

A statistical approach to signal denoising based on data-driven multiscale representation

Khuram Naveed^{a,*}, Muhammad Tahir Akhtar^b, Muhammad Faisal Siddiqui^a,
Naveed ur Rehman^c

^a Department of Electrical and Computer Engineering, COMSATS University Islamabad (CUI), Park Road, Islamabad, Pakistan

^b Department of Electrical and Computer Engineering, School of Engineering & Digital Sciences, Nazarbayev University, Kabanbay Batyr Avenue 53, Nur-Sultan City, Kazakhstan

^c Department of Engineering - Electrical and Computer Engineering, Aarhus University, Aarhus, Denmark

ARTICLE INFO

Article history:

Available online 23 October 2020

Keywords:

Variational mode decomposition (VMD)
Empirical distribution function (EDF)
Goodness of fit test (GoF) test
Cramer Von Mises (CVM) statistic

ABSTRACT

We develop a data-driven approach for signal denoising that utilizes variational mode decomposition (VMD) algorithm and Cramer Von Mises (CVM) statistic. In comparison with the classical empirical mode decomposition (EMD), VMD enjoys superior mathematical and theoretical framework that makes it robust to noise and mode mixing. These desirable properties of VMD materialize in segregation of a major part of noise into a few final modes while majority of the signal content is distributed among the earlier ones. To exploit this representation for denoising purpose, we propose to estimate the distribution of noise from the predominantly noisy modes and then use it to detect and reject noise from the remaining modes. The proposed approach first selects the predominantly noisy modes using the CVM measure of statistical distance. Next, CVM statistic is used locally on the remaining modes to test how closely the modes fit the estimated noise distribution; the modes that yield closer fit to the noise distribution are rejected (set to zero). Extensive experiments demonstrate the superiority of the proposed method as compared to the state of the art in signal denoising and underscore its utility in practical applications where noise distribution is not known a priori.

© 2020 Elsevier Inc. All rights reserved.

1. Introduction

Signals from various practical applications are subject to unwanted noise due to various physical limitations of acquisition systems, e.g., audio recording systems, lidar systems, EEG and ECG acquisition systems etc. Hence, to avoid any false decisions based on these noisy signals, it is necessary to remove the unwanted noise beforehand. In this regard, it is customary to assume the noise in time series data follows the additive white Gaussian noise (wGn) model. That problem of removal of additive wGn has been optimally solved for wide sense stationary signals, i.e., signals with perfectly known invariable statistics, using the Wiener filter. However, that approach may not be adequate in practical settings due to the following reasons. Firstly, majority of real life signals are nonstationary in that their attributes (statistics) change with time.

Secondly, the assumed wGn model may not always be used to characterize noise in time series data, e.g., EEG/ECG signals. Consequently, more evolved techniques capable of accounting for the non-stationarity of signal and non-Gaussianity of noise are required.

Discrete wavelet transform (DWT) is a multiscale method to process the non-stationary signals that exhibits property of sparse distribution of signal singularities within its coefficients. The noise coefficients, on the other hand, have lower amplitudes and uniform spread [1]. That allows to differentiate between signal and noise coefficients using a suitable threshold (e.g., universal threshold [2]) or statistical shrinkage functions, e.g., [3,4]. The foundations of these methods are built on prior assumptions about the (distribution) models of (true) signal and noise. In this regard, random noise due to data-acquisition and communication systems is conventionally modeled using the additive wGn model but that do not fully account for the factors contributing to the noise during the acquisition. However, specifying a signal model is challenging owing to the arbitrary nature of information generally found within the times series data. Moreover, specification of prior models restricts the efficacy of these methods in real world signals.

* Corresponding author.

E-mail addresses: khuram.naveed@comsats.edu.pk (K. Naveed), muhammad.akhtar@nu.edu.kz, akhtar@ieee.org (M.T. Akhtar), faisal_siddiqui@comsats.edu.pk (M.F. Siddiqui), naveed.rehman@eng.au.dk (N. ur Rehman).

This issue has been partially addressed within the framework presented in [5] that combines DWT with the goodness of fit (GoF) test. Hereafter, this approach is called as DWT-GoF method. It is worth mentioning that the DWT-GoF method requires only a prior noise model. Here, noise is expediently modeled as a zero-mean additive wGn that is, for example. The detection of wGn at multiple wavelet scales is facilitated by the fact that the DWT preserves the Gaussianity of noise. This essentially requires detection and rejection of wavelet coefficients fitting the Gaussian distribution for denoising. Henceforth, the DWT-GoF method [5] rejects noise from DWT scales by estimating the GoF of Gaussian distribution on the multiscale coefficients. An improved version of the DWT-GoF method has been proposed in [6,7] that employ GoF test along with the dual-tree complex wavelet transform (DTCWT), which is called the DT-GOF-NeighFilt method in the sequel. The key feature of the DT-GOF-NeighFilt method is to incorporate a novel neighborhood filtering technique to minimize the loss of signal details while rejecting the noise. Apart from the GoF test, other hypothesis testing tools such as False discovery rate (FDR), Bayesian local false discovery rate (BLFDR) are also used in combination with wavelet transforms for signal denoising [8].

Another avenue for multiscale denoising involves data-driven decomposition techniques. For instance, empirical mode decomposition (EMD) [9] that employs a data-driven approach to extract the principal oscillatory modes from a signal. Owing to EMD's ability to expand a signal into its inherent intrinsic mode functions (IMFs), EMD is considered well suited for processing the non-stationary signals generally encountered in practice. When employed for denoising, EMD aims at detecting the IMFs representing the (oscillatory) signal parts and rejecting the IMFs corresponding to the non-oscillatory noise. A wavelet-inspired interval-thresholding function is used for detecting the oscillatory signal parts from the noisy IMFs [10]. Specifically, the EMD-based interval thresholding (EMD-IT) [10] aims to detect the oscillations separated by two consecutive zero crossings. This is achieved by comparing the extrema of an interval against a threshold value leading to either retention or rejection of the whole interval.

Instead of performing thresholding, the work in [11] employed statistical tools to detect the relevant (signal) modes for a partial reconstruction of the denoised signal. However, these denoising approaches may result in suboptimal performance due to the mode mixing (i.e., manifestation of multiple IMFs within a single IMF) property of EMD and its sensitivity to noise and sampling. Essentially, the aforementioned shortcomings within EMD framework result in leakage of noise into a few signal modes which leads to their rejection resulting in suboptimal denoising. The lack of mathematical foundation of the EMD limits the chances of rectification of these issues within its framework.

The recently proposed variational mode decomposition (VMD) is based on optimization of a variational problem to obtain an ensemble of a fixed number of band limited IMFs (BLIMFs) [12]. Owing to its sound mathematical foundation, VMD successfully avoids mode mixing and is robust to noise and sampling unlike EMD [12,13]. From the view point of denoising, a very important feature of VMD is its ability to segregate the desired signal into a few initial BLIMFs while noise is mostly stashed into a few final BLIMFs. Hence, by rejecting the modes with noise, a good estimate of the true signal may be obtained by partial reconstruction.

A literature review shows that the existing VMD-based denoising approaches select relevant signal modes by comparing the probability distribution function (PDF) of an individual BLIMF against the PDF of the noisy signal. This is well founded because a distribution function is generally reflective of the signal present within the noisy data. An estimate of the signal present in a BLIMF may be obtained by measuring the closeness of its PDF with that of the noisy signal, for example, by employing Euclidean distance

[14], Bhattacharya distance [15], etc. Therein, the modes statistically close to the noisy signal are selected as relevant (signal) modes while largely dissimilar modes are rejected as noise. For a detailed study on the efficacy of various statistical distances for estimating relevant modes, the interested reader is referred to [16]. Moreover, the method in [17] selects relevant modes using the detrended fluctuation analysis (DFA) (originally used with EMD within the EMD-DFA method [11]) that estimates the randomness of data by observing the lack of trend. This method, hereafter referred as VMD-DFA [17], rejects the noisy BLIMFs and reconstructs the denoised signal based on the remaining modes.

The problem with these VMD-based methods is (a) the presence of noise within the selected (low frequency) signal modes that results in significant artifacts in the denoised signal and (b) the loss of high frequency signal details as part of the rejected modes leading to substandard denoising. This issue has not received much attention from researchers which may be because the interval thresholding techniques [10,18,19], widely used with EMD/MEMD, are not directly applicable to the VMD/MVMD modes owing to the arbitrary spread of noise power across the BLIMFs. On the other hand, the design of a wavelet like statistical thresholding technique for the BLIMFs of VMD requires prior knowledge of the distribution of noise (and/or signal). That is challenging because VMD, owing to its non-linearity, transforms a known noise distribution (e.g., Gaussian distribution) to an unknown one. Thus, discouraging the possibility of development of a statistical method for signal estimation.

To address this issue, we propose a novel statistical approach to signal denoising that exploits the nature of multiscale representation obtained using VMD to empirically estimate the distribution of noise residing in the BLIMFs of the noisy signal. This estimated noise distribution is then used to develop a nonlinear statistical signal estimation method using Cramer Von Mises (CVM) statistic from the BLIMFs of the VMD. Moreover, CVM statistic is also used to select the relevant signal modes which, in our case, involve the BLIMFs containing signal (with or without noise). That is a departure from the existing definition of relevant modes where only low frequency signal-only modes are selected to reconstruct an approximation of the original signal. Since our method selects all the modes containing signal, it yields a much closer estimate of the original signal because of its ability to reject noise from the selected (signal plus noise) modes using the CVM based thresholding function. We next list the main contributions of this work as follows:

- Development of an empirical procedure to estimate noise distribution model from within the noisy signal that is facilitated by the robustness of the VMD to noise and mode mixing that culminates in effective segregation of noise from the true-signal into a few final modes.
- Development of a signal estimation method based on the noise distribution that checks how closely the estimated noise distribution fits the local segments of the selected relevant signal BLIMFs.
- Use of the robust empirical distribution function (EDF)-based CVM distance for selecting the relevant signal modes ahead of the distances based on crude approximations of the probability density function (PDF).

To validate the performance of the proposed method, extensive computer simulations have been carried out for denoising a variety of benchmark signals corrupted by artificially generated Gaussian noise. Furthermore, the efficacy of the proposed method is demonstrated by denoising a (real) ECG signal corrupted by actual (non-Gaussian) sensor noise.

The rest of paper is organized as follows: Section 2 provides the background of the CVM statistic required to explain the proposed methodology that is subsequently presented in Section 3. Section 4 reports experiments analyzing the performance of our proposed method while Section 5 presents a few practical denoising examples. Finally, conclusion along with the future prospects of this work are discussed in Section 6.

2. Cramer Von Mises (CVM) statistic

CVM statistic [20] belongs to a class of statistical distances [21] that estimate how closely a dataset or observations follow a given distribution function. In this regard, the CVM statistic requires an estimate of the distribution of given observations, that is obtained using the EDF. It is worth mentioning that EDF happens to be a robust model of distribution even for small-sized data and is easy to compute [22]. More importantly, EDF is a discrete approximation of the cumulative distribution function (CDF) that means a distribution test is realized by testing how close an EDF of the data at hand is from the CDF of that reference distribution. This type of testing framework is termed as GoF test of a distribution on given dataset whereby EDF-based distances, e.g., Anderson Darling (AD) statistic [23], CVM statistic etc., are used to estimate the measure of fit of the reference CDF on the EDF of data at hand.

Given the CDF $E_0(z)$ corresponding to reference distribution and EDF $E(z)$ of given observations, the CVM statistic is given as follows

$$\Delta = \int_{-\infty}^{\infty} (E_0(z) - E(z))^2 d(E_0(z)), \quad (1)$$

where z denotes the support of the distribution function. Note that (1) involves an indefinite integration and is practically not computable. Therefore, its computable numerical adaptation is presented by D'Augustino in [22]

$$\Delta = \frac{1}{12L} + \sum_{t=1}^L \left(E_0(z'(t)) - \frac{(2t-1)}{L} \right), \quad (2)$$

where $z'(t)$ denotes a set of observations having finite length L , i.e., $t = 1, 2, \dots, L$.

The GoF test based on CVM statistic is realized by estimating the significance level or threshold λ that specifies the maximum value of the test statistic Δ (2) that is sufficient to suggest a *close-fit*. The GoF testing framework checks the following binary hypothesis:

$$\begin{aligned} \mathcal{H}_0 : \Delta &\leq \lambda, \\ \mathcal{H}_1 : \Delta &> \lambda, \end{aligned} \quad (3)$$

where $\mathcal{H}_0$ denotes the null hypothesis suggesting a *close-fit* of null (or reference) distribution on the given data while the $\mathcal{H}_1$ denotes the alternate hypothesis of *no-fit*. The threshold parameter λ is estimated by minimizing the probability of false alarm (P_{fa}), i.e., false detection rate of the alternate hypothesis $\mathcal{H}_1$ given the null hypothesis $\mathcal{H}_0$, and is mathematically stated as follows

$$P_{fa} = \text{Prob}(\mathcal{H}_1 | \mathcal{H}_0) = \text{Prob}(\Delta > \lambda | \mathcal{H}_0), \quad (4)$$

where $\text{Prob}(\cdot)$ denotes the probability of the event stated within the parenthesis. Here, the $P_{fa} = \alpha$, where α is kept very small, e.g., $\alpha = 10^{-2} - 10^{-4}$ to minimize the false detection of noise as signal [5,24].

3. Description of proposed approach

Consider the signal model

$$y(t) = x(t) + \psi(t), \quad (5)$$

where $y(t)$, $x(t)$ and $\psi(t)$ denote the noisy, true signal, and additive noise component, each of length N . It is customary to assume $\psi(t)$ being modeled as $\mathcal{N}(0, \sigma)$, i.e., zero-mean wGn process with variance σ^2 . However, the additive noise in the real-life signals may be non-Gaussian. Therefore, the existing denoising approaches developed on the assumption of wGn may have a limited scope for the practical signals. In this work, we propose to address the noise removal in practical signals using a robust multi-step procedure based on the VMD and CVM statistic. The main contribution of the proposed work is that it does not assume any prior model for the noise or signal. Instead, it employs a purely data-driven approach whereby noise distribution is estimated from within the noisy signal by exploiting its multiscale representation obtained using VMD. That is finally used within the proposed CVM based signal estimation (or noise rejection) framework from the selected VMD modes.

3.1. Variational mode decomposition of the noisy signal

VMD employs an entirely non-recursive approach to decompose the noisy signal $y(t)$, $\forall t = 1, \dots, N$, into K predefined modes BLIMFs $u_k(t)$, $\forall t = 1, 2, \dots, N$, and $k = 1, 2, \dots, K$. This is achieved by first finding the center frequencies w_k and then an ensemble of compact BLIMF by solving the following constrained variational problem [12]

$$\argmin_{\{u_k, w_k\}} \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|, \quad (6)$$

subject to

$$y(t) = \sum_{k=1}^K u_k(t), \quad (7)$$

where $\delta(t)$ denotes the Dirac distribution, and $*$ represents the linear convolution. In order to enforce the constraint that the aggregate of the total number of K modes $u_k(t)$, where $k = 1, \dots, K$, amount to the original signal; Lagrangian multipliers $\gamma(t)$ are used

$$\begin{aligned} \argmin_{\{u_k, w_k\}} \sum_{k=1}^K \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\| \\ + \left\| \sum_{k=1}^K u_k(t) - y(t) \right\|^2 + \left\langle \gamma(t), \sum_{k=1}^K u_k(t) - y(t) \right\rangle, \end{aligned} \quad (8)$$

where the quadratic data fidelity term $\left\| \sum_{k=1}^K u_k(t) - y(t) \right\|^2$ is used for its accelerated convergence and to ensure minimum squared error. This way, the center frequencies w_k required to find compact modes that successfully avoid mode mixing and are estimated by solving (8) using the alternating direction method of multipliers (ADMM). Further details of the algorithm can be found in [12].

VMD is naturally suited to denoising because it effectively segregates the true signal from noise such that signal details are mostly concentrated in a few initial BLIMFs. This is because of the well-posed variational problem in (6) that leads to the expansion of the noisy signal $y(t)$ as an ensemble of a set of K BLIMFs $\{u_k(t), \forall k = 1, \dots, K\}$ as given in (2). Given that $1 < k'_1 < k'_2 < K$, these BLIMFs may be largely divided into the following categories owing to the robust architecture within VMD to segregate signal and noise [16,17]:

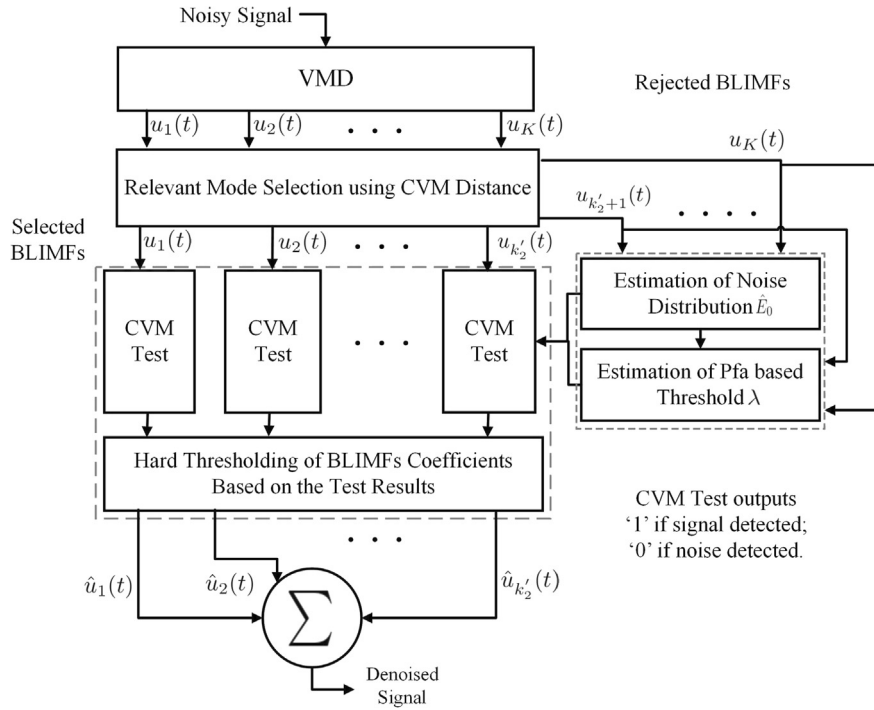


Fig. 1. Block diagram of the proposed denoising method.

- Initial modes $\{u_k(t), k < k'_1\}$ are composed of mostly signal content.
- Intermediate modes $\{u_k(t), k'_1 < k < k'_2\}$ mostly contain signal plus noise.
- Final modes $\{u_k(t), k > k'_2\}$ are composed of predominant noise.

The conventional VMD-based denoising approaches exploit this representation to perform the signal denoising whereby the modes with dominant signal, i.e., $\{u_k(t), k < k'_1\}$, are employed as relevant modes for a partial reconstruction of the denoised signal. The rest of the modes are simply rejected. This methodology results in a significant loss of the desired signal-details due to the rejection of intermediate (signal plus noise) modes, i.e., $u_k(t)$ for $k'_1 < k < k'_2$, along with the modes with dominant noise, i.e., $\{u_k(t) \text{ for } k > k'_2\}$.

In order to maximally preserve the desired signal information, the proposed framework only rejects the dominantly noise modes while the remaining modes are preserved as relevant signal modes using the CVM statistic. Hence, our definition of relevant modes includes the initial modes containing mostly signal and intermediate modes composed of both signal and noise. Subsequently, the selected relevant BLIMFs are cleansed of noise via a statistical thresholding function that operates by first estimating the noise distribution from the rejected noise modes which is then used to detect noise coefficients from selected modes using the CVM test. Finally, the denoised signal is partially reconstructed using the thresholded BLIMFs. This robust multistage procedure is depicted using the block diagram in Fig. 1 where each stage is explained in detail in the subsequent sections.

3.2. Relevant mode selection

This section details the process adopted to select BLIMFs containing signal information with or without noise, i.e., the relevant modes $\{u_k(t), k < k'_2\}$. This is achieved by detecting the BLIMFs entirely composed of noise, i.e., $\{u_k(t), k > k'_2\}$. In this regard, CVM distance based on EDF statistics is used to estimate the statistical distance between the noisy signal and the individual BLIMF.

3.2.1. Rationale

To identify relevant signal modes, conventional VMD denoising approaches investigate signal content in a BLIMF by computing some distance measure D'_k between the empirical PDFs of the noisy signal and the BLIMFs, as given below

$$D'_k = \text{Distance}\{p_y, p_{u_k}\}, \quad (9)$$

where p_y and p_{u_k} respectively denote the empirical PDFs of the noisy signal $y(t)$ and the k th BLIMF $u_k(t)$. The empirical PDFs p_y and p_{u_k} are estimated by dividing the data in hand into a finite number of bins leading to the construction of a PDF for these bins, e.g., use of kernel density function in [14–16]. The issue with this approach is that all the data elements in a bin are assigned the same probability as the probability of the bin where it resides. This compromises the individuality of the data points within the bin resulting in a less robust estimate of the distribution especially for small-sized data.

A robust estimate may be obtained by using the EDF (2) which is discrete approximation of the CDF of the data distribution. This makes EDF a robust estimator of data distribution even for small-sized data. Consequently, EDF is frequently used within the GoF-based hypothesis testing in various practical applications [24,25]. Therefore, we propose to use EDF-based distance to obtain a robust estimate of the distance between the noisy signal $y(t)$ and its BLIMFs $u_k(t)$

$$D_k = \text{Distance}\{E_y(z), E_{u_k}(z)\}, \quad (10)$$

where $E_y(z)$ and $E_{u_k}(z)$ respectively denote EDFs of the noisy signal and the k th BLIMF $u_k(t)$.

3.2.2. Estimating D_k using CVM statistic

In order to obtain an estimate $\hat{D}_k$ of the actual (statistical) distances D_k between the k th mode $u_k(t)$ and the noisy signal $y(t)$, both of size N , we use the CVM statistic as follows

$$\hat{D}_k = \frac{1}{12N} + \sum_{t=1}^N \left(\hat{E}_y(u_k(t)) - \frac{(2t-1)}{N} \right), \quad (11)$$

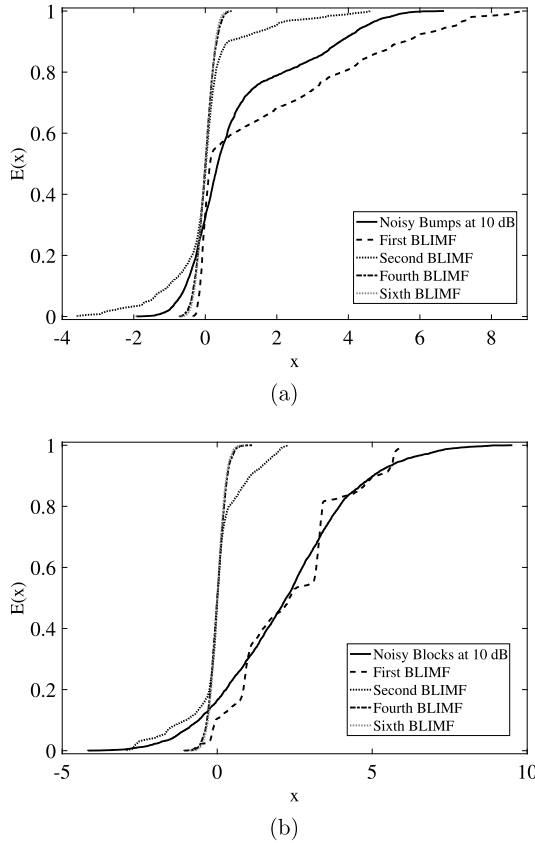


Fig. 2. Depiction of how CVM statistic estimates the closeness of the noisy signal from its BLIMFs. That is demonstrated by plotting EDFs of the noisy Bumps and Blocks signal along with EDF of their (selected) BLIMFs in (a) and (b) respectively.

where an estimate of the signal EDF $\hat{E}_y(z)$ is computed from the noisy signal $y(t)$ though

$$\hat{E}(z) = \frac{1}{N} \sum_{t=1}^N (y(t) \leq z). \quad (12)$$

The operation $(y(t) \leq z)$ results into a binary decision (i.e., 0 or 1) at every index t of the summation whereby, for a given z ; number of values of $y(t)$ less than or equal to z are accumulated.

In order to develop an insight on how CVM statistic estimates the distance between the noisy signal and the BLIMFs, consider Fig. 2 which plots EDFs of a few selected BLIMFs with the estimated reference EDF $\hat{E}_y(z)$ from the noisy signal. These results are obtained for the benchmark signals 'Bumps' and 'Blocks' (shown in Fig. 3(a) & (b) respectively). It is observed in Fig. 2, that the EDFs of first and second BLIMFs are closest to the EDF of noisy Bumps and Blocks signals that essentially means these initial modes are mostly signal. Contrarily, the EDFs of fourth and sixth BLIMFs are further from the reference EDF which means these higher modes have lesser signal content and more noise.

3.2.3. Criteria for selection of relevant modes

The above-detailed discussion indicates that the relevant signal modes may be selected by evaluating slopes of the distances between the consecutive BLIMFs [16,17]. Naturally, significant change in slope between the two adjacent BLIMFs means rapid decline of signal content when moving from earlier to the latter. This ensures that signal will decline further in the forthcoming BLIMFs with the increase in noise. Consequently, the existing methods [16,17] employ maximum slope in the distance-curve to determine a threshold, k'_1 , to select the relevant modes containing signal details

$$k'_1 = \underset{k}{\operatorname{argmax}} (S_k, S_{k+1}, \dots, S_K), \quad (13)$$

where S_k denotes the slope of the distances of k th and $(k+1)$ th BLIMFs, computed via

$$S_k = |D_{k+1} - D_k|. \quad (14)$$

Consider Fig. 4 which plots CVM distances of the modes of two benchmark (noisy) signals 'Bumps' and 'Heavy Sine' (shown in Fig. 3(c) & (d) respectively). It is seen from Fig. 4, that the CVM distances D_k corresponding to the 'Heavy Sine' signal show maximum slope between the first and second BLIMFs, and declines massively when moving to the second BLIMF and the subsequent ones. This essentially means that most of the signal content is concentrated in the first BLIMF. A similar observation can be made for the 'Bumps' signal. It is seen that the maximum slope is observed between the second and third BLIMF, and decreases rapidly in the latter modes. This shows that the signal content is largely concentrated in first two BLIMFs.

The above-detailed procedure, however, selects only dominantly signal modes as the relevant ones $\{u_k(t), k < k'_1\}$ for partial reconstruction of the denoised signal. By this definition, the rejected noise modes $\{u_k(t), k > k'_1\}$ include the intermediate signal plus noise modes $\{u_k(t), k'_1 < k < k'_2\}$ that causes loss of signal details. To address this issue, we suggest rejection of only purely noise modes $\{u_k(t), k > k'_2\}$ and retention of all the modes containing signal (with or without noise) $\{u_k(t), k < k'_2\}$. The selection of relevant modes according to new definition, i.e., modes containing signal (with or without noise) $\{u_k(t), k < k'_2\}$, requires estimation of mode index k'_2 that indicates the start of purely noise modes.

For this purpose, we alter the criteria discussed above by dividing the CVM distance curve (plotted in Fig. 4 for instance) into transient and stable regions where earlier relates to the purely signal modes while latter relates to the modes with noise. It is observed that the CVM distance plotted in Fig. 4 have transient phase before the maximum slope and the region after that can be categorized as the stable phase for all the BLIMFs. Since, the transient phase ends with maximum slope that can be seen as the threshold for selecting the signal only modes while an estimate of modes containing signal plus noise may be obtained from the stable phase. We suggest looking for maximum slope within the stable region (highlighted in red in Fig. 4, for instance) that indicates the point of maximum change between partially noisy modes to purely noisy modes. Understandably, maximum slope in the stable region separates the purely noise modes and the modes with signal. Based on the above discussion, maximum slope in the stable region may be obtained as follows

$$k'_2 = \max (S_k, S_{k+1}, \dots, S_K), \quad k = k'_1 + \dots, K \quad (15)$$

where k'_1 is obtained from (13) and k'_2 denotes the index of the mode that is followed by noise only modes. Mathematically, the thresholding criteria for selection of relevant modes and rejection noise only modes are then given below

$$\begin{cases} \{u_k(t), k \leq k'_2\} \in \text{Relevant modes with signal,} \\ \{u_k(t), k > k'_2\} \in \text{Rejected modes with noise.} \end{cases} \quad (16)$$

3.3. Estimation of noise distribution from the rejected modes

In this step, the distribution model governing noise within the VMD BLIMFs is estimated empirically from the rejected modes $\{u_k(t), k > k'_2\}$ which are mostly composed of noise. This step of the estimation of noise EDF is an essential part of the proposed

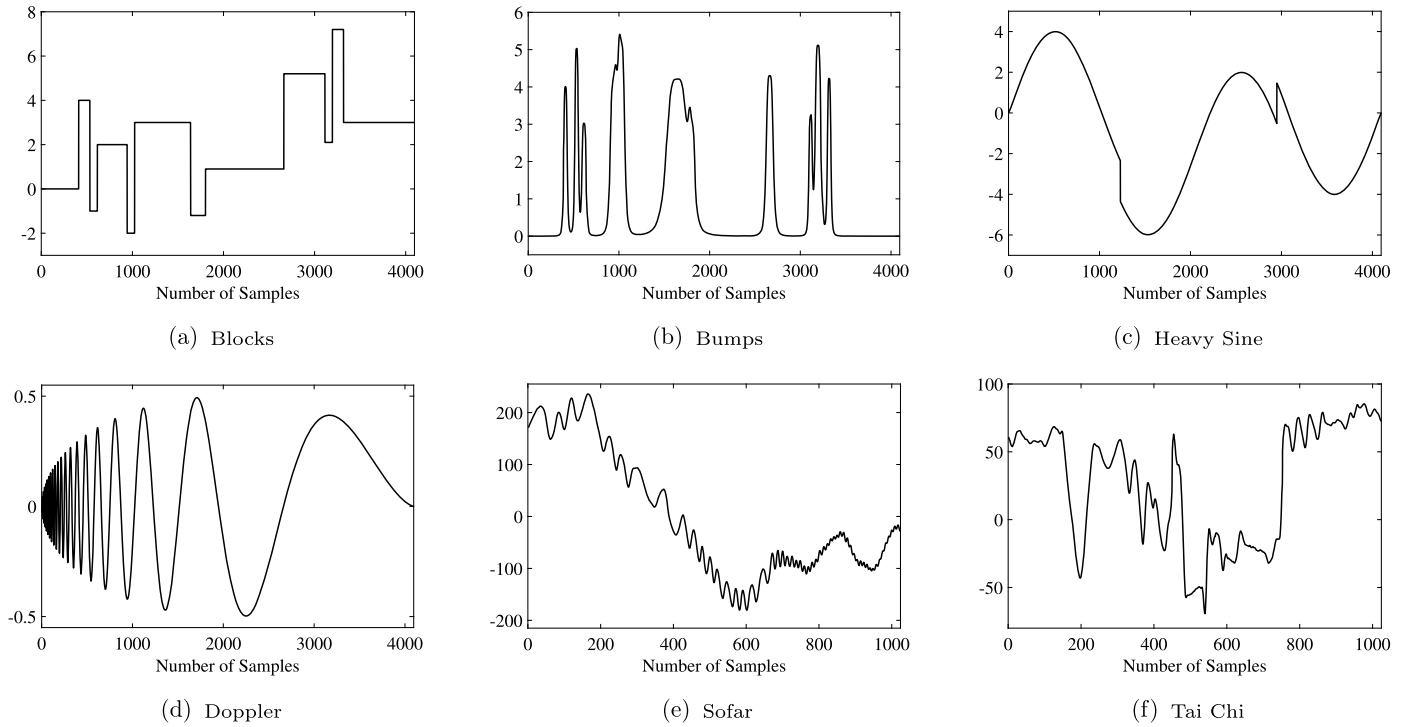
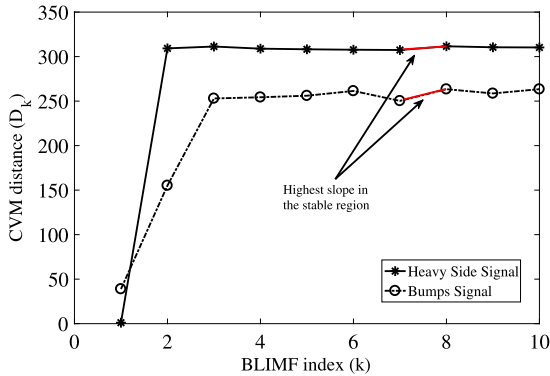


Fig. 3. Test signals used for experimentation in this work.

Fig. 4. An illustration of variation of CVM distance D_k for index k of the BLIMFs of a noisy 'Bumps' signal (dotted line) and 'Heavy Sine' signal (solid line). (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

approach as indicated by block 'CVM Test' in Fig. 1. The process adopted for empirical estimation of the noise distribution is illustrated in detail in Fig. 5 whereby, given a large sized dataset from an unknown distribution; a good estimate of its CDF may be empirically obtained using the ensemble average of the EDF of its local segments. That is a standard procedure used by statisticians to empirically estimate the unknown distribution function governing a given dataset [26–28].

Let $u_k^{(n)}(t) = \{u_k(t), k > k'_2\}$, $\forall t = 1, \dots, N$ denote the rejected noise BLIMFs to differentiate these noise modes from the relevant signal modes. First, we divide each rejected noise mode $u_k^{(n)}(t)$ into M non-overlapping segments/vectors $\mathbf{u}_{jk}^{(n)} = \{u_k^{(n)}(t) \mid t = j - \frac{L}{2}, \dots, j + \frac{L}{2}\}$ of equal size $L + 1$ and centered around an index j (depicted using boxes drawn on the rejected noise coefficients at the bottom left of Fig. 5). Next, the EDF $E_{jk}^{(n)}(z)$ of the noise segment $\mathbf{u}_{jk}^{(n)}$ from the k th rejected BLIMF $u_k^{(n)}(t)$, is computed as follows

$$E_{jk}^{(n)}(z) = \frac{1}{L+1} \sum_{t=1}^{L+1} (u_k^{(n)}(t) \leq z), \quad k > k'_2. \quad (17)$$

This way, EDFs of all of the non-overlapping segments are computed using (17). This step is depicted using the EDF blocks in Fig. 5.

Finally, by ensemble averaging the EDFs of all the segments from the rejected modes yields a close estimate $\hat{E}_0(z)$ of the actual noise CDF

$$\hat{E}_0(z) = \frac{1}{M(K - k'_2)} \sum_{k=k'_2+1}^K \sum_{j=1}^M E_{jk}^{(n)}(z), \quad (18)$$

where the accuracy of the estimate $\hat{E}_0(z) \approx E_0(z)$ increases with increase in number of segments, i.e., the length of the dataset. Fig. 5 plots the resulting estimate $\hat{E}_0(z)$ of the noise CDF $E_0(z)$ that is obtained from the rejected modes of the noisy signal corrupted by additive wGn at SNR = 10 dB.

3.4. Thresholding relevant modes using CVM test

This section describes the CVM statistic-based testing framework used to reject noise from the selected relevant modes $\{u_k(t), k \leq k'_2\}$. The aim here is to reject the coefficients corresponding to noise $\psi(t)$ without losing those corresponding to the true signal $x(t)$. In this regard, detection of noise coefficients from the selected modes is defined as a local hypothesis testing problem by selecting a local segment $\mathbf{u}_{jk} = \{u_k(t) \mid t = j - L/2, \dots, j + L/2\}$ of size $L + 1$, around each coefficient $u_k(j)$, from a selected BLIMF $\{u_k(t), k \leq k'_2\}$, as follows

$$\begin{aligned} \hat{\mathcal{H}}_0 : \mathbf{u}_{jk} &\in \psi(t), \\ \hat{\mathcal{H}}_1 : \mathbf{u}_{jk} &\in x(t), \end{aligned} \quad (19)$$

where $\hat{\mathcal{H}}_0$ and $\hat{\mathcal{H}}_1$ respectively denote the null and alternate hypothesis of our VMD-based denoising approach.

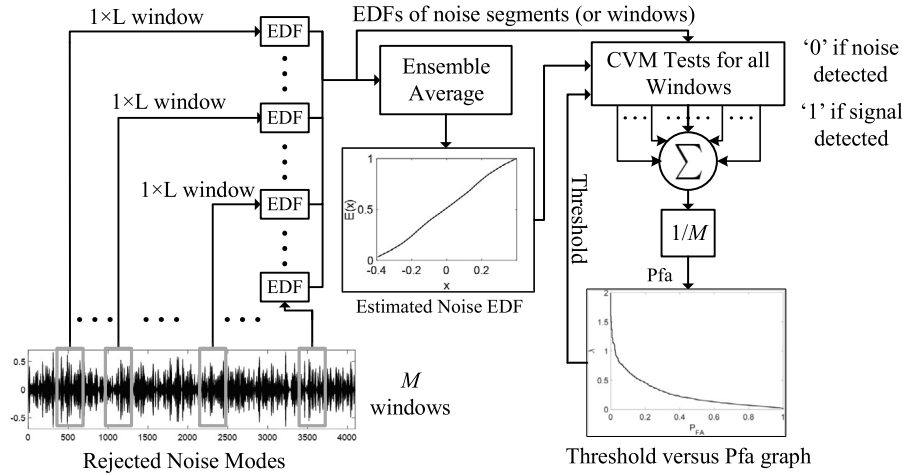


Fig. 5. Depiction of the procedure adopted for estimation of noise EDF and threshold selection.

In order to test the hypothesis given in (19), i.e., to check the possibility that $\mathbf{u}_{jk} \in \psi(t)$; the EDF $E_{jk}(z)$ of the local segment $\mathbf{u}_{jk}$ is computed based on (17) and then the GoF of $E_{jk}(z)$ is evaluated/tested on the estimated noise EDF $\hat{E}_0$, where

$$\begin{aligned} \mathcal{H}_0 : E_{jk}(z) \sim \hat{E}_0(z) &\Rightarrow \Delta_{jk} \leq \lambda_k, \\ \mathcal{H}_1 : E_{jk}(z) \not\sim \hat{E}_0(z) &\Rightarrow \Delta_{jk} > \lambda_k. \end{aligned} \quad (20)$$

Here, the symbol $\sim$ denotes the *close-fit* and $\not\sim$ denotes *no-fit* of EDFs which is decided based on the value of the CVM distance Δ_{jk} between $E_{jk}(z)$ and $\hat{E}_0(z)$ computed using (2). To achieve that, a threshold λ_k is estimated such that the false detections of null hypothesis $\mathcal{H}_0$, referred to as false alarms, are minimized. In essence, λ_k indicates the maximum possible value of distance Δ_{jk} required to suggest a *close-fit* between the two EDFs. Therefore, a *close-fit* $E_{jk}(z) \sim \hat{E}_0(z)$ is detected when Δ_{jk} is within the specified bounds, i.e., $\Delta_{jk} \leq \lambda_k$. On the other hand, the case of *no-fit* $E_{jk}(z) \not\sim \hat{E}_0(z)$ is obtained when the distance exceeds the specified bound of threshold, i.e., $\Delta_{jk} > \lambda_k$.

3.4.1. Selection of threshold based on rejected noise modes

Within GoF tests, the threshold or critical value λ_k is selected for each mode k that serves as an upper bound on the CVM distances Δ_{jk} . Generally, λ is selected for very small value of P_{fa} (4) which ensures least false rejections of the null hypothesis $\mathcal{H}_0$. That means false detections of *no-fit* (i.e., $\mathcal{H}_1$) when the reference EDF actually fits the given data samples (i.e., $\mathcal{H}_0$) are minimized. In the context of the testing problem (21), very small P_{fa} means threshold is selected such that it minimizes the false detection of noise (i.e. $\hat{\mathcal{H}}_0$) as true signal (i.e. $\hat{\mathcal{H}}_1$). Hence, this requirement of minimum P_{fa} fits right in the denoising problem since the goal in denoising is to ensure maximum noise is removed which can be achieved by minimizing the P_{fa} while attempting to maximize the preservation of the true signal.

Conventionally, threshold selection is performed by accumulating the probabilities of false detection, i.e., $P_{fa} = \text{Prob}(\Delta_{jk} > \lambda | \hat{\mathcal{H}}_0)$. Given the PDF $p(u_k^{(n)}(t))$ of rejected noise modes $u_k^{(n)}(t)$, P_{fa} is defined as follows

$$\begin{aligned} P_{fa} &= \text{Prob}(\hat{\mathcal{H}}_1 | \hat{\mathcal{H}}_0) \\ &= \int_{\{u_k^{(n)}(t); \Delta_{jk} > \lambda | \hat{\mathcal{H}}_0\}} p(u_k^{(n)}(t)) du_{jk}^\psi(t), \end{aligned} \quad (21)$$

where $\{u_k^{(n)}(t); \Delta_{jk} > \lambda | \hat{\mathcal{H}}_0\}$ are noise coefficients from the rejected BLIMFs $\{u_k(t), k > k'_2\}$ yielding false alarms, i.e., falsely suggesting signal detection.

An estimate $\hat{p}(u_k^{(n)}(t))$ of the noise PDF $p(u_k^{(n)}(t))$ may be obtained by computing the derivative of the empirically estimated noise CDF $\hat{E}_0(t)$ (18)

$$\hat{p}(u_k^{(n)}(t)) = \frac{d\hat{E}_0(z)}{dt}. \quad (22)$$

Consequently, an empirical adaptation of (21) to our denoising problem may be obtained through

$$P_{fa} = \int_{\{u_k^{(n)}(t); \Delta_{jk} > \lambda | \hat{\mathcal{H}}_0\}} \frac{d\hat{E}_0(z)}{dt} du_k^\psi(t). \quad (23)$$

Since, the detection of the range of coefficients $\{u_k^{(n)}(t); \Delta_{jk} > \lambda | \hat{\mathcal{H}}_0\}$ is central to the computation of P_{fa} using (24) for a given threshold λ_k . The proposed empirical approach estimates the coefficients $\{u_k^{(n)}(t); \Delta_{jk} > \lambda\}$ from the rejected noise modes $\{u_k(t), k > k'_2\} \in \psi(t)$.

To begin with, a range of candidate thresholds are selected and the noise coefficients within each rejected mode $u_k^{(n)}(t)$ are divided into M windows having size $L + 1$. Next, each candidate threshold λ is used within the CVM test for applying it on the coefficients from each window. Therein, CVM statistic Δ_{jk} between the EDF $E(z)$ of a window of noise coefficients and the reference noise EDF $\hat{E}_0(z)$ is computed through (2) followed by hypothesis testing based on (20). For each λ , probability of false alarm P_{fa} is computed by recording the instances of erroneously detecting noise segments (from rejected modes) as signal and then dividing the accumulated false alarms by total number of windows M .

This way, a threshold versus P_{fa} table is estimated by computing the P_{fa} for all the candidate thresholds λ . The relationship between the threshold λ and the P_{fa} , obtained from the rejected BLIMFs of a given input noisy signal having SNR = 10 dB, is graphically shown in Fig. 5 (bottom right). In general, higher number of false alarms (i.e., higher P_{fa}) are observed for lower thresholds but with increase in threshold value, the P_{fa} decreases.

Afterwards, a threshold is selected for a given P_{fa} from the estimated threshold λ versus P_{fa} table. Here, it is important to consider the trade-off between the P_{fa} and the probability of true-signal detection P_d when selecting a threshold for noise reduction. Generally, P_d also decreases with decrease in P_{fa} , i.e., signal is falsely (or erroneously) rejected as noise for lower P_{fa} , see [29]

for more insight into this matter. Therefore, to avoid loss of signal from the initial BLIMFs (which are mostly composed of signal), higher P_{fa} is selected to keep P_d high as well, i.e., signal is not falsely rejected as noise. On the other hand, lower P_{fa} is chosen for latter modes (with dominant noise component) to reject maximum noise. To that end, a scale adaptive separate threshold λ_k is selected for each of the relevant mode $u_k(t)$ using the following decaying function

$$P_{fa}^{(k)} = e^{-k+1}, \quad (24)$$

where $P_{fa}^{(k)}$ denotes the false alarm probability of the k th mode. The decaying function (24) assigns higher P_{fa} to initial signal modes (i.e., to recover maximum signal) and smaller P_{fa} to latter noise BLIMFs (i.e., to reject maximum noise).

3.4.2. Thresholding function

Traditional hard-thresholding function detects noise coefficients based on their smaller amplitudes via a threshold that is well adapted to wavelet denoising. However, the use of a thresholding function that detects noise based on amplitude difference among coefficients is not properly motivated for VMD denoising. This is because VMD denoising methods do not exploit sparsity of multiscale decomposition, instead it estimates the local trend to detect the signal content from noise [10,17]. The proposed approach works on the same principle of estimating local trend in a BLIMF to detect and reject noise. In this work, the local trend is estimated using the EDF $E_{tk}(z)$ of local segment u_{tk} that is selected around each coefficient at location t . Subsequently, it is checked whether the $E_{tk}(z)$ is close to that of noise $\hat{E}_0(z)$ by estimating the (CVM) distance Δ_{tk} between the two EDFs using (11) and then it is compared against the threshold λ_k

$$\hat{u}_k(t) = \begin{cases} 0 & \text{if } \Delta_{tk} \leq \lambda_k, \\ u_k(t), & \text{if } \Delta_{tk} > \lambda_k. \end{cases} \quad (25)$$

Finally, the denoised signal is reconstructed from the thresholded relevant BLIMFs $\{\hat{u}_k, k < k'_2\}$, as follows

$$\hat{x}(t) = \sum_{k=1}^{k'_2} \hat{u}_k(t), \quad k < k'_2, \quad (26)$$

where $\hat{x}$ is an estimate of the true signal $x(t)$ obtained using the proposed approach. In the rest of the paper, we will refer to the proposed approach as **VMD-CVM**.¹

4. Simulation results and discussion

This section demonstrates the effectiveness of the performance of the proposed VMD-CVM method against the state of the art in signal denoising. In this regard, the used comparative state of the art methods includes a good mix of wavelet and data driven decomposition (i.e., EMD and VMD) based methods, given as follows:

- DTCWT-Thr [30]: Employs neighborhood based thresholding along with the quasi-translation invariant DTCWT.
- BLFDR [8]: Performs noise shrinkage by employing the false discovery rate (FDR) [31] within the Bayesian framework.
- EMD-IT [10]: Performs interval thresholding on the EMD modes of noisy signal to reject noise.

- EMD-DFA [11]: Performs partial reconstruction by rejecting the EMD modes exhibiting detrend (i.e., randomness) using the DFA.
- VMD-DFA [17]: Performs partial reconstruction by rejecting the VMD-BLIMFs exhibiting detrend (i.e., randomness) using the DFA.
- DWT-GOF [5]: Tests the normality DWT coefficients to detect and reject noise.
- DT-GoF-NeighFilt [6]: Exploits the quasi-translation invariance of the DTCWT using the normality test and neighborhood classification based filtering for effective noise removal.

The quantitative evaluation of performance of the comparative methods has been carried out using the signal-to-noise ratio (SNR) and mean squared error (MSE). On the contrary, the qualitative analysis has been performed by visually demonstrating the closeness of the denoised signal from the original one.

The test datasets employed within experiments include both real and synthetic signals. Among those, we employed Donoho's 'Blocks', 'Bumps', 'Heavy Sine', and 'Doppler' as synthetic signals which are respectively plotted in Fig. 3(a)-(d). The real signals include 'Sofar' (Fig. 3(e)) that records the oceanographic float drift of the water flowing through Mediterranean sea and 'Tai-Chi' (plotted in Fig. 3(f)) that tracks the human body movements in a Tai-Chi sequence using a 3D sensor attached to the ankles.

The reported experiments study and analyze different aspects of the proposed method when compared against the state of the art is signal denoising. For this purpose, noisy signals are generated by adding wGn (at varying input SNRs) to the (clean) test signals shown in Fig. 3. Those are subsequently denoised using the comparative methods and resulting quantitative measures of performance (i.e., SNR and MSE) are reported in tabular and graphical forms. Within these experiments, the parameters concerning the state of the art methods are set in accordance with the recommendations made by the authors in their corresponding articles for their best performance. In case of proposed method, we use window length $L = 32$ for local application of the CVM test on the BLIMFs of VMD where the number of decomposition levels/modes K using VMD are chosen to be $K = 10$. This choice of K has been finalized based on an experiment reported below where the impact of the number of decomposition levels of VMD has been studied on the denoising performance of our method.

4.1. Number of modes versus output SNR

The selection of the number modes K is a critical problem in VMD based applications since it impacts the overall outcome of the algorithm. Consequently, we study the impact of the choice of number of modes K on the denoising performance of the proposed VMD-CVM method. Moreover, we also compare the performance of our method against the state of the art VMD-DFA method for each choice of K . The presented experiment performs denoising of the noisy signals at input SNR = 0 & 10 dB for each $K = 4, \dots, 12$. In this experiment, we used two input signals namely Bumps and Sofar signals and the resulting output SNRs are respectively plotted in Fig. 6(a & b) for each choice of K . Note that the output SNRs reported in this experiment are the mean of $J = 20$ iterations.

Observe from solid lines in Fig. 6 that the choice of number of modes K does not significantly impact the performance of the proposed method. Albeit, a slight increase in the output SNR (corresponding to the 10 dB input SNR) is visible in Fig. 6(a) when noisy signal is decomposed into a large number of modes. That may be due to the fact that the intermediate modes can be detected accurately when noisy signal is decomposed into higher number of BLIMFs. Consequently, we use $K = 10$ in our method for the experiments reported below.

¹ For reproducibility of research, the MATLAB code for the proposed method is available online at <https://data.mendeley.com/datasets/g2wv44fsb3/draft?a=6a07a1d3-a4e0-4b05-80e6-0eedbbcd4080>.

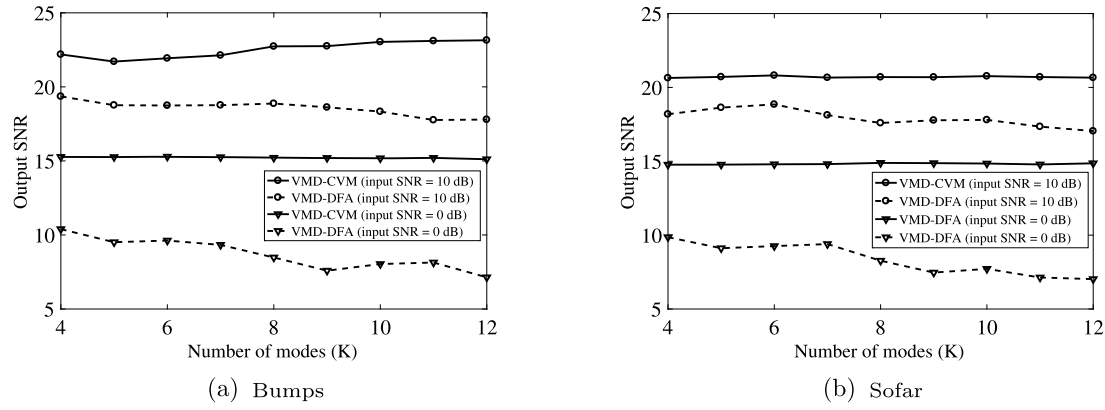


Fig. 6. Depiction of how output SNRs of two VMD based methods (i.e., proposed VMD-CVM and VMD-DFA) vary with change in number decomposition levels or modes k for (a) Bumps signal and (b) Sofar signal at input SNR = 0 and 10 dB respectively.

Table 1

Output SNR values obtained from various denoising methods for varying input SNR levels.

Input signal		Blocks				Bumps				Sofar				Tai-Chi			
Input SNR		-5	0	5	10	-5	0	5	10	-5	0	5	10	-5	0	5	10
EMD-DFA	SNR	-2.00	2.83	6.39	7.52	2.78	7.77	12.85	17.11	-1.43	3.24	8.41	13.88	-1.38	2.97	8.79	13.46
	MSE	13.96	4.64	2.01	1.55	2.78	7.77	12.85	17.11	2.01e4	0.68e4	0.21e4	0.05e4	0.36e4	0.13e4	0.03e4	0.01e4
VMD-DFA	SNR	2.77	7.67	12.71	17.31	2.95	8.11	12.74	18.56	2.86	7.66	12.72	17.89	2.80	7.43	12.98	18.19
	MSE	4.65	1.50	0.47	0.16	1.65	0.50	0.17	0.04	0.75e4	0.24e4	0.07e4	0.02e4	1.37e3	0.47e3	0.13e3	0.04e3
EMD-IT	SNR	5.38	9.90	14.57	18.88	6.06	10.29	15.19	19.83	4.53	9.31	13.56	17.43	3.73	8.53	12.84	17.29
	MSE	2.56	0.90	0.30	0.11	0.80	0.30	0.09	0.03	0.52e4	0.17e4	0.06e4	0.02e4	1.15e3	0.36e3	0.13e3	0.04e3
DTCWT-Thr	SNR	3.37	10.71	15.36	19.01	6.14	10.95	16.12	21.05	-5.05	0.11	4.92	10.09	-4.97	0.06	5.13	10.75
	MSE	4.05	0.74	0.25	0.11	0.78	0.26	0.07	0.02	0.46e5	0.14e5	0.04e5	0.01e5	0.82e4	0.25e4	0.08e4	0.02e4
BLFDR	SNR	-0.62	5.33	12.09	17.88	-0.19	6.26	12.90	19.58	1.59	8.26	13.81	18.16	1.16	7.35	12.99	18.17
	MSE	10.18	2.58	0.54	0.14	3.39	0.76	0.16	0.03	1.01e4	0.41e4	0.06e4	0.02e4	2.00e3	0.48e3	0.13e3	0.03e3
DWT-GoF	SNR	5.08	9.37	13.92	17.83	7.02	11.50	16.40	20.98	4.56	9.66	14.21	18.95	4.20	8.88	12.80	15.37
	MSE	2.73	1.01	0.35	0.14	0.64	0.23	0.07	0.02	0.50e4	0.15e4	0.05e4	0.01e4	1.00e3	0.33e3	0.13e3	0.07e3
DT-GOF-NeighFilt	SNR	11.20	15.30	17.10	19.27	10.73	15.16	18.52	23.28	10.91	14.25	16.59	17.99	9.78	12.42	15.41	18.29
	MSE	0.66	0.26	0.13	0.08	0.25	0.09	0.04	0.01	1.18e3	0.54e3	0.31e3	0.22e3	0.27e3	0.15e3	0.07e3	0.03e3
Proposed VMD-CVM	SNR	11.30	15.22	17.21	19.86	10.92	15.38	18.49	23.19	10.28	14.67	17.35	20.19	10.04	12.58	16.19	19.38
	MSE	0.65	0.26	0.13	0.07	0.25	0.08	0.04	0.01	1.35e3	0.49e3	0.26e3	0.14e3	0.25e3	0.14e3	0.06e3	0.03e3

In this experiment, we also demonstrate the impact of selection of intermediate modes and the efficacy of the CVM based signal extraction framework from these modes which are otherwise rejected as noise. For this purpose, we compare the denoising performance of our VMD-CVM method against the recent VMD-DFA method. Clearly, the proposed VMD-CVM method outperforms the VMD-DFA method by a significant margin for all choices of K . Moreover, the VMD-DFA is highly sensitive to the choice of number of modes K and its performance deteriorates with increase in K . That may be attributed to its crude methodology where intermediate modes are rejected as noise.

4.2. Input SNR versus output SNR

Table 1 reports output SNR and MSE values of the denoised signals obtained by employing the comparative methods on the noisy signals at varying input SNRs, i.e., = -5 dB, 0 dB, 5 dB and 10 dB. In this regard, we use the synthetic 'Bumps' and 'Blocks' signals (of 2^{12} sample size) and real 'Sofar' and 'Tai Chi' signals (of 2^{10} sample size) as input datasets. Moreover, each of the reported output SNR and MSE values in the Table 1 are mean of $J = 20$ realizations whereby highest output SNR and the corresponding MSE values are highlighted in bold.

It can be seen from the Table 1 that the proposed VMD-CVM method outperforms the state of the art methods for almost all

input signals, except for a few cases where the DT-GOF-NeighFilt yields better performance than the proposed VMD-CVM. The rest of the data-driven denoising methods based on VMD or EMD fall behind the wavelet denoising methods used in this study. This superior performance of the proposed method owes to the robust multistage procedure that recovers signal within the noisy modes which are rejected as noise in other VMD-based methods.

An important observation in Table 1 is the highest output SNR values of the proposed VMD-CVM for the 'Blocks' signal despite its piece-wise constant nature. This is significant because VMD/EMD-based denoising methods conventionally fail to extract the details of piece-wise constant signals. This result demonstrates the efficacy of the robust architecture within the proposed VMD based approach. Moreover, for the 'Sofar' and 'Tai Chi' signals with a fair bit of complexity due to sharp and subtle variations, the proposed VMD-CVM yields highest output SNRs for input SNR ≥ 0 dB. Here, the margin of difference between the SNR values from proposed method and the DT-GOF-NeighFilt is large enough to be more than 5% of the best. This result again demonstrates the efficacy and robustness of our method for complex signals.

4.3. Signal length vs. output SNR

Signal denoising is a lossy process that means a fraction of the sample points corresponding to true signal are rejected as noise.

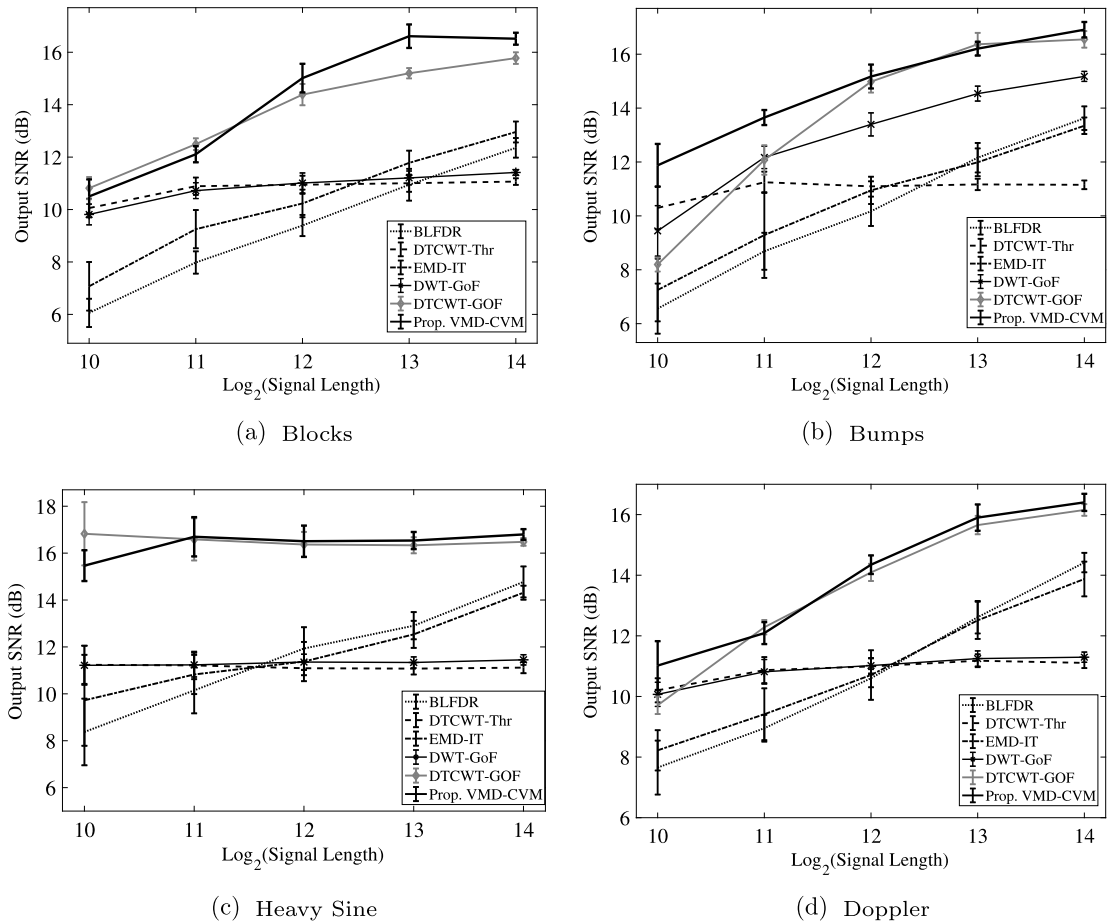


Fig. 7. Error bar plots depicting output SNR levels obtained from different denoising methods for (a) Blocks, (b) Bumps, (c) Heavy sine, (d) Doppler, at input SNR = 0 dB and varying signal lengths $N = 2^{10} \rightarrow 2^{14}$.

That can be a problem because, in case of low sampling rate, loss of signal samples may result in poor reconstruction of the denoised signal. On the contrary, at higher sampling rate (or large sample size) loss of a few sample points may not be an issue due to denoising owing to the fact that more number of samples are available to represent the same information now. Hence, change in sampling rate (or signal length) is an important parameter to analyze the performance of denoising methods.

Consequently, in this experiment, we demonstrate efficacy of the proposed VMD-CVM method against the state of the art methods with varying lengths $N = 2^{10}$ to 2^{14} . In this regard, we use all four of the synthetic signals which are corrupted by noise at input noise of 0 dB. The comparison of the resulting output SNRs is presented graphically using error-bar plot as shown in Fig. 7. That not only displays the mean of the output SNRs over $J = 20$ iterations but also gives an estimate of maximum variation in SNR values during these iterations.

Fig. 7(a)-(d) respectively plots the denoising results for Blocks, Bumps, Heavy Sine and Doppler signals where it can be observed that proposed method generally outperforms the state of the art methods irrespective of the signal length. The proposed VMD-CVM and DT-GoF-NeighFilt show out of the box performance when compared with other state of the state of the art methods. Therein, the VMD-CVM method consistently beats the DT-GoF-NeighFilt method except at a few instances of lower sample rate or smaller signal length, see Fig. 7(a),(c). However, this is not a pattern because, for other two input signals, proposed method yields much better results for lower signal lengths as well. Hence, this trend in

results does not undermine the efficacy of our method and may be attributed to the nature of input signal under study.

4.4. Qualitative performance analysis

The qualitative analysis demonstrates how closely the denoised signals from various methods resemble their corresponding true or noise-free signals. Generally, this is shown by plotting the denoised signals along with the original (noise-free) one that enables the reader to visualize how well the denoising methods extract signal-details from the noisy signal. To that end, we plot denoised 'Bumps' and 'Taichi' signals along with the original ones respectively in Fig. 8 and Fig. 9 whereby the corresponding noisy signals are also shown for comparison. We compare the visual results of the proposed VMD-CVM method against the top comparative state of the art methods namely BLFDR, EMD-IT, VMD-DFA, GoF-DWT and DT-GOF-NeighFilt. Denoised signals were obtained by respectively denoising the noisy 'Bumps' (shown in Fig. 8(a)) and noisy 'Tai Chi' signal (shown in Fig. 9(a)) where noisy version of input SNR = 10 dB is shown in gray while true signal is shown in dark black.

It can be observed from the Fig. 8 that the proposed VMD-CVM method yields best estimate of the original signals. The denoised 'Bumps' signals from DWT-GoF and DT-GOF-NeighFilt yield very close estimate of the original signals as can be seen from Fig. 8(c & d) but both these methods suffer from artifacts. Thus overshadowing their efficiency of extracting signal details. On the contrary, the proposed VMD-CVM yields an equally close estimate of the original signal but without artifacts, see Fig. 8(e). More visible (undesired) spikes are found in the denoised 'Bumps' signals by EMD-IT, see

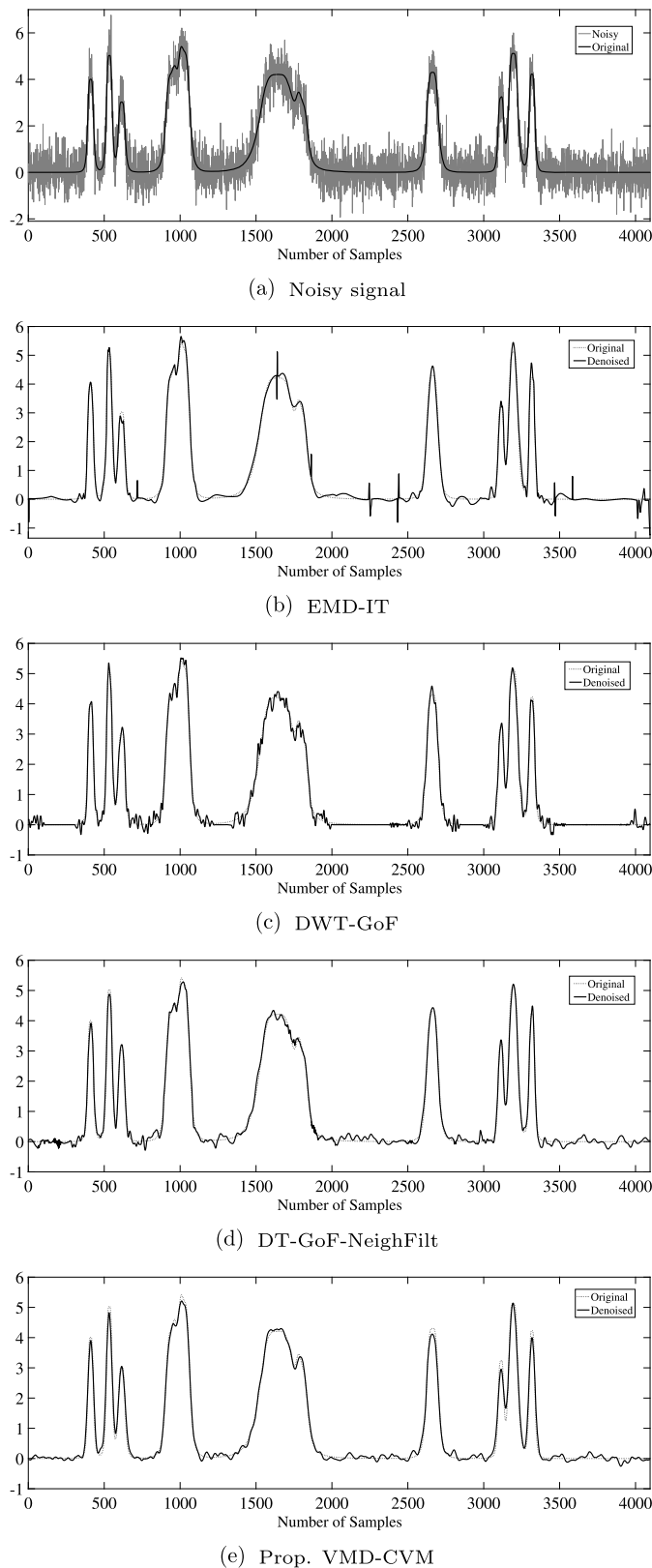


Fig. 8. Visual comparison of denoised signals from various methods for 'Bumps' signal at input SNR = 10 dB and signal lengths $N = 2^{12}$.

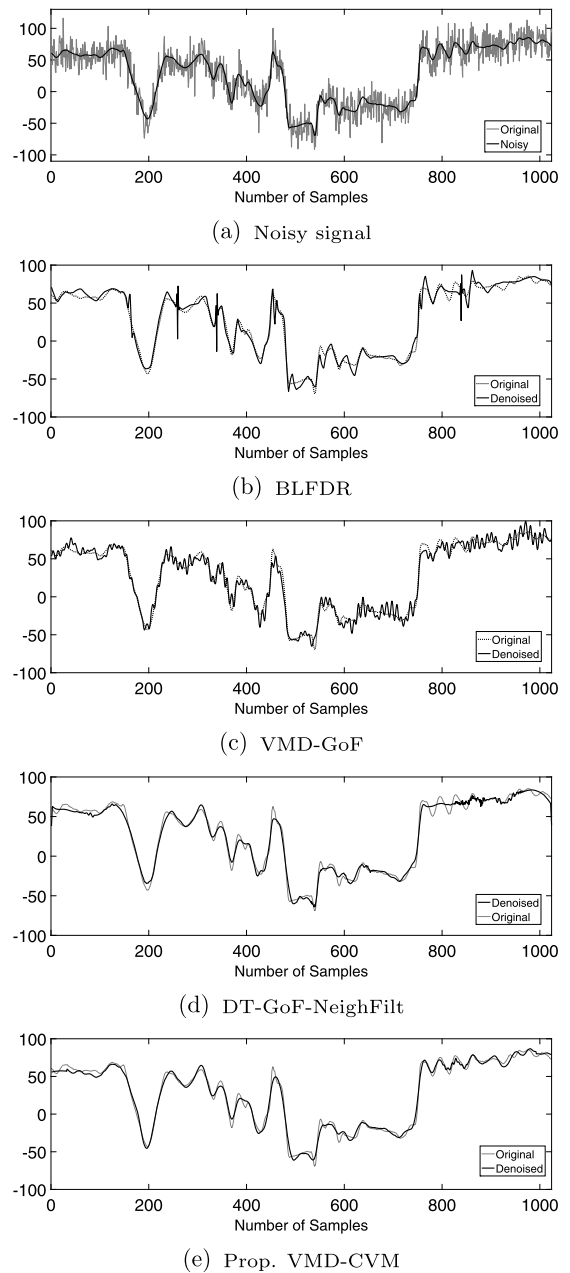


Fig. 9. Visual comparison of denoised signals from various methods for 'Tai Chi' signal at input SNR = 10 dB and signal lengths $N = 2^{10}$.

from Fig. 8(b), whereby the overall quality of the denoised signal has been adversely effected by artifacts.

The denoised versions of the 'Tai Chi' signal from the comparative methods are plotted in Fig. 9(b)-(d) at input SNR = 10 dB. A striking result here corresponds to the VMD-DFA that completely fails to recover the peaks and highly varying parts of the signal and instead introduces undesired oscillations. That is owing to the fact that VMD-DFA only uses initial few modes for reconstruction of the denoised signal while discarding high frequency signal parts as part of rejected modes. Another factor, in this regard, may be the complex structure of the 'Tai Chi' signal that is composed of subtle variations with higher frequencies which are challenging to extract in presence of random noise. A better estimate of the true signal is obtained by the DT-GOF-NeighFilt that largely recovers the variations while doing away with noise. Though, it fails to capture the subtle variations specially in the last half of the denoised signal, see Fig. 9(e). The best estimate of the 'Tai Chi' signal is

obtained by the proposed VMD-CVM that captures the subtle variations throughout the signal as can be observed from see Fig. 9(f). Apart from the sharp changes situated in the middle of this signal, the proposed method recovers all the details of the real 'Tai Chi' signal demonstrating its effectiveness for complex real world signals.

Furthermore, the VMD-DFA yields exaggerated variations as artifacts in the aftermath of denoising process, see Fig. 9(c). That is owing to its partial reconstruction nature where relevant modes were selected to reconstruct the denoised signal and the presence of noise within the selected modes was ignored. It is observed from Fig. 9(e) that the proposed approach does not suffer from this issue because our method performs thresholding on the selected relevant modes to reject the coefficients exhibition noise-like-statistics. Consequently, the reconstruction of the denoised signal based on cleansed thresholded BLIMFs successfully avoids the artifacts otherwise seen the results of the VMD-DFA method.

5. Denoising ECG signal corrupted by sensor noise

In this section, we present denoising results of the proposed method when applied to an ECG signal corrupted by the sensor noise. The raw ECG signal in this regard is taken from [32] that is corrupted by actual sensor noise (during acquisition) with unknown distribution model. To further exaggerate the noise we added an additive wGn of at input SNR = 20 dB to account for the thermal noise component due to electronic devices. As a result, the noise mostly obscures the useful information within the subtle variations of the ECG signal as can be observed from Fig. 10(a) where the noisy ECG signal is plotted in gray along with a clean model of the raw ECG signal. The clean version is also available in [32] for use as the ground truth.

To address this challenging problem we used the proposed method owing to its ability to estimate the distribution of noise/artifacts from within the noisy signal. The resulting denoised signal is plotted in Fig. 10(b) (in dark black) where clean signal is also shown (in gray) in the background the denoised signal. Here, the effectiveness of our method is shown by demonstrating how closely the denoised version follows the clean ECG signal. Evidently, from Fig. 10(b), the denoised version closely follows the clean signal in the background because it recovered important details including sharp peaks and slower variations. Despite the presence of the noise artifacts, observed near the sharp peaks, the overall quality of the recovered ECG remains intact verifying the efficacy of the proposed method in suppressing sensor noise while retaining the subtle variations which were previously hidden in the sensor noise.

6. Conclusions and discussion

In this paper, we have addressed the problem of noise removal from the practical signals whereby the noise is considered to be governed by unknown probability distribution. We propose to exploit the desirable properties of VMD to estimate the EDF of noise from within the noisy signal. As stated earlier, VMD possesses ability to segregate the signal and noise in separate group of BLIMFs owing to its robustness to noise and mode mixing. First, we detect the group of BLIMFs predominantly composed of noise using CVM statistic followed by the empirical estimation of noise EDF from these rejected modes. Subsequently, the estimated distribution is used as a means to detect and reject the noise coefficients (the coefficients fitting the estimated noise EDF) from the remaining modes. The estimation of GoF of the reference noise EDF on the local segment has been performed by the CVM-GoF test.

The effectiveness of the proposed method has been demonstrated by comparing its performance against the state of art meth-

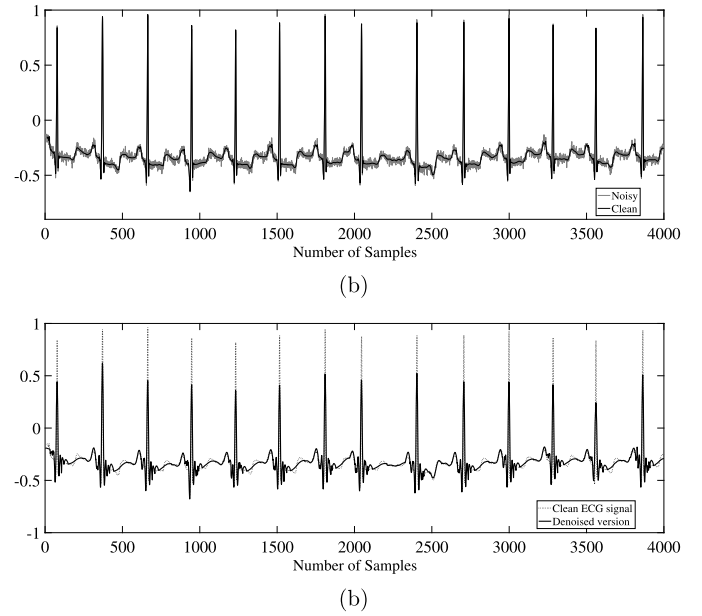


Fig. 10. Denoising results on a raw ECG signal obtained from [32]. The subfigure (a) shows the noisy ECG signal (in gray) along with its clean model (in black); while the subfigure (b) shows the denoised version by the proposed VMD-CVM method (in black) along with the clean signal model as reference (in gray).

Table 2

Computational time (in seconds) required to complete one iteration of the selected denoising methods.

	Sofar (of size 2^{10})	Bumps (of size 2^{12})
EMD-DFA	6.21	29.48
VMD-DFA	9.82	42.99
DT-GoF-NeighFilt	1.64	3.88
Proposed VMD-CVM	11.11	47.17

ods. It has been observed that the proposed method comprehensively outperformed the rest of the methods considered in this paper. This improvement in denoising performance by proposed method has not been achieved at too much increase in the computational cost when compared to the existing VMD/EMD-based methods. In fact, as can be observed from Table 2, the computation time of the proposed denoising methods on MATLAB, when running on a Core-i7 HP-Spectre-360X computer, is nearly at par with existing VMD or EMD based methods. The slight increase in the computation time in our VMD-CVM method may be explained by the additional CVM based data driven signal estimation process and consideration of intermediate modes (both of which are not considered in EMD-DFA and VMD-DFA). The DT-GoF-NeighFilt yields much smaller execution time because of the built-in optimized implementation of the wavelet transforms within MATLAB and smaller number of samples per scale (due to decimation operations) when compared to VMD or EMD based denoising methods.

Moreover, we also demonstrated the efficacy of the proposed method in addressing the problem of removal of sensor noise governed by some unknown distribution. For this purpose, we took the example of ECG signals (corrupted with sensor noise) where it has been shown that the proposed method successfully removes the sensor noise in ECG signals. The future prospects of this work may include its use in other practical applications for removal of noise before the signal processing part, e.g., denoising of Lidar signals, vibration signals from conditioning systems of heavy mechanical systems etc.

CRediT authorship contribution statement

Khuram Naveed: Conceptualization, Methodology, Experiments, Writing - original draft, Formal analysis, Programming, Software, Investigation, Visualization. **Muhammad Tahir Akhtar:** Conceptualization, Writing - reviewing and editing, Supervision, Visualization. **Muhammad Faisal Siddiqui:** Methodology, Writing - reviewing and editing, Formal analysis, Investigation, Visualization. **Naveed ur Rehman:** Conceptualization, Methodology, Experiments, Supervision, Writing - reviewing and editing, Formal analysis, Investigation, Visualization, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Authors would like to acknowledge anonymous reviewers for providing many critical and insightful comments on the manuscript. This has greatly helped in improving the contents as well as quality of presentation. The work of Dr. Akhtar has been partially funded from the Faculty Development Competitive Research Grants Program of Nazarbayer University under the Grant Number 110119FD4525.

References

- [1] S. Mallat, A Wavelet Tour of Signal Processing, Elsevier, 1999.
- [2] D.L. Donoho, I.M. Johnstone, G. Kerkycharian, D. Picard, Wavelet shrinkage: asymptopia?, *J. R. Stat. Soc. B* (1995) 301–369.
- [3] F. Abramovich, T. Sapatinas, B.W. Silverman, Wavelet thresholding via a Bayesian approach, *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 60 (4) (1998) 725–749.
- [4] F. Abramovich, Y. Benjamini, Adaptive thresholding of wavelet coefficients, *Comput. Stat. Data Anal.* 22 (4) (1996) 351–361.
- [5] N. ur Rehman, S.Z. Abbas, A. Asif, A. Javed, K. Naveed, D.P. Mandic, Translation invariant multi-scale signal denoising based on goodness-of-fit tests, *Signal Process.* 131 (2017) 220–234.
- [6] K. Naveed, B. Shaukat, N. ur Rehman, Dual tree complex wavelet transform-based signal denoising method exploiting neighbourhood dependencies and goodness-of-fit test, *R. Soc. Open Sci.* 5 (9) (2018) 180436.
- [7] K. Naveed, B. Shaukat, N. ur Rehman, Signal denoising based on dual tree complex wavelet transform and goodness of fit test, in: 2017 22nd International Conference on Digital Signal Processing (DSP), IEEE, 2017, pp. 1–5.
- [8] I. Lavrik, Y. Young Jung, F. Ruggeri, B. Vidakovic, Bayesian false discovery rate wavelet shrinkage: theory and applications, *Commun. Stat., Simul. Comput.* 37 (6) (2008) 1086–1100.
- [9] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond. A, Math. Phys. Eng. Sci.* 454 (1971) (1998) 903–995.
- [10] Y. Kopsinis, S. McLaughlin, Development of emd-based denoising methods inspired by wavelet thresholding, *IEEE Trans. Signal Process.* 57 (4) (2009) 1351–1362.
- [11] A. Mert, A. Akan, Detrended fluctuation thresholding for empirical mode decomposition based denoising, *Digit. Signal Process.* 32 (2014) 48–56.
- [12] K. Dragomiretskiy, D. Zosso, Variational mode decomposition, *IEEE Trans. Signal Process.* 62 (3) (2014) 531–544.
- [13] M. Ali, A. Khan, N.U. Rehman, Hybrid multiscale wind speed forecasting based on variational mode decomposition, *Int. Trans. Electr. Energy Syst.* 28 (1) (2018) e2466.
- [14] G. Ren, J. Jia, X. Jia, J. Han, Variational Mode Decomposition Denoising Combined with the Euclidean Distance for Diesel Engine Vibration Signal, 2017.
- [15] H. Li, J. Chang, F. Xu, Z. Liu, Z. Yang, L. Zhang, S. Zhang, R. Mao, X. Dou, B. Liu, Efficient lidar signal denoising algorithm using variational mode decomposition combined with a whale optimization algorithm, *Remote Sens.* 11 (2) (2019) 126.
- [16] W. Ma, S. Yin, C. Jiang, Y. Zhang, Variational mode decomposition denoising combined with the Hausdorff distance, *Rev. Sci. Instrum.* 88 (3) (2017) 035109.
- [17] Y. Liu, G. Yang, M. Li, H. Yin, Variational mode decomposition denoising combined the detrended fluctuation analysis, *Signal Process.* 125 (2016) 349–364.
- [18] H. Hao, H. Wang, N. Rehman, A joint framework for multivariate signal denoising using multivariate empirical mode decomposition, *Signal Process.* 135 (2017) 263–273.
- [19] N. ur Rehman, B. Khan, K. Naveed, Data-driven multivariate signal denoising using Mahalanobis distance, *IEEE Signal Process. Lett.* 26 (9) (2019) 1408–1412.
- [20] H. Cramér, On the composition of elementary errors: first paper: mathematical deductions, *Scand. Actuar. J.* 1928 (1) (1928) 13–74.
- [21] M.A. Stephens, Edf statistics for goodness of fit and some comparisons, *J. Am. Stat. Assoc.* 69 (347) (1974) 730–737.
- [22] R. D'Agostino, Goodness-of-Fit-Techniques, Routledge, 2017.
- [23] T.W. Anderson, D.A. Darling, A test of goodness of fit, *J. Am. Stat. Assoc.* 49 (268) (1954) 765–769.
- [24] S. Lei, H. Wang, L. Shen, Spectrum sensing based on goodness of fit tests, in: Electronics, Communications and Control (ICECC), 2011 International Conference on, IEEE, 2011, pp. 485–489.
- [25] K. Naveed, S. Ehsan, K.D. McDonald-Maier, N. Ur Rehman, A multiscale denoising framework using detection theory with application to images from cmos/ccd sensors, *Sensors* 19 (1) (2019) 206.
- [26] M.P. McAssey, An empirical goodness-of-fit test for multivariate distributions, *J. Appl. Stat.* 40 (5) (2013) 1120–1131.
- [27] N. ur Rehman, K. Naveed, S. Ehsan, K. McDonald-Maier, Multi-scale image denoising based on goodness of fit (gof) tests, in: 2016 24th European Signal Processing Conference (EUSIPCO), IEEE, 2016, pp. 1548–1552.
- [28] K. Naveed, B. Shaukat, S. Ehsan, K.D. McDonald-Maier, N. ur Rehman, Multiscale image denoising using goodness-of-fit test based on edf statistics, *PLoS ONE* 14 (5) (2019) e0216197.
- [29] M.K. Steven, Fundamentals of Statistical Signal Processing: Detection Theory, 1998.
- [30] G. Chen, W.-P. Zhu, Signal denoising using neighbouring dual-tree complex wavelet coefficients, *IET Signal Process.* 6 (2) (2012) 143–147.
- [31] F. Abramovich, Y. Benjamini, Thresholding of wavelet coefficients as multiple hypotheses testing procedure, in: Wavelets and Statistics, Springer, 1995, pp. 5–14.
- [32] B.H. Tracey, E.L. Miller, Nonlocal means denoising of ecg signals, *IEEE Trans. Biomed. Eng.* 59 (9) (2012) 2383–2386.



Khuram Naveed received his PhD degree in 2020 from COMSATS University Islamabad (CUI), Islamabad, Pakistan. He is currently working as a lecturer in the Department of Electrical and Computer Engineering at CUI Islamabad where he has also worked as research associate from May 2011 to Jan 2014. He was a visiting researcher at Tandon School of Engineering, New York University (NYU), New York USA from Jan 2018 – July 2018. His research interests include signal and image processing using multiscale and multi-component signal decomposition methods. His recent research focuses on the design of multivariate signal processing techniques and machine learning applications.



Muhammad Tahir Akhtar received his PhD degree in Electronic Engineering from the Tohoku University, Sendai, Japan, in 2004, MSc in Systems Engineering from the Quaid-i-Azam University, Islamabad, Pakistan in 1999, and BSc Electrical (Electronics and Communication) Engineering from the University of Engineering & Technology, Taxila, Pakistan in 1997.

He is currently working as an Associate Professor at the School of Engineering and Digital Sciences, Nazarbayer University, Nur-Sultan City, Kazakhstan. From 2014 to 2017, he was an Associate Professor at the COMSATS University Islamabad, Pakistan. From 2008 to 2014, he was an Assistant Professor at the University of Electro-Communications, Tokyo, Japan, and a Special Visiting Researcher at the Tokyo Institute of Technology, Tokyo, Japan. He was a visiting researcher at Institute of Sound and Vibration Research (ISVR), University of Southampton, UK (Dec. 2008–Feb. 2009), and at Institute for Neural Computations (INC), University of California San Diego (Nov. 2010–Mar. 2011). Prior to that, he was a COE postdoctoral fellow at Tohoku University, Sendai Japan (2004–2005), and has worked as an Assistant Professor at the United Arab Emirates University, UAE (2006–2008).

His research interests include adaptive signal processing, active noise control, blind source separation, and biomedical signal processing. Dr. Akhtar has published about 95 papers in the peer-reviewed international journals and conference proceedings. He won Best Student Paper at the

IEEE 2004 Midwest Symposium on Circuits and Systems, Hiroshima, Japan, and student paper award (with Marko Kanadi) at 2010 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing. He is on the editorial board of *Advances in Mechanical Engineering*, and has served (2011-2013) as a co-editor for the newsletter of Asia-Pacific Signal & Information Processing Association (APSIPA). Currently he is member IEEE Signal Processing Society, IEEE Industrial Electronic Society, and he is a Senior Member IEEE.



Muhammad Faisal Siddiqui has been working as an Assistant Professor at Department of Electrical & Computer Engineering, COMSATS University Islamabad, Pakistan since July 2016. He is a Computer Engineer with expertise in Digital System Design and Medical Imaging. He received his PhD from Electrical Engineering Department, Faculty of Engineering,

University of Malaya, Kuala Lumpur, Malaysia. He received MS (EE) and BS (CE) degrees from COMSATS, Institute of Information Technology, Islamabad, Pakistan. His research interests are FPGA based digital system designing, biomedical image processing, embedded systems, application specific architectural design and real-time systems.



Naveed ur Rehman obtained his PhD from Imperial College London in 2011 and is currently working as Assistant Professor in the Department of Engineering at Aarhus University, Denmark. His research interests mainly include data-driven multi-scale signal decomposition methods, time-frequency analysis, optimization based methods in signal processing and applications of machine learning.