about clean signals. Furthermore, we discuss conditions for optimal solution for both supervised and unsupervised denoising frameworks.

Our contributions can be summarized as follows: (1) a unified linear matrix form with flexible structure for threshold function and a concrete OTRP function are proposed; (2) both supervised and unsupervised denoising frameworks are established; (3) conditions for guaranteeing optimal solution for minimizing problems are discussed and provided.

The paper is organized as follows. In Section 2, both supervised and unsupervised denoising frameworks are introduced. In Section 3, frameworks employing proposed OTRP function are compared with soft and hard based denoising methods and comparison results are presented. Conclusions are made in the final.

## 2. Proposed Approaches

**2.1. Problem Settings and Denoising Scheme.** In time domain, it is assumed that the clean signal  $\mathbf{x}$  is additively corrupted by noise  $\mathbf{e}$  to produce the noisy signal  $\mathbf{y}$  in the form of

$$\mathbf{y} = \mathbf{x} + \mathbf{e}, \quad (1)$$

where  $\mathbf{y} = \{y_i\}_{i=1}^N$ ,  $\mathbf{x} = \{x_i\}_{i=1}^N$ , and  $\mathbf{e} = \{e_i\}_{i=1}^N$  are discrete time series. Additive Gaussian white noise (AGWN) is only considered, with a zero mean and a  $\sigma^2$  variance; that is,  $e_i \sim \mathcal{N}(0, \sigma^2)$ .

We only consider orthonormal wavelet transform (OWT), which keeps energy conservation. In OWT domain, the AGWN is still Gaussian. Relationship of wavelet coefficients in vector form is

$$\mathbf{Y} = \mathbf{X} + \mathbf{E}, \quad (2)$$

where  $\mathbf{Y} = \mathbf{W}\mathbf{y}$ ,  $\mathbf{X} = \mathbf{W}\mathbf{x}$ , and  $\mathbf{E} = \mathbf{W}\mathbf{e}$ .  $\mathbf{W}$  is discrete wavelet transform (DWT) matrix. Each vector  $\mathbf{V}$  ( $\mathbf{V} = \mathbf{Y}, \mathbf{X}, \mathbf{E}$ ) is stacked by both detail and approximation wavelet coefficients as similarly shown in Figure 1. Each detail is a subband. Let  $\mathbf{V}_j$  ( $1 \leq j \leq J$ ) denote the  $j$ th detail subband with a vector length of  $K_j$ .  $J$  represents the total number of detail subbands and also the number of decomposition level. There holds the relationship  $K_j = N/2^j$ .

We measure the quality of denoising by mean-square error (MSE) defined as

$$\text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \varepsilon \left\{ \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 \right\}, \quad (3)$$

where  $\varepsilon(\cdot)$  is the expectation operator. An estimate of MSE is given as

$$\widehat{\text{MSE}}(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{N} \left[ (\mathbf{x} - \hat{\mathbf{x}})^T (\mathbf{x} - \hat{\mathbf{x}}) \right] = \frac{1}{N} \sum_{i=1}^N (x_i - \hat{x}_i)^2. \quad (4)$$

By considering OWT, energy conservation property guarantees that the total MSE is a weighted sum of MSE in each individual subband, which is in the form of

$$\text{MSE}(\mathbf{x}, \hat{\mathbf{x}}) = \text{MSE}(\mathbf{X}, \hat{\mathbf{X}}) = \sum_{j=1}^J \omega_j \times \text{MSE}(\mathbf{X}_j, \hat{\mathbf{X}}_j), \quad (5)$$

where  $\omega_j$  is a weight of corresponding subband. It only needs to minimize each individual subband MSE to get minimum of total MSE. In the rest of paper, we ignore subband index

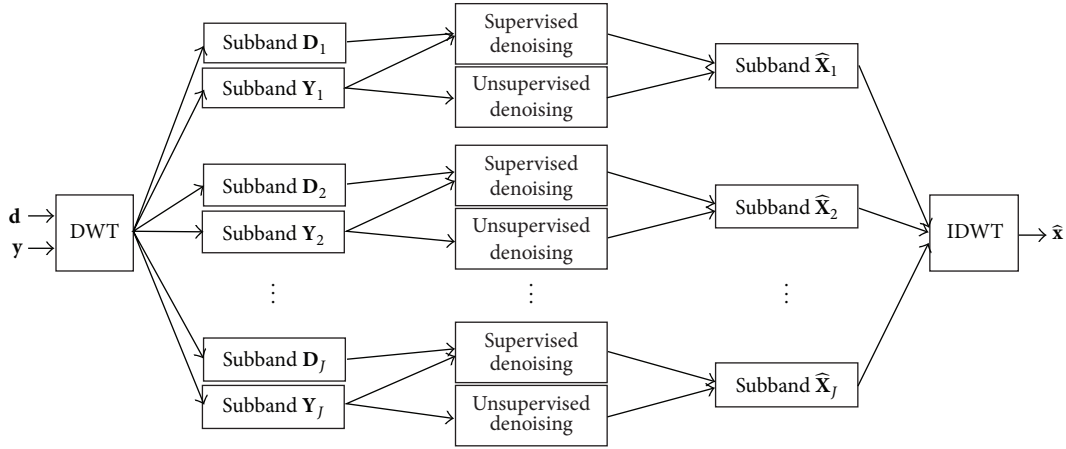


FIGURE 2: Proposed wavelet domain denoising schemes.

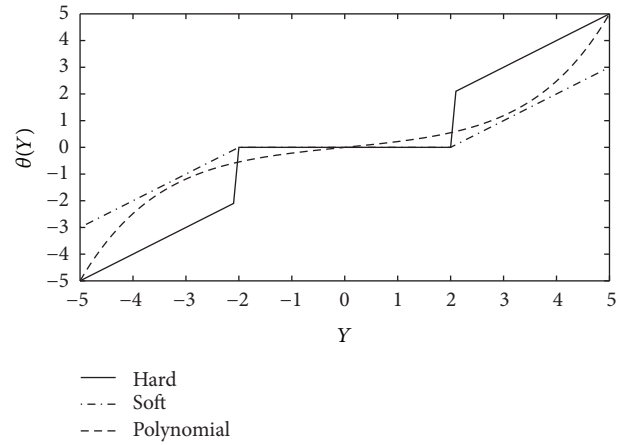
$j$  and give a unified notation  $\text{MSE}(\mathbf{X}, \hat{\mathbf{X}})$  to represent MSE in any subband.

Wavelet domain denoising generally consists of a three-stage procedure. First, perform DWT on noisy signal. Then a threshold function is applied to wavelet domain coefficients. Finally, denoised signal is obtained through an inverse DWT (IDWT). We adopt this three-stage procedure to perform our proposed denoising methods. Each subband is adaptively denoised by supervised or unsupervised denoising method as shown in Figure 2. That makes our approaches work in a subband adaptive manner. There are two input signals in Figure 2 where  $\mathbf{d}$  represents the desired signal for supervised denoising and  $\mathbf{y}$  denotes noisy signal. Corresponding uppercase letters of  $\mathbf{d}$  and  $\mathbf{y}$  denote their subbands. Both methods output an estimate for clean signal  $\mathbf{x}$ .

**2.2. Proposed Threshold Function.** The general problem can be formulated as a construction of threshold function that minimizes the MSE. By employing linear combination idea, an estimate of a clean signal can be represented by a linear combination of atoms; atoms mean the columns of a matrix. Involving wavelet transform, an estimate of clean wavelet coefficients can be represented by a linear combination of atoms which are decided by noisy wavelet coefficients. A general threshold function can be demonstrated as

$$\boldsymbol{\theta} = \boldsymbol{\Psi}\mathbf{a}, \quad (6)$$

where  $\boldsymbol{\theta}$  is an estimate for clean wavelet coefficients.  $\boldsymbol{\Psi} = [\boldsymbol{\varphi}_1 \ \boldsymbol{\varphi}_2 \ \cdots \ \boldsymbol{\varphi}_M]$  is a matrix composed of  $M$  atoms and each atom  $\boldsymbol{\varphi}_m = [\varphi_m(Y_1) \ \varphi_m(Y_2) \ \cdots \ \varphi_m(Y_K)]^T$  is decided by a function of noisy wavelet coefficients. Dimension of an atom is decided by subband wavelet coefficients number  $K_j = N/2^j$ : that is,  $K = K_j$  when denoising subband  $j$ . Besides, atom dimension  $K$  does not have direct relationship with atom number  $M$ . A more deep connection is discussed in Section 2.5. Furthermore, atom number  $M$  has nothing to do with detail subband number or decomposition level  $J$ .  $\mathbf{a} = [a_1 \ a_2 \ \cdots \ a_M]^T$  is a weight vector with each element having a weight coefficient for corresponding atom,


 FIGURE 3: Visual comparison among soft and hard threshold function and a realization of a three-order  $\theta(Y)$  with coefficients  $[a_3 \ a_2 \ a_1] = [0.00064 \ 0.016 \ 0.2]$ .

so  $\boldsymbol{\theta}$  is a weighted sum of atoms and can be denoted as  $\boldsymbol{\theta} = [\theta(Y_1) \ \theta(Y_2) \ \cdots \ \theta(Y_K)]^T$ .

In this paper, the adopted concrete threshold function is OTRP function with order  $M$ , denoted as

$$\theta(Y) = \sum_{m=1}^M a_m Y^{2m-1}. \quad (7)$$

This polynomial function could fit almost any curves along with varying polynomial order. The most important advantages are its differentiability and its linear flexible structure over soft and hard threshold functions. Clearly, (7) conforms with (6) where atom is  $\boldsymbol{\varphi}_m = [Y_1^{2m-1} \ Y_2^{2m-1} \ \cdots \ Y_K^{2m-1}]^T$ , ( $1 \leq m \leq M$ ). A comparison among soft and hard threshold function and a realization of a three-order  $\theta(Y)$  are shown in Figure 3.

**2.3. Supervised Denoising.** In supervised denoising, a desired signal  $\mathbf{d}$  is available; thus, a direct MSE minimization problem

can be carried out. As (5) indicates, total MSE is a weighted sum of MSE in each individual subband. It only needs to minimize a subband MSE to determine its flexible parameter  $\mathbf{a}$  in (6). With prior knowledge about desired signal  $\mathbf{d}$ , a direct MSE can be written as

$$\begin{aligned} \text{MSE}(\mathbf{D}, \boldsymbol{\theta}) &= \varepsilon \left\{ \|\mathbf{D} - \boldsymbol{\theta}\|_2^2 \right\} = \varepsilon \left\{ (\mathbf{D} - \boldsymbol{\theta})^T (\mathbf{D} - \boldsymbol{\theta}) \right\} \\ &= \varepsilon \left\{ \mathbf{D}^T \mathbf{D} - \mathbf{D}^T \boldsymbol{\theta} - \boldsymbol{\theta}^T \mathbf{D} + \boldsymbol{\theta}^T \boldsymbol{\theta} \right\} \\ &= \varepsilon \left\{ \mathbf{D}^T \mathbf{D} \right\} - 2\varepsilon \left\{ \boldsymbol{\theta}^T \mathbf{D} \right\} + \varepsilon \left\{ \boldsymbol{\theta}^T \boldsymbol{\theta} \right\} \quad (8) \\ &= \varepsilon \left\{ \mathbf{D}^T \mathbf{D} \right\} - 2\mathbf{a}^T \varepsilon \left\{ \boldsymbol{\Psi}^T \mathbf{D} \right\} \\ &\quad + \mathbf{a}^T \varepsilon \left\{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right\} \mathbf{a}. \end{aligned}$$

The optimal parameter  $\mathbf{a}_{\text{opt}}$  can be obtained through  $\partial \text{MSE}(\mathbf{D}, \boldsymbol{\theta}) / \partial \mathbf{a} = 0$  on condition that  $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$  is positive semidefinite and invertible. These conditions are discussed in Section 2.5. We first derive the optimal solution form as

$$\frac{\partial \text{MSE}(\mathbf{D}, \boldsymbol{\theta})}{\partial \mathbf{a}} = 2\varepsilon \left\{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right\} \mathbf{a} - 2\varepsilon \left\{ \boldsymbol{\Psi}^T \mathbf{D} \right\} = 0 \implies \quad (9)$$

$$\mathbf{a}_{\text{opt}} = \varepsilon \left\{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right\}^{-1} \varepsilon \left\{ \boldsymbol{\Psi}^T \mathbf{D} \right\}. \quad (10)$$

As seen from (10), the optimal parameter  $\mathbf{a}_{\text{opt}}$  is affected by matrix  $\boldsymbol{\Psi}$  and desired signal  $\mathbf{D}$ . The estimates of  $\varepsilon \left\{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right\}$  and  $\varepsilon \left\{ \boldsymbol{\Psi}^T \mathbf{D} \right\}$  are obtained via an estimate of expectation operator in (4):

$$\begin{aligned} \widehat{\varepsilon} \left\{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right\} &= \begin{bmatrix} \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_1^T \boldsymbol{\varphi}_1 \right\} & \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_1^T \boldsymbol{\varphi}_2 \right\} & \cdots & \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_1^T \boldsymbol{\varphi}_M \right\} \\ \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_2^T \boldsymbol{\varphi}_1 \right\} & \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_2^T \boldsymbol{\varphi}_2 \right\} & \cdots & \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_2^T \boldsymbol{\varphi}_M \right\} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_M^T \boldsymbol{\varphi}_1 \right\} & \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_M^T \boldsymbol{\varphi}_2 \right\} & \cdots & \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_M^T \boldsymbol{\varphi}_M \right\} \end{bmatrix} \\ &= \frac{1}{K} \begin{bmatrix} \boldsymbol{\varphi}_1^T \boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_1^T \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_1^T \boldsymbol{\varphi}_M \\ \boldsymbol{\varphi}_2^T \boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_2^T \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_2^T \boldsymbol{\varphi}_M \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\varphi}_M^T \boldsymbol{\varphi}_1 & \boldsymbol{\varphi}_M^T \boldsymbol{\varphi}_2 & \cdots & \boldsymbol{\varphi}_M^T \boldsymbol{\varphi}_M \end{bmatrix} \quad (11) \\ &= \frac{1}{K} \boldsymbol{\Psi}^T \boldsymbol{\Psi}, \\ \widehat{\varepsilon} \left\{ \boldsymbol{\Psi}^T \mathbf{D} \right\} &= \begin{bmatrix} \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_1^T \mathbf{D} \right\} \\ \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_2^T \mathbf{D} \right\} \\ \vdots \\ \widehat{\varepsilon} \left\{ \boldsymbol{\varphi}_M^T \mathbf{D} \right\} \end{bmatrix} = \frac{1}{K} \begin{bmatrix} \boldsymbol{\varphi}_1^T \mathbf{D} \\ \boldsymbol{\varphi}_2^T \mathbf{D} \\ \vdots \\ \boldsymbol{\varphi}_M^T \mathbf{D} \end{bmatrix} = \frac{1}{K} \boldsymbol{\Psi}^T \mathbf{D}. \end{aligned}$$

$\widehat{\varepsilon} \left\{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right\}^{-1} = K \left( \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right)^{-1}$  is obtained from  $\widehat{\varepsilon} \left\{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right\} = \boldsymbol{\Psi}^T \boldsymbol{\Psi} / K$ ; thus, coefficient  $1/K$  is eliminated by  $K$ . So, the estimate of  $\mathbf{a}_{\text{opt}}$  is obtained as

$$\mathbf{a}_{\text{opt}}^{\text{est}} = \left( \boldsymbol{\Psi}^T \boldsymbol{\Psi} \right)^{-1} \left( \boldsymbol{\Psi}^T \mathbf{D} \right). \quad (12)$$

**2.4. Unsupervised Denoising.** In practice, when a desired signal is not available, a direct MSE minimization is impossible. Constructing an estimate of MSE seems a reasonable choice. A practical approach is Stein's unbiased risk estimate [6]. This part mainly introduces a SURE based unsupervised denoising.

**2.4.1. Stein's Unbiased Risk Estimate (SURE).** This section mainly introduces Stein's theorem stated in [6, 20] and its tailored version.

**Theorem 1.**  $\mathbf{Y} = (Y_1 \ Y_2 \ \cdots \ Y_K)$  is a normal random vector with mean  $\mathbf{X} = (X_1 \ X_2 \ \cdots \ X_K)$  and identity as covariance matrix. Let  $\mathbf{Y} + \mathbf{h}(\mathbf{Y})$  be an estimate of  $\mathbf{X}$ , where function

$$\begin{aligned} \mathbf{h}(\mathbf{Y}) &= (h_1(\mathbf{Y}) \ h_2(\mathbf{Y}) \ \cdots \ h_K(\mathbf{Y})) : \mathbb{R}^K \longrightarrow \mathbb{R}^K, \\ h_k(\mathbf{Y}) &: \mathbb{R}^K \longrightarrow \mathbb{R}^1, \quad (13) \\ &\quad \forall j = 1, 2, \dots, k \end{aligned}$$

is (weakly) differentiable. If

$$\nabla \cdot \mathbf{h}(\mathbf{Y}) = \sum_{k=1}^K \left( \frac{\partial h_k}{\partial Y_k} \right), \quad (14)$$

$$\varepsilon \left\{ \nabla \cdot \mathbf{h}(\mathbf{Y}) \right\} < \infty,$$

then

$$\begin{aligned} R &= \varepsilon \left\{ \|\mathbf{Y} + \mathbf{h}(\mathbf{Y}) - \mathbf{X}\|^2 \right\} \\ &= K + \varepsilon \left\{ \|\mathbf{h}(\mathbf{Y})\|^2 + 2\nabla \cdot \mathbf{h}(\mathbf{Y}) \right\}. \end{aligned} \quad (15)$$

Stein's unbiased risk estimate of  $\text{MSE}(\mathbf{Y} + \mathbf{h}(\mathbf{Y}), \mathbf{X})$  is

$$\widehat{R} = K + \|\mathbf{h}(\mathbf{Y})\|^2 + 2\nabla \cdot \mathbf{h}(\mathbf{Y}). \quad (16)$$

The proposed structure of threshold function does not strictly satisfy conditions of  $\mathbf{h}(\mathbf{Y})$  in Theorem 1. Because each element of  $\boldsymbol{\theta}$  in (6) is  $\mathbb{R} \rightarrow \mathbb{R}$  mapping while each element in  $\mathbf{h}(\mathbf{Y})$  is  $\mathbb{R}^K \rightarrow \mathbb{R}^1$  mapping. Therefore, a tailored version of SURE for  $\boldsymbol{\theta}$  is stated in the following theorem.

**Theorem 2.**  $\mathbf{Y} = (Y_1 \ Y_2 \ \cdots \ Y_K)$  is a normal random vector with mean  $\mathbf{X} = (X_1 \ X_2 \ \cdots \ X_K)$  and  $\sigma^2 \mathbf{I}$  is covariance matrix. Let  $\boldsymbol{\theta} = (\theta(Y_1) \ \theta(Y_2) \ \cdots \ \theta(Y_K))$  be an estimate of  $\mathbf{X}$ , where  $\theta(y) : \mathbb{R} \rightarrow \mathbb{R}$  is (weakly) differentiable. If

$$\nabla \cdot \boldsymbol{\theta} = \sum_{k=1}^K \theta'(Y_k), \quad (17)$$

$$\varepsilon \left\{ \nabla \cdot \boldsymbol{\theta} \right\} < \infty,$$

then

$$\begin{aligned} \text{MSE}(\mathbf{X}, \boldsymbol{\theta}) &= \varepsilon \left\{ \mathbf{X}^T \mathbf{X} \right\} - 2\varepsilon \left\{ \boldsymbol{\theta}^T \mathbf{Y} \right\} + 2\sigma^2 \varepsilon \left\{ \nabla \cdot \boldsymbol{\theta} \right\} \\ &\quad + \varepsilon \left\{ \boldsymbol{\theta}^T \boldsymbol{\theta} \right\}. \end{aligned} \quad (18)$$

The unbiased estimate of  $\text{MSE}(\mathbf{X}, \boldsymbol{\theta})$  is

$$\widehat{R} = \frac{1}{K} \mathbf{X}^T \mathbf{X} + \frac{1}{K} \left( -2\boldsymbol{\theta}^T \mathbf{Y} + 2\sigma^2 \mathbf{1}^T \boldsymbol{\theta}' + \boldsymbol{\theta}^T \boldsymbol{\theta} \right). \quad (19)$$

*Proof.* Combining (2) and (8), the MSE is represented as

$$\begin{aligned} \text{MSE}(\mathbf{X}, \boldsymbol{\theta}) &= \varepsilon \{ \mathbf{X}^T \mathbf{X} \} - 2\varepsilon \{ \boldsymbol{\theta}^T \mathbf{X} \} + \varepsilon \{ \boldsymbol{\theta}^T \boldsymbol{\theta} \} \\ &= \varepsilon \{ \mathbf{X}^T \mathbf{X} \} - 2\varepsilon \{ \boldsymbol{\theta}^T (\mathbf{Y} - \mathbf{E}) \} + \varepsilon \{ \boldsymbol{\theta}^T \boldsymbol{\theta} \} \\ &= \varepsilon \{ \mathbf{X}^T \mathbf{X} \} - 2\varepsilon \{ \boldsymbol{\theta}^T \mathbf{Y} \} + 2\varepsilon \{ \boldsymbol{\theta}^T \mathbf{E} \} \\ &\quad + \varepsilon \{ \boldsymbol{\theta}^T \boldsymbol{\theta} \}. \end{aligned} \quad (20)$$

The only unknown part in the above equation is  $\varepsilon \{ \boldsymbol{\theta}^T \mathbf{E} \}$  while equality  $\varepsilon \{ E_k \theta(Y_k) \} = \sigma^2 \varepsilon \{ \theta'(Y_k) \}$ ,  $1 \leq k \leq K$  proved in [18] ensures that  $\varepsilon \{ \boldsymbol{\theta}^T \mathbf{E} \}$  can be transformed to a known part. Concrete deduction formula is in the following equation:

$$\begin{aligned} \varepsilon \{ \boldsymbol{\theta}^T \mathbf{E} \} &= \varepsilon \left\{ \sum_{k=1}^K E_k \theta(Y_k) \right\} = \sum_{k=1}^K \varepsilon \{ E_k \theta(Y_k) \} \\ &= \sigma^2 \sum_{k=1}^K \varepsilon \{ \theta'(Y_k) \} = \sigma^2 \varepsilon \left\{ \sum_{k=1}^K \theta'(Y_k) \right\} \\ &= \sigma^2 \varepsilon \{ \nabla \cdot \boldsymbol{\theta} \}. \end{aligned} \quad (21)$$

So, MSE can be rewritten as

$$\begin{aligned} \text{MSE}(\mathbf{X}, \boldsymbol{\theta}) &= \varepsilon \{ \mathbf{X}^T \mathbf{X} \} - 2\varepsilon \{ \boldsymbol{\theta}^T \mathbf{Y} \} + 2\sigma^2 \varepsilon \{ \nabla \cdot \boldsymbol{\theta} \} \\ &\quad + \varepsilon \{ \boldsymbol{\theta}^T \boldsymbol{\theta} \}. \end{aligned} \quad (22)$$

Estimate the expectation of  $\varepsilon(\mathbf{X}^T \mathbf{X})$ ,  $\varepsilon(\boldsymbol{\theta}^T \mathbf{Y})$ ,  $\varepsilon(\nabla \cdot \boldsymbol{\theta})$ , and  $\varepsilon(\boldsymbol{\theta}^T \boldsymbol{\theta})$  via

$$\begin{aligned} \widehat{\varepsilon}(\mathbf{X}^T \mathbf{X}) &= \frac{1}{K} \mathbf{X}^T \mathbf{X}, \\ \widehat{\varepsilon}(\boldsymbol{\theta}^T \mathbf{Y}) &= \frac{1}{K} \boldsymbol{\theta}^T \mathbf{Y}, \\ \widehat{\varepsilon}(\nabla \cdot \boldsymbol{\theta}) &= \frac{1}{K} \mathbf{1}^T \boldsymbol{\theta}', \\ \widehat{\varepsilon}(\boldsymbol{\theta}^T \boldsymbol{\theta}) &= \frac{1}{K} \boldsymbol{\theta}^T \boldsymbol{\theta}. \end{aligned} \quad (23)$$

It is obvious that (19) is an unbiased estimate of  $\text{MSE}(\mathbf{X}, \boldsymbol{\theta})$ .

Proof is completed.  $\square$

**2.4.2. Coefficients Determination.** From (6), we first obtain element derivative of  $\boldsymbol{\theta}$  in the form of

$$\boldsymbol{\theta}'(Y_k) = \mathbf{a}^T [\varphi'_1(Y_k) \ \varphi'_2(Y_k) \ \cdots \ \varphi'_M(Y_k)]^T. \quad (24)$$

So, the formula of  $\nabla \cdot \boldsymbol{\theta}$  can be deduced as

$$\begin{aligned} \nabla \cdot \boldsymbol{\theta} &= \sum_{k=1}^K \boldsymbol{\theta}'(Y_k) \\ &= \mathbf{a}^T \left[ \sum_{k=1}^K \varphi'_1(Y_k) \ \sum_{k=1}^K \varphi'_2(Y_k) \ \cdots \ \sum_{k=1}^K \varphi'_M(Y_k) \right]^T \\ &= \mathbf{a}^T [\mathbf{1}^T \boldsymbol{\varphi}'_1 \ \mathbf{1}^T \boldsymbol{\varphi}'_2 \ \cdots \ \mathbf{1}^T \boldsymbol{\varphi}'_M]^T = \mathbf{a}^T (\mathbf{1}^T \boldsymbol{\Psi}')^T. \end{aligned} \quad (25)$$

Thus,

$$\varepsilon \{ \nabla \cdot \boldsymbol{\theta} \} = \mathbf{a}^T \varepsilon \{ (\mathbf{1}^T \boldsymbol{\Psi}')^T \}. \quad (26)$$

Combining (6) and (18) and (26),  $\text{MSE}(\mathbf{X}, \boldsymbol{\theta})$  can be rewritten as

$$\begin{aligned} \text{MSE}(\mathbf{X}, \boldsymbol{\theta}) &= \varepsilon \{ \mathbf{X}^T \mathbf{X} \} - 2\mathbf{a}^T \varepsilon \{ \boldsymbol{\Psi}^T \mathbf{Y} \} \\ &\quad + 2\sigma^2 \mathbf{a}^T \varepsilon \{ (\mathbf{1}^T \boldsymbol{\Psi}')^T \} + \mathbf{a}^T \varepsilon \{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \} \mathbf{a}. \end{aligned} \quad (27)$$

$\mathbf{a}_{\text{opt}}$  is obtained through  $\partial \text{MSE}(\mathbf{X}, \boldsymbol{\theta}) / \partial \mathbf{a} = 0$  on condition that  $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$  is positive semidefinite and invertible. These conditions the same as in supervised denoising are discussed in Section 2.5. Here, we first derive the optimal solution form as

$$\begin{aligned} \frac{\partial \text{MSE}(\mathbf{X}, \boldsymbol{\theta})}{\partial \mathbf{a}} &= 2\varepsilon \{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \} \mathbf{a} - 2\varepsilon \{ \boldsymbol{\Psi}^T \mathbf{Y} \} + 2\sigma^2 \varepsilon \{ (\mathbf{1}^T \boldsymbol{\Psi}')^T \} = 0 \implies \end{aligned} \quad (28)$$

$$\mathbf{a}_{\text{opt}} = \varepsilon \{ \boldsymbol{\Psi}^T \boldsymbol{\Psi} \}^{-1} \left( \varepsilon \{ \boldsymbol{\Psi}^T \mathbf{Y} \} - \sigma^2 \varepsilon \{ (\mathbf{1}^T \boldsymbol{\Psi}')^T \} \right). \quad (29)$$

An estimate of  $\mathbf{a}_{\text{opt}}$  is obtained via Theorem 2. From (6), we first obtain pointwise derivative of  $\boldsymbol{\theta}$  that can be denoted as

$$\boldsymbol{\theta}' = \boldsymbol{\Psi}' \mathbf{a}. \quad (30)$$

Then (19) can be rewritten as

$$\begin{aligned} \widehat{R} &= \frac{1}{K} \mathbf{X}^T \mathbf{X} \\ &\quad + \frac{1}{K} \left( -2\mathbf{a}^T \boldsymbol{\Psi}^T \mathbf{Y} + 2\sigma^2 \mathbf{1}^T \boldsymbol{\Psi}' \mathbf{a} + \mathbf{a}^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \mathbf{a} \right) \\ &= \frac{1}{K} \mathbf{X}^T \mathbf{X} \\ &\quad + \frac{1}{K} \left( -2\mathbf{a}^T \boldsymbol{\Psi}^T \mathbf{Y} + 2\sigma^2 \mathbf{a}^T (\mathbf{1}^T \boldsymbol{\Psi}')^T + \mathbf{a}^T \boldsymbol{\Psi}^T \boldsymbol{\Psi} \mathbf{a} \right). \end{aligned} \quad (31)$$

The result can be found by setting the derivative over  $\mathbf{a}$  to zero. Still positive semidefinite and invertible must be satisfied for  $\boldsymbol{\Psi}^T \boldsymbol{\Psi}$ . Estimate for optimal solution is achieved by

$$\frac{\partial \widehat{R}}{\partial \mathbf{a}} = \frac{1}{K} \left( -2\boldsymbol{\Psi}^T \mathbf{Y} + 2\sigma^2 (\mathbf{1}^T \boldsymbol{\Psi}')^T + 2\boldsymbol{\Psi}^T \boldsymbol{\Psi} \mathbf{a} \right) = 0 \implies \quad (32)$$

$$\mathbf{a}_{\text{opt}}^{\text{est}} = (\boldsymbol{\Psi}^T \boldsymbol{\Psi})^{-1} \left( \boldsymbol{\Psi}^T \mathbf{Y} - \sigma^2 (\mathbf{1}^T \boldsymbol{\Psi}')^T \right). \quad (33)$$

**2.4.3. Noise Variance Estimate.** As it can be seen from (29) and (33), noise variance  $\sigma$  is needed in unsupervised denoising. In practice, there are various methods that can provide estimate for AGWN variance, such as the most popular median absolute deviation (MAD) [3], SVD-based [21], and block-based [22, 23] methods. MAD implements simply and achieves accurate estimate. For simplicity, we adopt MAD to estimate noise variance. The estimate is denoted as

$$\widehat{\sigma} = \frac{\text{MAD}}{0.6745}, \quad (34)$$

$$\text{MAD} = \text{Median} \{ |Y_1 - \bar{Y}|, \dots, |Y_K - \bar{Y}| \},$$

where  $\bar{Y}$  is the mean value in corresponding subband.



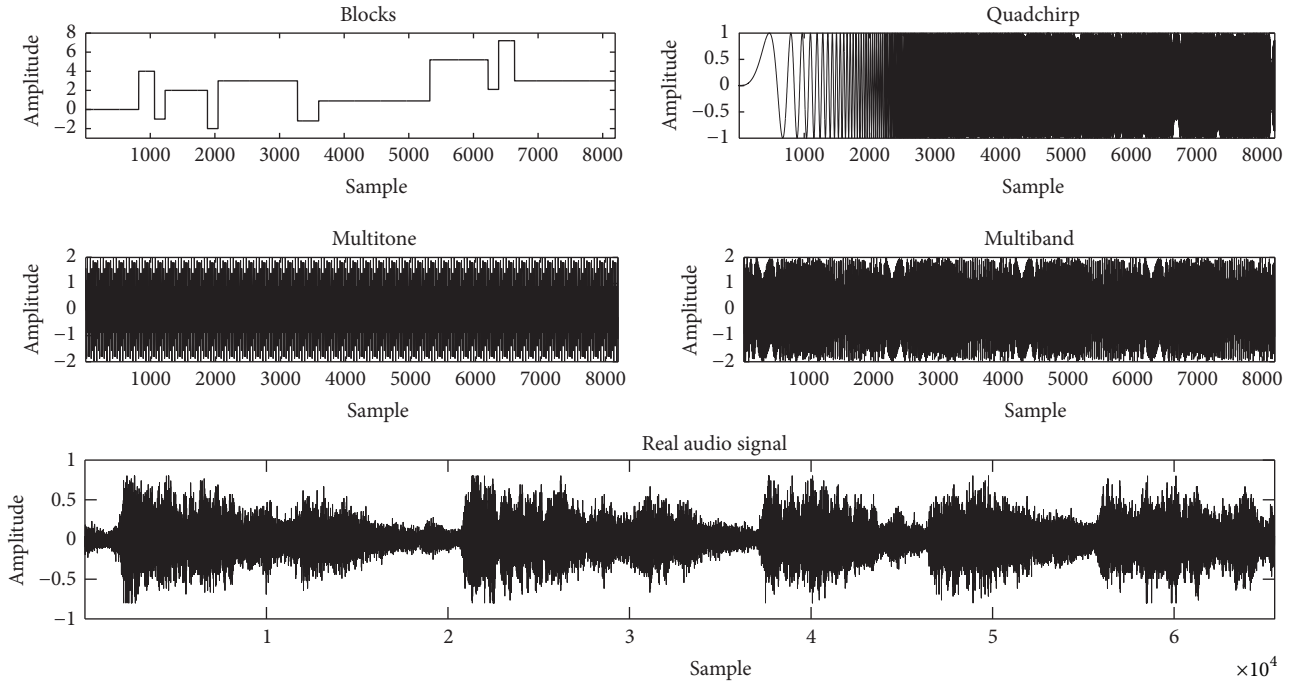


FIGURE 4: Clean “blocks,” “quadchirp,” “multitone,” and “multiband,” each signal with length of 8192 and audio signal with a length of 65536.

**2.5. Optimal Solution Guarantee Discussion.** Optimal solution assurance conditions for both supervised and unsupervised denoising are discussed in this section. To guarantee optimal solution of  $\mathbf{a}$  by setting  $\partial \text{MSE}(\mathbf{X}, \boldsymbol{\theta}) / \partial \mathbf{a}$  to zero, the matrix  $\Psi^T \Psi$  must be positive semidefinite. According to the definition of positive semidefinite matrix,  $\mathbf{A}$  is positive semidefinite when  $\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$  for any vector  $\mathbf{x}$ . So, it is easy to verify positive semidefinite of  $\Psi^T \Psi$ ; that is, for any vector  $\mathbf{x}$ , there always exists

$$\mathbf{x}^T \Psi^T \Psi \mathbf{x} = \langle \Psi \mathbf{x}, \Psi \mathbf{x} \rangle = \|\Psi \mathbf{x}\|_2^2 \geq 0. \quad (35)$$

Thus, it can be concluded that  $\Psi^T \Psi$  is always positive semidefinite. Positive semidefinite property guarantees that  $\text{MSE}(\mathbf{X}, \boldsymbol{\theta})$  is convex while invertible of  $\Psi^T \Psi$  makes the optimal solution of  $\mathbf{a}$  in the form of (10), (12) and (29), (33). It must be full rank to ensure invertible  $\Psi^T \Psi$ , so we can deduce that  $\Psi$  must be full row or column rank because of  $\text{rank}(\Psi) = \text{rank}(\Psi^T \Psi)$ . In practice, we do not use too many atoms to form matrix  $\Psi$  and the number of noisy coefficients  $K$  in a subband is much greater than atom number  $M$  (i.e.,  $K \gg M$ ). In other words, with row number far greater than column number,  $\Psi$  is a very “thin” matrix. Thus, matrix  $\Psi$  tends to be full column rank and matrix  $\Psi^T \Psi$  tends to be full rank.

### 3. Experiments and Results

**3.1. Experiment Settings.** Standard experiment settings for latter simulations are given. The adopted signals in simulations are “blocks,” “quadchirp,” “multitone,” and “multiband”

with AGWN. In order to explain the superiority, the proposed methods are also tested by real audio signal. All those clean signals are shown in Figure 4 with each signal having length of 8192 for the first four signals and a length of 65536 for audio signal. The adopted orthogonal wavelet is “sym8” and decomposition level is  $J = 5$ . So the  $j$ th detail subband has a vector length of  $N/2^j$  ( $1 \leq j \leq 5$ ), where  $N$  can be 8192 or 65536.

In supervised denoising, coefficient  $\mathbf{a}$  is determined by matrix  $\Psi$  and desired subband  $\mathbf{D}$  in each individual subband. Desired subbands come from desired signal by OWT. Desired signal can be clean or noisy. In practice, it may only have access to noisy signal with known noise level; thus, we adopt it as desired signal to determine coefficient  $\mathbf{a}$  and then use  $\mathbf{a}$  to derive the denoised version of noisy signal with unknown noise level. A remarkable notice is that the desired signal and noisy signal are, respectively, the same clean signal corrupted by known and unknown noise level. In unsupervised denoising, coefficient  $\mathbf{a}$  is determined by matrix  $\Psi$ , noisy signal subband  $\mathbf{Y}$ , and noise variance. There are no requirements for clean signal or desired signal. For simplicity, an estimate for noise variance is provided by MAD.

Applying the proposed OTRP function as concrete threshold function, supervised and unsupervised denoising methods are compared with several other available techniques, that is, VisuShrink, SureShrink, MiniMaxShrink, and BayesShrink. Wavelab 850 is applied to perform the other four techniques; their noise variance estimates are also provided by MAD. Denoising quality is measured by both MSE and SNR, which are defined as

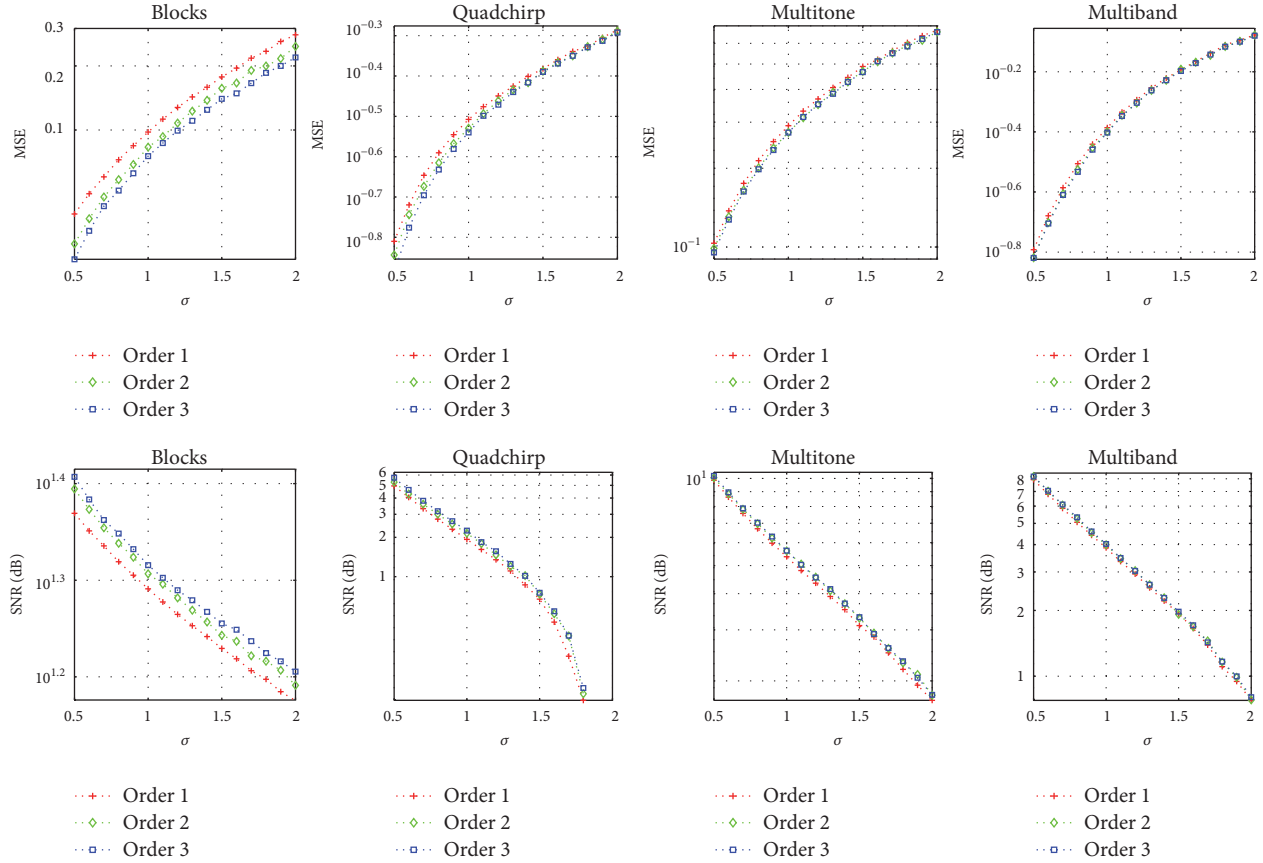


FIGURE 5: PO1 to PO3 comparisons in supervised denoising condition.

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (\hat{x}_i - x_i)^2, \quad (36)$$

$$\text{SNR} = 10 \log \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N (\hat{x}_i - x_i)^2}.$$

**3.2. Polynomial Order Simulation.** Before carrying out the final denoising, proper polynomial order needs to be explored. Due to the fact that polynomial order is equal to atom number of matrix  $\Psi$ , it needs to select proper polynomial order to guarantee full column rank of  $\Psi$  and thus full rank of  $\Psi^T \Psi$ . Although it has already been discussed in Section 2.5 that subband wavelet coefficients number is far greater than polynomial order and  $\Psi^T \Psi$  tends to be full rank, simulations are still needed. Whether  $\Psi^T \Psi$  is full rank or not is determined by polynomial order and subband wavelet coefficients. Thus, a probability statistical table is made through simulations. The achieved probabilities make up Table 1 where PO represents polynomial order and SB denotes subband. The probabilities reveal how often  $\Psi^T \Psi$  is full rank in different PO for different signals corrupted with different noise level.

Six orders of PO are tested in Table 1. Due to 5-level wavelet decomposition, there are five detail SBs to be

processed. Each SB corresponds to a  $\Psi^T \Psi$  and different PO corresponds to different dimension of  $\Psi^T \Psi$ . So, there are 30 combinations shown in Table 1 header where each PO contains five SBs. In  $x$ -axis direction, Table 1 reflects how PO influences matrix full rank probabilities. In  $y$ -axis direction of Table 1, four signals are tested with each corrupted by different noise variance  $\sigma$ . This reflects how noise level influences matrix full rank probabilities. For example, the italic number shown in Table 1 means matrix full rank probability is 100%. This matrix is determined by PO2 and “blocks” signal’s SB3 in  $\sigma = 1.6$  noise level. All probabilities are obtained by counting matrix full rank times in a 100-trial.

From Table 1, PO1–PO3 always ensure 100% matrix full rank for four signals in all tested noise level. Saying 100% matrix full rank for a signal we mean 100% matrix full rank for every subband of this signal. PO4 only cannot ensure “blocks” due to almost 0% probability for SB5 (bold underlined numbers, bold numbers indicate the probabilities under 100%). For PO5 and PO6, there are more and more bold numbers than PO4. In horizontal direction, the trend is higher PO lower probability for satisfying matrix full rank. In vertical direction, higher noise level tends to lead lower probability for full rank, but PO is a dominant factor. From statistical analysis, we restrict our option from PO1 to PO3.

Visual comparisons of PO1 to PO3 in supervised denoising are shown in Figure 5. As it can be seen, in both MSE and





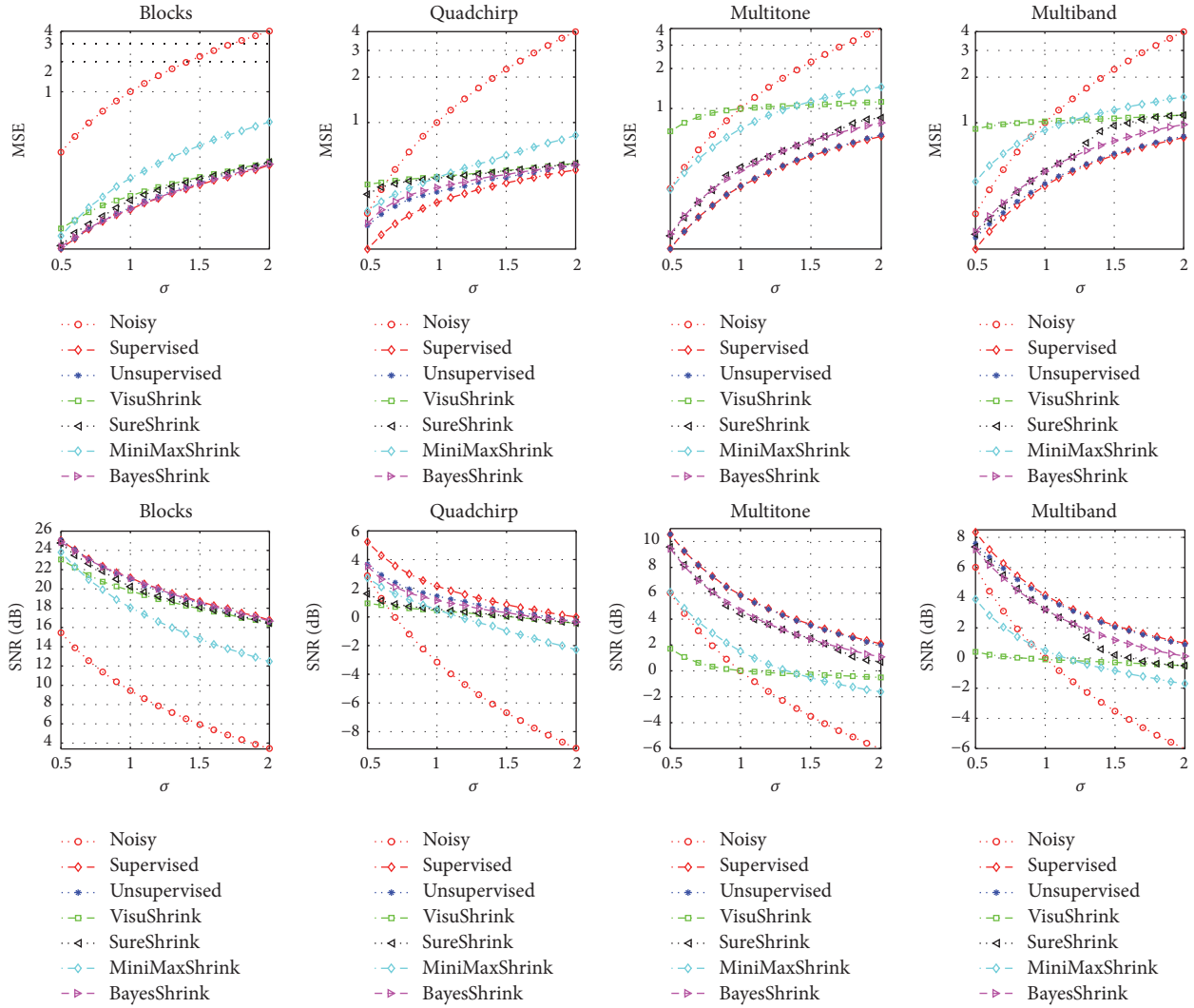


FIGURE 6: Comparison of both MSE and SNR changing with noise variance  $\sigma$  for “blocks,” “quadchirp,” “multitone,” and “multiband” denoising.

SNR sense PO1 is inferior to PO2 and PO3 in all four signals. PO2 is inferior to PO3 in blocks and almost equals PO3 in the three remaining signals. In unsupervised denoising, similar conclusions can be made and are not repetitive state. In the end, we choose PO2 for final denoising simulation.

**3.3. Denoising Performance Comparison.** As mentioned above, comparisons are made among supervised and unsupervised denoising methods and VisuShrink, SureShrink, MiniMaxShrink, and BayesShrink. In both MSE and SNR sense, denoising performance comparisons in global and local views with different noise variance  $\sigma$  are presented in Figure 6.  $\sigma$  changes in range of 0.5–2.0 with step 0.1 and each data point plotted in Figure 6 is a 50-average result. A remarkable notice is that we adopt a clean signal corrupted by noise in  $\sigma = 0.5$  level as desired signal for supervised denoising. Moreover, MSEs in Figure 6 are plotted in log scale and SNRs in Figure 6 are plotted in linear scale.

From global and local views of Figure 6, supervised denoising method achieves better MSE and SNR quality than unsupervised denoising method and BayesShrink for all four noisy signals. All these three methods have better MSE and SNR performance than VisuShrink, SureShrink, and MiniMaxShrink. In noisy “blocks” processing, BayesShrink is slightly better than unsupervised denoising method when  $\sigma$  is greater than 0.9 while poorer when  $\sigma$  is less than 0.9 in both MSE and SNR sense. However, unsupervised denoising method has better MSE and SNR performance than BayesShrink in noisy quadchirp, multitone, and multiband processing. Thus, a conclusion can be made that the proposed supervised and unsupervised denoising methods show their advantages.

Our approaches are also validated for real-world signals. We test supervised and unsupervised denoising approaches by an AGWN corrupted real audio signal. In both MSE and SNR sense, comparisons with changing  $\sigma$  in range of 0.1–0.5

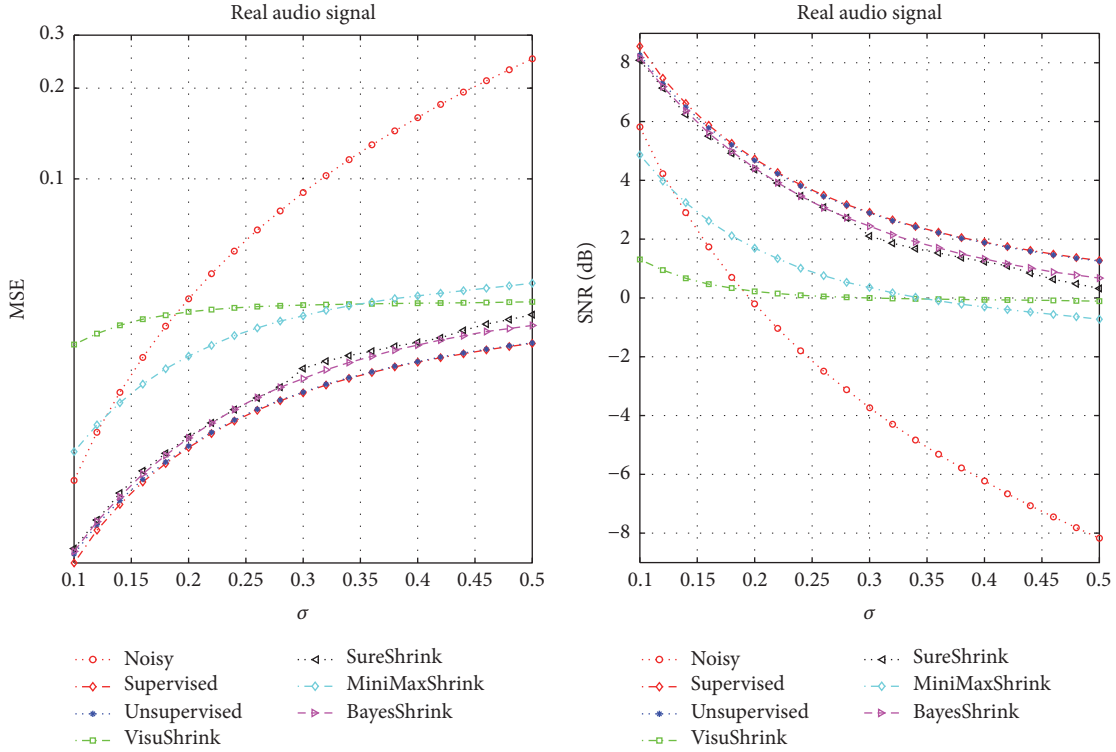


FIGURE 7: Denoising performance comparisons in both MSE and SNR sense with changing noise variance  $\sigma$  for real audio signal denoising.

are shown in Figure 7 and each data point plotted is still a 50-average result. We adopt clean audio signal corrupted by noise in  $\sigma = 0.5$  level as desired signal for supervised denoising. In addition, MSEs are plotted in log scale and SNRs are plotted in linear scale. From global and local views of Figure 7, the proposed supervised and unsupervised denoising methods have smaller MSEs and higher SNRs than VisuShrink, SureShrink, MiniMaxShrink, and BayesShrink. A conclusion can still be made that the proposed supervised and unsupervised denoising methods show their advantages in real-world condition.

*Remark.* All denoising methods except for supervised denoising involved in Figures 6 and 7 require noise variance  $\sigma$ , and estimates of it are achieved by MAD for all those methods. So, proposed unsupervised method and the other four techniques work under the same condition, which reflects superiority of the proposed method. In practice, MAD with simple implementation and estimate accuracy do satisfy our requirements.

**3.4. Noise Effect Analysis.** In this section, different noise levels of desired signal for supervised denoising are analyzed. It only aims at real audio signal. In simulation, desired signal noise levels are set in range  $\sigma = 0.1$  to  $\sigma = 2.5$  with equal step 0.4. Both MSE and SNR results are shown in Figure 8 and each data point plotted in Figure 8 is averaged by 50 times. It must be noticed that  $\sigma$  on  $x$ -axis represents noise level of noisy signal to be processed. All desired signal noise levels are compared with BayesShrink denoising method. From

Figure 8 it can be seen that all  $\sigma = 0.1$  to  $\sigma = 2.5$  desired signal noise levels have better denoising performance than BayesShrink in MSE and SNR sense when  $\sigma \geq 0.18$  on  $x$ -axis. While when  $\sigma < 0.18$  on  $x$ -axis BayesShrink does better than  $\sigma = 2.1$  and  $\sigma = 2.5$  and worse than  $\sigma = 0.1$  to  $\sigma = 1.7$  in MSE and SNR sense. In addition, denoising performance becomes worse when desired signal noise level gets higher. In conclusion, if some noisy desired signal (even higher noise level than the signal to be processed) is available at hand, it is still possible to achieve good denoising performance.

## 4. Conclusions

In order to utilize the effectiveness of wavelet domain denoising, two subband adaptive denoising schemes were proposed in this paper. For any subband, a unified linear matrix form with flexible structure of threshold function was proposed and OTRP function was proposed for a concrete realization. Based on the unified linear flexible structure threshold function, both supervised and unsupervised subband adaptive denoising frameworks were established. A direct MSE minimization was conducted in supervised denoising while Stein's unbiased risk estimate as MSE estimate was minimized in unsupervised denoising for flexible coefficients determination. Conditions to obtain optimal coefficients were further discussed.

Applying the concrete OTRP function, simulations for polynomial order, denoising performance, and noise effect were conducted. A full rank probability statistical table was generated in polynomial order simulation. The table reflects

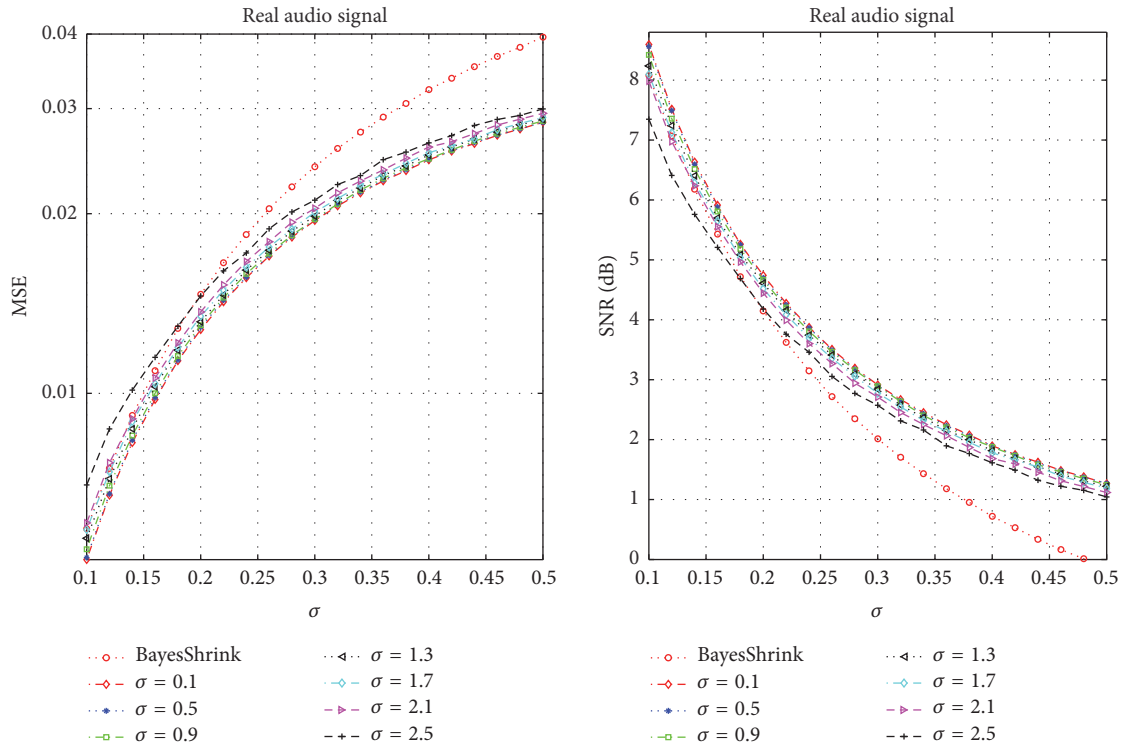


FIGURE 8: Different desired signal noise level and BayesShrink denoising performance comparisons in both MSE and SNR sense with changing noise variance  $\sigma$  for real audio signal denoising.

that PO1–PO3 always ensure full rank of  $\Psi^T\Psi$ . By choosing PO2, denoising performance was compared among proposed supervised and unsupervised denoising methods and VisuShrink, SureShrink, MiniMaxShrink, and BayesShrink. Comparison results demonstrate that the proposed methods have advantages over those soft and hard threshold function based denoising schemes in both simulation and real-world condition. Final simulation of noise effect analysis for supervised denoising demonstrates that supervised denoising method is robust for different desired signal noise level.

**Competing Interests**

The authors declare that there is no conflict of interests.

**Acknowledgments**

This work is fully supported by National Science and Technology Major Project of the Ministry of Science and Technology of China (no. Y6D8020801).

**References**

[1] X.-P. Zhang, “Thresholding neural network for adaptive noise reduction,” *IEEE Transactions on Neural Networks*, vol. 12, no. 3, pp. 567–584, 2001.  
 [2] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *Journal of the American Statistical Association*, vol. 90, no. 432, pp. 1200–1224, 1995.

[3] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.  
 [4] D. L. Donoho, “De-noising by soft-thresholding,” *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.  
 [5] S. G. Chang, B. Yu, and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000.  
 [6] C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.  
 [7] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, “Image denoising using scale mixtures of Gaussians in the wavelet domain,” *IEEE Transactions on Image Processing*, vol. 12, no. 11, pp. 1338–1351, 2003.  
 [8] H. Rabbani and M. Vafadoost, “Wavelet based image denoising based on a mixture of Laplace distributions,” *Iranian Journal of Science and Technology, Transaction B: Engineering*, vol. 30, no. 6, pp. 711–733, 2006.  
 [9] A. Achim and E. Kuruoghlu, “Image denoising using alpha-stable distributions in the complex wavelet domain,” *IEEE Signal Processing Letter*, vol. 12, no. 1, pp. 17–20, 2005.  
 [10] H.-Y. Gao, “Wavelet shrinkage denoising using the non-negative garrote,” *Journal of Computational and Graphical Statistics*, vol. 7, no. 4, pp. 469–488, 1998.  
 [11] C. He, J. Xing, J. Li, Q. Yang, and R. Wang, “A new wavelet thresholding function based on hyperbolic tangent function,” *Mathematical Problems in Engineering*, vol. 2015, Article ID 528656, 10 pages, 2015.  
 [12] L. Jing-Yi, L. Hong, Y. Dong, and Z. Yan-Sheng, “A new wavelet threshold function and denoising application,” *Mathematical*

- Problems in Engineering*, vol. 2016, Article ID 3195492, 8 pages, 2016.
- [13] T.-H. Yi, H.-N. Li, and X.-Y. Zhao, "Noise smoothing for structural vibration test signals using an improved wavelet thresholding technique," *Sensors*, vol. 12, no. 8, pp. 11205–11220, 2012.
  - [14] Z. Liu, Y. Mi, and Y. Mao, "An improved real-time denoising method based on lifting wavelet transform," *Measurement Science Review*, vol. 14, no. 3, pp. 152–159, 2014.
  - [15] X.-P. Zhang and M. D. Desai, "Adaptive denoising based on SURE risk," *IEEE Signal Processing Letters*, vol. 5, no. 10, pp. 265–267, 1998.
  - [16] M. Nasri and H. Nezamabadi-pour, "Image denoising in the wavelet domain using a new adaptive thresholding function," *Neurocomputing*, vol. 72, no. 4–6, pp. 1012–1025, 2009.
  - [17] G.-W. Wu, C.-M. Wang, J.-D. Bao, Y. Chen, and Y.-P. Hu, "A wavelet threshold de-noising algorithm based on adaptive threshold function," *Journal of Electronics and Information Technology*, vol. 36, no. 6, pp. 1340–1347, 2014.
  - [18] F. Luisier, T. Blu, and M. Unser, "A new SURE approach to image denoising: interscale orthonormal wavelet thresholding," *IEEE Transactions on Image Processing*, vol. 16, no. 3, pp. 593–606, 2007.
  - [19] C. B. Smith, S. Aghaian, and D. Akopian, "A wavelet-denoising approach using polynomial threshold operators," *IEEE Signal Processing Letters*, vol. 15, pp. 906–909, 2008.
  - [20] G. Yu, S. Mallat, and E. Bacry, "Audio denoising by time-frequency block thresholding," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1830–1839, 2008.
  - [21] W. Liu and W. Lin, "Additive white Gaussian noise level estimation in SVD domain for images," *IEEE Transactions on Image Processing*, vol. 22, no. 3, pp. 872–883, 2013.
  - [22] D.-H. Shin, R.-H. Park, S. Yang, and J.-H. Jung, "Block-based noise estimation using adaptive Gaussian filtering," *IEEE Transactions on Consumer Electronics*, vol. 51, no. 1, pp. 218–226, 2005.
  - [23] A. Amer and E. Dubois, "Fast and reliable structure-oriented video noise estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 1, pp. 113–118, 2005.





# Hindawi

Submit your manuscripts at  
<https://www.hindawi.com>

