# ECG Baseline Estimation and Denoising With Group Sparse Regularization

**HAO SHI, RUIXIA LIU, CHANGFANG CHEN, MINGLEI SHU, AND YINGLONG WANG**
Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

Corresponding author: Yinglong Wang (wangylscsc@126.com)

**ABSTRACT** Baseline wander (BW) and electrocardiogram (ECG) noise reduction play an important role in ECG data analysis and disease diagnosis. This article introduces a sparse optimization method, which takes into account the group sparse characteristics of the signal, and combines low-pass filter to denoise the ECG signal and estimate the baseline. Derived from the classic total variation (TV) denoising method, a denoising method considering the structural characteristics of ECG signals is proposed. This method uses a band matrix to represent the sparse optimization problem, and adopts majorization-minimization (MM) algorithm to optimize the solution of the convergence problem. Through data comparison and detailed analysis, we first compares the method with two TV denoising methods. Then, the proposed method is validated in the MIT-BIH arrhythmia database of ECG signals, and compared with nonlocal means (NLM) and complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN) methods. The simulation experiment results show that the proposed algorithm has lower root mean square error (RMSE) and higher signal-to-noise ratio improvement (SNR_imp).

**INDEX TERMS** ECG denoising, baseline estimation, sparse optimization, group sparsity penalty.

## I. INTRODUCTION

Electrocardiogram (ECG) is an important tool for cardiovascular disease diagnosis, while during signal acquisition process, it is inevitably subject to various interferences, including random noises and baseline wander (BW). BW is usually caused by human breathing and movement. This is an inevitable common problem in the process of collecting ECG information. Its noise frequency is very low, usually less than 0.7Hz, which belongs to low-frequency information interference and often overlaps with the ST-segment component of ECG.

Research on the denoising of ECG signals and the correction of BW has always been a hot topic. For BW correction, the use of traditional nonlinear filters can easily cause waveform distortion. For this problem, empirical mode decomposition (EMD) was proposed to solve the problem by virtue of its good time scale characteristics [1]. In [2], the combination of EMD method and low-pass filter (LPF) was proposed, and in [3], a denoising method combining EMD and morphological filtering is proposed. More BW denoising works are presented in [4]–[7]. In view of the complex decomposition of

The associate editor coordinating the review of this manuscript and approving it for publication was Naveed ur Rehman.

EMD problems, a correction method combined with Hilbert transform is proposed in [6], but the details of the signal waveform are still incomplete. In [7], the method of EMD combined with mean-median filter has a good effect on the estimation of BW.

On the other hand, as one of the most effective methods, ECG denoising adaptive filtering has been recently studied in [8]–[11]. Due to the P-waves and QRS-waves in the ECG signals are similar to some wavelet bases, therefore, the wavelet transform is used to locate the peaks, and then a method of threshold noise reduction is proposed [12]–[15]. However, these methods will reduce the QRS peak of the ECG signals, After denoising, there will be underestimation, leading to the loss of ECG signal characteristics and useful information.

Sparse processing, as a new direction of signal processing, is now widely used in [16]–[18], especially the total variation (TV) denoising. TV is a commonly used penalty function sparse signal processing, which is widely used for denoising [19]–[22], reconstruction [23], nonlinear decomposition [24], [25], compression induction [26], deconvolution [27], [28]. However, TV noise reduction also has some disadvantages. After the signal is denoised, it usually produces step-like artifacts and underestimates the peak

value. Among the sparse methods, Ref. [29] proposed a sparse representation-based noise reduction and ECG signal BW correction algorithm. This method uses the input ECG records to train a redundant dictionary, and uses a well-trained dictionary to sparsely represent the ECG records to achieve signal noise reduction. The statistical method divides the atoms in the dictionary into clean signals and noises. However, this method has high algorithm complexity and takes a long time. In [30], the TV method using sparse derivative was designed to reduce the noise of the chromatogram, and the baseline was designed as a low-pass signal. In [31], an improved sparse assisted signal smoothing (SASS) formula was proposed to remove unsmooth points in the signal. In [32], [33], the sparse penalty function was further improved based on [30] to solve the problem of peak underestimation, but it cannot effectively eliminate BW interference. In [34], it was proposed to use an asymmetric penalty function to improve the sparsity of ECG signals, and considered the information of the positive and negative peaks of the signal to achieve precise punishment. In addition, signals with different differential orders have different denoising results. Normally, the more obvious the sparsity of high-order differential information, the better the denoising effect. However, since high-order difference information will lose more information components, it is still necessary to explore the choice of order. The author verifies that the sparsity characteristics of the third-order difference combined with the asymmetric penalty of TV noise reduction results have achieved good results through experimental results of different orders. It is worth mentioning that although certain achievements have been witnessed in TV problem combined with asymmetric penalty functions, the study of TV problems with different penalty functions still has great significance.

In this paper, we consider the group sparsity characteristics of the signal or signal derivative, and proposes a new TV denoising algorithm (GSTV). Specifically, we assume that the larger value derivatives of the signal are not independently distributed, and usually have a distribution within a certain range. From this perspective, the signal point value changes rapidly and has a cluster attribute or a group attribute. The proposed new TV noise reduction method is similar to the traditional TV method. Both of them are based on minimizing a convex cost function, and use multiplier optimization minimization (MM) to solve the optimization problem of construction [35]. In addition, a sparse-based 'fusion lasso signal approximator' is proposed in [36]. Different from TV and GSTV, it is a compound penalty called CTV. This method not only considers the sparse characteristics of the signal, but also includes the sparse characteristics of the first-order difference of the signal. This composite function can be regarded as a special example of articles [37], [38], where two or more regularization functions are used to promote sparse signal expression.

Considering that in the existing technology, BW correction and noise removal are addressed separately, researchers have designed a method that integrates these two parts to improve efficiency. In [5], some researchers applied wavelet transform to locate the wave, and then set a threshold function to remove noise, but it could not effectively remove changing BW interference. In [29], using dictionary training to sparsely represent the signal, this method takes a long time and reduces the peak of QRS-waves.

In order to further reduce the influence of BW interference and the peak underestimation, a denoising method based on LPF filter and group sparsity is proposed. This article lists its main contributions.

1) The characteristics of ECG signals are analyzed from the perspective of sparsity. Using the group sparsity penalty function cannot only improve the sparsity of ECG signals, but also improve the convexity of the objective optimization function.

2) The performance of group-based sparse denoising algorithm has been thoroughly tested, which is a derivative of the TV algorithm. We conducted a series of comparative experiments to explore the denoising performance algorithm based on group sparsity threshold. The relationship between parameter settings and performance is verified through experiments. The experimental results also show that the denoising algorithm based on the group sparsity threshold has a better denoising performance than the TV and CTV algorithms.

3) Use LPF filter and group sparse algorithm to correct BW and denoise ECG signals, effectively improve the ECG signal ST-segment elevation and peak underestimation problems. In addition, in comparison with the method in [34], we verify in detail the denoising performance under different differential orders and analyze the experimental data. Combining the group sparse TV denoising method, it has a certain degree of scalability in the sparse field denoising.

The structure of the paper is as follows. The relevant pre-knowledges is in Section II. In Section III, several optimization algorithms are compared through analysis and derivation. The simulation results are discussed in detail in Section IV. Section V draws the conclusion.

## II. PRE-KNOWLEDGES
### A. NOTATION

In this paper, lowercase and uppercase boldface for vectors and matrices are adopted respectively. The $N$-point signal $\mathbf{x}$ with the length of $N$ is represented by the vector $\mathbf{x} = [x(0), x(1) \ldots, x(N-1)]^T \in \mathbb{R}^N$, where $[\cdot]^T$ is the transpose. The signal is expressed in the discrete domain, so the first-order difference characteristic of the signal can be expressed as

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \ddots & \ddots \\ & & & -1 & 1 \end{bmatrix}. \tag{1}$$

The first difference of signal $\mathbf{x}$ can be represented by $\mathbf{Dx}$, where the magnitude of $\mathbf{D}$ is $(N-1) \times N$.

A $K$-point group of the vector $v$ can be expressed as

$$\mathbf{v}_{n,K} = [v(n), \ldots, v(n+K-1)] \in \mathbb{R}^N. \qquad (2)$$

This is the $K$ points of the vector $v$ starting from index $n$. Except these, The $\ell_1$ norm is

$$\|x\|_1 = \sum_n |x_n|. \qquad (3)$$

The soft-threshold function in [39] is defined as

$$\text{soft}(x, T) = \begin{cases} 0, & |x| < T \\ x - T(x/|x|), & |x| > T \end{cases} \qquad (4)$$

for $x \in \mathbb{C}$ and $T > 0$. This is the soft threshold function expression of a continuous signal. In the discrete domain for a signal $x(n)$ or vector $\mathbf{x}$, the $\text{soft}(x, T)$ is an application to every signal point.

## B. SPARSITY CHARACTERISTIC OF ECG SIGNALS

The sparsity of a signal means that most of its components are close to zero. As shown in Fig. 1, a complete ECG signal consists of T, Q, R, S and P waves. In the P-R and S-T intervals, the characteristics of the signal are relatively flat and sparse. Considering the sparsity characteristics, comparing the results after the difference in Fig. 2, it can be seen that the sparsity of the first-order difference is higher, and that of the third-order difference is basically zero.
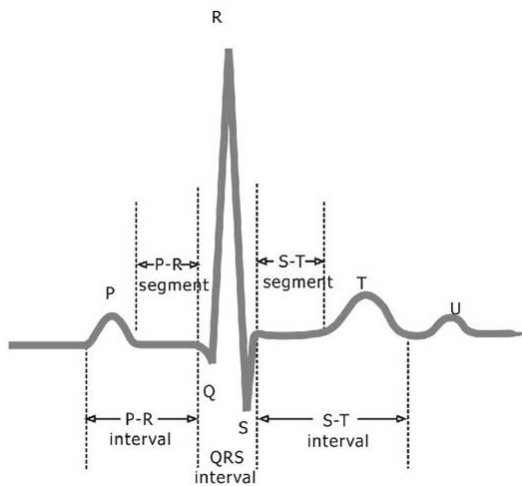


**FIGURE 1.** A complete ECG signal waveform.

## C. GROUP-SPARSE TOTAL VARIATION DENOISING

Sparse derivative denoising refers to the problem of estimating the sparse characteristic or the derivative approximate sparse signal $\mathbf{x}$ based on the observation of noise. The model is

$$\mathbf{y} = \mathbf{x} + \mathbf{w}, \qquad (5)$$

where $\mathbf{w}$ stands for noise component. In [40], the $\ell_1$ norm is a convex proxy for sparsity. For discrete data, the first-order
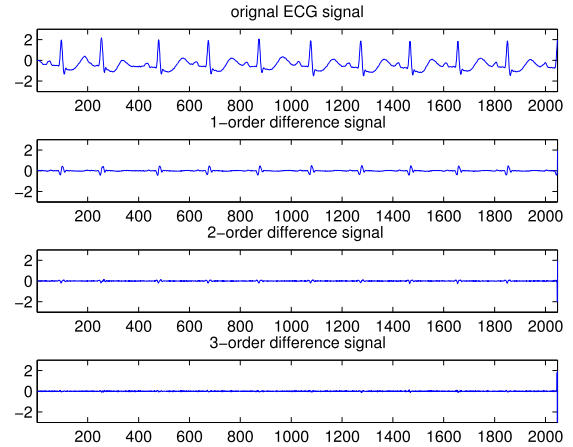


**FIGURE 2.** Comparisons of ECG signal differential results.

difference is the simplest approximation. To estimate component $\mathbf{x}$ it can consider the minimization of $\|\mathbf{Dx}\|_1$. Assuming $\mathbf{g}$ is Gaussian white noise with variance $\sigma^2$, a suitable data fidelity constraint for $\mathbf{x}$ is $\|\mathbf{y} - \mathbf{x}\|_2^2 \leq N\sigma^2$. That is to solve the constrained optimization problem

$$\arg \min_{\mathbf{x}} \ \|\mathbf{Dx}\|_1, \qquad (6a)$$

$$\text{such that } \|\mathbf{y} - \mathbf{x}\|_2^2 \leq N\sigma^2. \qquad (6b)$$

For a suitable regularization coefficient $\lambda$, change to a unconstrained optimization problem.

$$\arg \min_x \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{Dx}\|_1 \right\}, \qquad (7)$$

we describe the solution to problem (7) as TV $(\mathbf{y}, \lambda)$,

$$\text{TV}(\mathbf{y}, \lambda) := \arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{Dx}\|_1 \right\}. \qquad (8)$$

In addition, here, as described in the introduction, it is assumed that the derivative (first-order difference, $\mathbf{Dx}$) has group sparsity behavior. The x estimation in Eq. (5), we can introduce a function $\phi(\mathbf{x})$ to promote group sparsity and replace the constraint problem in Eq. (7)

$$\arg \min_x \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \phi(\mathbf{Dx}) \right\}. \qquad (9)$$

Joint Eq. (2), mark $\mathbf{v} = \mathbf{Dx}$, so $\phi(\mathbf{v})$ has the following expression

$$\phi(\mathbf{v}) = \sum_n \left[ \sum_{k=0}^{K-1} |v(n+k)|^2 \right]^{1/2}. \qquad (10)$$

The $\phi(\mathbf{v})$ regularizer is used to promote group sparsity [40]–[43], the parameter $K$ is the size of group. Note that, if $K = 1$, $\phi(\mathbf{v}) = \|\mathbf{v}\|_1$ and problem Eq. (9) is TV denoising problem. If $K > 1$, the function $\phi(\mathbf{v})$ is a convex measure of group sparsity. We refer to problem Eq. (9) as group-sparse total variation (GSTV) denoising.

$$\text{GSTV}(\mathbf{y}, \lambda, K) := \arg \min_x \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda \phi \|\mathbf{Dx}\|_1 \right\}. \qquad (11)$$

### D. FUSED LASSO SIGNAL APPROXIMATOR

TV and GSTV, consider the sparse characteristics of signal derivatives and the structural characteristics of the signal respectively. Next we consider a more complex situation. If both the signal $x$ and its derivative are all sparse, then the denoising problem is more appropriately formulated through Eq.(7) as

$$\arg \min_{\mathbf{x}} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda_0 \|\mathbf{x}\|_1 + \lambda_1 \|\mathbf{Dx}\|_1 \right\}. \quad (12)$$

This is an example that considers a compound penalty function, with two regulators regularizers used to promote signal recovery [37], [38], [44], [45]. The problem Eq. (12) is called the 'fused lasso signal approximator'. Here, it is described as compound penalty total variation (CTV) denoising.

$$\text{CTV}(\mathbf{y}, \lambda_0, \lambda_1) := \arg \min_{x} \left\{ \frac{1}{2} \|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda_0 \|\mathbf{x}\|_1 + \lambda_1 \|\mathbf{Dx}\|_1 \right\}. \quad (13)$$

Note that, in Proposition 1 in [36] shows that problem Eq. (12) is equivalent to Eq. (7) in the sense that the solution to Eq. (12) can be obtained explicitly from the solution to Eq. (7). So the solution of Eq. (12) can be

$$\mathbf{x} = \text{soft}(\text{TV}(\mathbf{y}, \lambda_1), \lambda_0). \quad (14)$$

It is not necessary to use a new algorithm for solving problem Eq. (12) here, it is enough to have a method to solve TV denoising.

### E. MAJORIZATION-MINIZATION

In solving convex optimization problems, the MM algorithm replaces difficult minimization problems with simple problems [35]. To minimize the $F(x)$, the MM approach defines an iterative algorithm via:

$$\mathbf{x}_{(i+1)} = \arg \min_{\mathbf{x}} G\left(\mathbf{x}, \mathbf{x}_{(i)}\right), \quad (15)$$

where $i$ is the iteration index, $i > 0$, and the function $G_i(\mathbf{x})$ is any convex majorizer of $F(\mathbf{x})$ (i.e., $G_i(\mathbf{x}) > F(\mathbf{x}) \forall \mathbf{x}$), and when $F(\mathbf{x})$ at $\mathbf{x}_i$ (i.e., $G_i(\mathbf{x}_i) = F(\mathbf{x}_i)$). Initialize $x_0$, and then iterate Eq. (15) until it converges to the minimum value of $F(\mathbf{x})$. Learn more details in [35] and references therein.

### F. HIGHER-ORDER HIGH-PASS FILTER

The use of linear filters, applied to the processing of one-dimensional signals, must require the design of the filter structure to meet zero-phase non-causality, Because it can avoid some unnecessary distortion. It can be achieved by calculating some band matrices. A discrete-time filter is described by the difference equation

$$\sum_i a\,(i)\,y\,(n-i) = \sum_i b\,(i)\,x\,(n-i), \quad (16)$$

where $y(n)$ and $x(n)$ are the output and input signals respectively.

Signal processing problems based on sparsity are usually expressed by finite-length signals, especially in the case of TV noise reduction. The difference Eq. (15) for finite-length signals, we write

$$\mathbf{Ay} = \mathbf{Bx}, \quad (17)$$

where $\mathbf{A}$ and $\mathbf{B}$ are banded matrices. The output is $y$ equals to

$$\mathbf{y} = \mathbf{A}^{-1}\mathbf{Bx}. \quad (18)$$

In this paper, we set $\mathbf{H} = \mathbf{A}^{-1}\mathbf{B}$, $\mathbf{H}$ represents the high-pass filter. Regarding the setting of the filter, we mentioned that the filter $\mathbf{H}$ and $\mathbf{LPF}$ need to meet the zero-phase characteristic, so as to avoid unnecessary losses in the signal processing process. The matrices $\mathbf{A}$ and $\mathbf{B}$ are required to have specific properties. This kind of filter is proposed in [40], the parameter is by order $o$ and cutoff frequency $fc$.

Let $P$ be reversal matrix, the filter should satisfy

$$\mathbf{PHP} = \mathbf{P}, \quad (19)$$

where the dimension of $\mathbf{P}$ is determined by the dimensions of $\mathbf{H}$. If $\mathbf{A}$ and $\mathbf{B}$ satisfy

$$\mathbf{PAP} = \mathbf{A} \quad \text{and} \quad \mathbf{PBP} = \mathbf{B}. \quad (20)$$

then $\mathbf{H} = \mathbf{A}^{-1}\mathbf{B}$ satisfies Eq. (19).

Examples of recursive zero-phase high-pass filters can be obtained in Ref. [40] Section II.

### G. SYSTEM MODEL

A complicated ECG signals includes the original clean ECG signals, BW signals and noises.

$$\mathbf{y} = \mathbf{x} + \mathbf{f} + \mathbf{w}, \quad (21)$$

where $\mathbf{x}$ is the original clean ECG signals, $\mathbf{f}$ represents the BW signals, and $\mathbf{w}$ is the Gaussian white noises with variance is $\sigma^2$.

Our goal is to estimate the BW signals $\mathbf{f}$, and clean ECG signals $\mathbf{x}$. Observed in $\mathbf{y}$ seek estimates

$$\hat{\mathbf{x}} \approx \mathbf{x}, \quad \hat{\mathbf{f}} \approx \mathbf{f}. \quad (22)$$

Given an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$, we will estimate $\mathbf{f}$ as

$$\hat{\mathbf{f}} := \mathbf{LPF}\left(\mathbf{y} - \hat{\mathbf{x}}\right), \quad (23)$$

where $\mathbf{LPF}$ is low-pass filter. Therefore, we seek $\hat{\mathbf{x}}$, Using Eq. (23) in Eq. (22), we can know

$$\mathbf{LPF}\left(\mathbf{y} - \hat{\mathbf{x}}\right) \approx \mathbf{f}. \quad (24)$$

Using Eq. (21) in Eq. (24)

$$\mathbf{LPF}\left(\mathbf{y} - \hat{\mathbf{x}}\right) \approx \mathbf{y} - \mathbf{x} - \mathbf{w}. \quad (25)$$

Using Eq. (22) in Eq. (25)

$$\mathbf{LPF}\left(\mathbf{y} - \hat{\mathbf{x}}\right) \approx \mathbf{y} - \hat{\mathbf{x}} - \mathbf{w}, \quad (26)$$

or

$$\left(\mathbf{y} - \hat{\mathbf{x}}\right) - \mathbf{LPF}\left(\mathbf{y} - \hat{\mathbf{x}}\right) \approx \mathbf{w}. \quad (27)$$

The left side of Eq. (27) is a high-pass filter $\mathbf{y} - \hat{\mathbf{x}}$. $\mathbf{H} := \mathbf{I} - \mathbf{LPF}$, so Eq. (27) equals to

$$\mathbf{H}\left(\mathbf{y} - \hat{\mathbf{x}}\right) \approx \mathbf{w}. \tag{28}$$

The Eq. (27) contains the data $\mathbf{y}$, the estimate $\hat{\mathbf{x}}$ that we seek to determine. The noise component $\mathbf{w}$ can be used to derive an estimate $\hat{\mathbf{x}}$.

To solve the problem Eq. (21), generally, convex optimization techniques are used to estimate transient components from observed signals. According to the Eq. (7), Eq. (9) and Eq. (12), we can solve this problem. After obtaining the estimate of $\hat{\mathbf{x}}$, using Eq.(24), $\mathbf{f}$ can be obtained.

## III. OPTIMIZATION ALGORITHM

Regarding the problem of the form like Eq. (8), Eq. (11), Eq. (13), non-differentiable phenomena often appear in image processing/signal processing tasks (compressed sensing, deconvolution, reconstruction, ect.). In this papaer, we use the MM as the solution algorithm [35]. We change variable

$$\mathbf{x} = \mathbf{R}\mathbf{u}, \tag{29}$$

where $\mathbf{R}$ is a matrix of size $N \times (N-1)$).

$$\mathbf{R} := \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ 1 & 1 & 0 & & \\ \vdots & & & \ddots & \\ 1 & 1 & \cdots & 1 & 0 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}. \tag{30}$$

This is a cumulative sum.

$$\mathbf{D}\mathbf{R} = \mathbf{I}, \tag{31}$$

$\mathbf{R}$ is a discrete anti-derivative. So

$$\mathbf{D}\mathbf{x} = \mathbf{D}\mathbf{R}\mathbf{u} = \mathbf{u}. \tag{32}$$

Introduce the filter introduced in Section II. D, the matrix $\mathbf{B}$ equals to

$$\mathbf{B} = \mathbf{B}_1\mathbf{D}. \tag{33}$$

where $\mathbf{B}_1$ is a banded matrix. Using Eqs. (18), (31), (32), we can get

$$\mathbf{H}\mathbf{R} = \mathbf{A}^{-1}\mathbf{B}\mathbf{R} = \mathbf{A}^{-1}\mathbf{B}_1\mathbf{D}\mathbf{R} = \mathbf{A}^{-1}\mathbf{B}_1. \tag{34}$$

### A. GROUP SPARSE PENALTY MODEL

We use MM algorithm to minimize objective function.

Using Eq. (30) in Eq. (8), Eq. (11) gives respectively

$$\mathbf{x}^* = \arg\min_{\mathbf{u}} \left\{ F(\mathbf{u}) = \frac{1}{2}\|\mathbf{H}(\mathbf{y} - \mathbf{R}\mathbf{u})\|_2^2 + \lambda\|\mathbf{u}\|_1 \right\}. \tag{35}$$

$$\mathbf{x}^* = \arg\min_{\mathbf{u}} \left\{ F(\mathbf{u}) = \frac{1}{2}\|\mathbf{H}(\mathbf{y} - \mathbf{R}\mathbf{u})\|_2^2 + \lambda\phi(\mathbf{u}) \right\}. \tag{36}$$

Reference Eq. (2), the penalty function $\phi(\mathbf{u})$ equals to

$$\phi(\mathbf{u}) = \sum_n \|\mathbf{u}_{n,k}\|_2. \tag{37}$$

To find a majorizor of $F(\mathbf{x})$ in Eq. (36), firstly find a majorizor of $\phi(\mathbf{u})$. Note that

$$\frac{1}{2\|\mathbf{z}\|_2}\|\mathbf{u}\|_2^2 + \frac{1}{2}\|\mathbf{z}\|_2 \geqslant \|\mathbf{u}\|_2, \tag{38}$$

for all $\mathbf{u}$ and $\mathbf{z} \neq \mathbf{0}$ with equality when $\mathbf{u} = \mathbf{z}$. Using Eq. (38) for each group, a majorizor of $\phi(\mathbf{u})$ is given by

$$g(\mathbf{u}, \mathbf{z}) = \frac{1}{2}\sum_n\left[\frac{1}{\|\mathbf{z}_{n,K}\|_2}\|\mathbf{u}_{n,K}\|_2^2 + \|\mathbf{z}_{n,K}\|_2\right], \tag{39}$$

with

$$g(\mathbf{u}, \mathbf{z}) \geqslant \phi(\mathbf{u}), \qquad g(\mathbf{z}, \mathbf{z}) = \phi(\mathbf{z}), \tag{40}$$

provided $\|\mathbf{z}_{n,K}\|_2 \neq 0$ for all $n$. Note that $g(\mathbf{u}, \mathbf{z})$ is quadratic in $\mathbf{u}$. It can be written as

$$g(\mathbf{u}, \mathbf{z}) = \frac{1}{2}\mathbf{u}^T\Lambda(\mathbf{z})\mathbf{u} + C, \tag{41}$$

where $C$ is not related with $\mathbf{v}$. $\Lambda(\mathbf{z})$ is a diagonal matrix

$$[\Lambda(\mathbf{z})] = \sum_{j=0}^{K-1}\left[\sum_{k=0}^{K-1}|z(n-j+k)|^2\right]^{-1/2}. \tag{42}$$

At the $i$-th iteration, with Eq. (41), a majorizor of $F(\mathbf{u})$ is given by

$$\begin{aligned}
G_i(\mathbf{u}) &= \frac{1}{2}\|\mathbf{H}(\mathbf{y} - \mathbf{R}\mathbf{u})\|_2^2 + \lambda\, g(\mathbf{u}, \mathbf{z}) \\
&= \frac{1}{2}\|\mathbf{H}(\mathbf{y} - \mathbf{R}\mathbf{u})\|_2^2 + \frac{\lambda}{2}\mathbf{u}^T\Lambda_i(\mathbf{z})\mathbf{u} + \lambda C \\
&= \frac{1}{2}\left\|\mathbf{A}^{-1}\mathbf{B}\mathbf{y} - \mathbf{A}^{-1}\mathbf{B}_1\mathbf{u}\right\|_2^2 + \frac{\lambda}{2}\mathbf{u}^T\Lambda_i(\mathbf{z})\mathbf{u} + \lambda C \\
&= \frac{1}{2}(\mathbf{A}^{-1}\mathbf{B}\mathbf{y} - \mathbf{A}^{-1}\mathbf{B}_1\mathbf{u})^T(\mathbf{A}^{-1}\mathbf{B}\mathbf{y} - \mathbf{A}^{-1}\mathbf{B}_1\mathbf{u}) \\
&\quad + \frac{\lambda}{2}\mathbf{u}^T\Lambda_i(\mathbf{z})\mathbf{u} + \lambda C,
\end{aligned} \tag{43}$$

i.e., $\quad G(\mathbf{u}, \mathbf{z}) \geq F(\mathbf{u}), \ G(\mathbf{z}, \mathbf{z}) = F(\mathbf{z})$.

To minimize $F(\mathbf{x})$, using Eq. (15) after iterative is

$$\mathbf{u}_{i+1} = (\mathbf{B}_1^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{B}_1 + \lambda\Lambda_i(\mathbf{z})^{-1})^{-1}\mathbf{B}^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{B}\mathbf{y}, \tag{44}$$

As the number of iterations increases, many values of $\mathbf{u}_i$ will be 0, and some entries of $\Lambda_i^{-1}$ will reach infinity. Refer to [46], solve the problem by matrix inverse lemma:

$$\begin{aligned}
&(\mathbf{B}_1^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{B}_1 + \lambda\Lambda(\mathbf{z})^{-1})^{-1} \\
&= \frac{1}{\lambda}\Lambda_i(\mathbf{z}) - \frac{1}{\lambda}\Lambda_i(\mathbf{z})\mathbf{B}_1^T(\lambda\mathbf{A}\mathbf{A}^T + \mathbf{B}_1\Lambda_i(\mathbf{z})\mathbf{B}_1^T)^{-1}\mathbf{B}_1\Lambda_i(\mathbf{z}),
\end{aligned} \tag{45}$$

The indicated matrix is banded because $\mathbf{A}$, $\mathbf{B}$ and $\Lambda_i$ are all banded. Using MM update

$$\mathbf{b} \leftarrow \frac{1}{\lambda}\mathbf{B}_1^T(\mathbf{A}\mathbf{A}^T)^{-1}\mathbf{B}\mathbf{y}$$

$$\Lambda_i \leftarrow \sum_{j=0}^{K-1}\left[\sum_{k=0}^{K-1}|z_i(n-j+k)|^2\right]^{-1/2}$$

$$\mathbf{u}_{i+1} \leftarrow \Lambda_i[\mathbf{b} - \mathbf{B}_1^T(\lambda\mathbf{A}\mathbf{A}^T + \mathbf{B}_1\Lambda_i\mathbf{B}_1^T)^{-1}\mathbf{B}_1\Lambda_i\mathbf{b}]. \tag{46}$$

And when $K = 1$, its TV Denoising, a special form of GSTV. The complete algorithm is proposed in Algorithm 1.

---

**Algorithm 1** LPF / GSTV

**Input**: $\mathbf{y}, K, \lambda > 0$
**Output**: $\mathbf{f}, \mathbf{x}$
1. $\mathbf{u} \leftarrow \mathbf{D}\mathbf{y}$
2. $\mathbf{b} \leftarrow (1/\lambda) \mathbf{B}_1^T \left(\mathbf{AA}^T\right)^{-1} \mathbf{B}\mathbf{y}$
3. **repeat**
4. $\quad \Lambda \leftarrow \sum_{j=0}^{K-1} \left[\sum_{k=0}^{K-1} |z(n-j+k)|^2\right]^{-1/2}$
5. $\quad \mathbf{M} \leftarrow \lambda \mathbf{AA}^T + \mathbf{B}_1 \Lambda \mathbf{B}_1^T$
6. $\quad \mathbf{u} \leftarrow \Lambda \left[\mathbf{b} - \mathbf{B}_1^T \mathbf{M}^{-1} \mathbf{B}_1 \Lambda \mathbf{b}\right]$
7. **until** convergence
8. $\mathbf{f} \leftarrow (\mathbf{y} - \mathbf{x}) - \mathbf{A}^{-1} \mathbf{B} (\mathbf{y} - \mathbf{x})$
9. $\mathbf{x} \leftarrow \mathbf{R}\mathbf{u}$
10. **return** $\mathbf{f}, \mathbf{x}$

---

### B. COMPOUND SPARSE PENALTY MODEL

In the optimization solution of Eq. (13), the regularization term constraint problem includes the first-order difference form of the signal. As in Section IV.A, we will estimate the BW signals $f$ by applying a low-pass filter to $(\mathbf{y} - \hat{\mathbf{x}})$. To estimate $\hat{\mathbf{x}}$, instead of solving Eq. (13), we solve

$$\arg\min_{\mathbf{x}} \left\{\frac{1}{2}\|\mathbf{H}(\mathbf{y}-\mathbf{x})\|_2^2 + \lambda_0\|\mathbf{x}\|_1 + \lambda_1\|\mathbf{D}\mathbf{x}\|_1\right\}. \quad (47)$$

To solve the optimization problem above, it needs to be pointed out. If the MM algorithm is used to solve Eq. (47), the final iterative equations need to solve $N$ equations, where $N$ is the signal length. Therefore, for CTV denoising solution, the alternating direction method of multipliers (ADMM) optimization method is adopted [35], [42], [47], it is also a method for solving dual problems [48], [49].

As in [35], we apply 'variable splitting' to decouple the terms of the cost function. so Eq. (47) can be rewritten as the constrained problem:

$$\arg\min_{\mathbf{x},\mathbf{p}} \left\{\frac{1}{2}\|\mathbf{H}(\mathbf{y}-\mathbf{x})\|_2^2 + \lambda_0\|\mathbf{p}\|_1 + \lambda_1\|\mathbf{D}\mathbf{p}\|_1\right\} \quad (48a)$$

$$\text{such} \qquad \mathbf{p} = \mathbf{x}. \quad (48b)$$

Applying ADMM to Eq. (47) yields the iterative algorithm:

$$\mathbf{x} \leftarrow \arg\min_{\mathbf{x}} \|\mathbf{H}(\mathbf{y}-\mathbf{x})\|_2^2 + \mu\|\mathbf{p}-\mathbf{x}-\mathbf{d}\|_2^2 \quad (49a)$$

$$\mathbf{p} \leftarrow \arg\min_{\mathbf{p}} \lambda_0\|\mathbf{p}\|_1 + \lambda_1\|\mathbf{D}\mathbf{x}\|_1 + 0.5\mu\|\mathbf{p}-\mathbf{x}-\mathbf{d}\|_2^2 \quad (49b)$$

$$\mathbf{d} \leftarrow \mathbf{d} - (\mathbf{p}-\mathbf{x}) \quad (49c)$$

$$\text{Go to Eq. (48a).} \quad (49d)$$

It should be noted that in the algorithm Eq. (49), the parameter $\mu$ needs to be a positive number. The value of $\mu$ will not affect the convergence of the algorithm, but it will affect the convergence speed. Parameters $\mathbf{d}$ and $\mathbf{p}$ also need to be determined before the loop, but because the cost function is a convex function, the algorithm will converge to the minimum, so we set $\mathbf{d}$ and $\mathbf{p}$ to be zero vectors of the same size as $\mathbf{y}$.

The solution to Eq. (49a) can be expressed as

$$\mathbf{x} \leftarrow \left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{I}\right)^{-1} \left(\mathbf{H}^T\mathbf{H}\mathbf{y} + \mu(\mathbf{p}-\mathbf{d})\right). \quad (50)$$

From Eq. (18), we write

$$\mathbf{H}^T\mathbf{H}\mathbf{y} = \mathbf{B}^T \left(\mathbf{AA}^T\right)^{-1} \mathbf{B}\mathbf{y}. \quad (51)$$

Introducing the matrix inverse lemma, we can get

$$\left(\mathbf{H}^T\mathbf{H} + \mu\mathbf{I}\right)^{-1} = \frac{1}{\mu}\left[\mathbf{I} - \mathbf{B}\left(\mu\mathbf{AA}^T + \mathbf{BB}^T\right)^{-1}\mathbf{B}\right]. \quad (52)$$

Using Eqs. (51) and Eq. (52) in Eq.(50), line Eq. (49a) is implemented as

$$\mathbf{g} \leftarrow \frac{1}{\mu}\mathbf{B}^T \left(\mathbf{AA}^T\right)^{-1} \mathbf{B}\mathbf{y} + (\mathbf{p}-\mathbf{d}), \quad (53a)$$

$$\mathbf{x} \leftarrow \mathbf{g} - \mathbf{B}^T \left(\mu\mathbf{AA}^T + \mathbf{BB}^T\right)^{-1} \mathbf{B}\mathbf{g}. \quad (53b)$$

Note that because $\mathbf{y}$ will not be updated in Eq. (49), the first term on the right side of Eq. (53b) only needs to be calculated once; so it can be precomputed prior to the iteration. Using Eq. (14), the problem Eq. (49b) equals to

$$\mathbf{u} \leftarrow \text{soft}(\text{TV}(\mathbf{x}+\mathbf{d}, \lambda_1/\mu), \lambda_0/\mu), \quad (54)$$

with the above support, the ADMM algorithm Eq. (49) can be implemented for LPF/CTV.

---

**Algorithm 2** LPF/CTV

**Input**: $\mathbf{y}, \lambda_0 > 0, \lambda_1 > 0, \mu > 0$
**Output**: $\mathbf{f}, \mathbf{x}$
1. $\mathbf{b} \leftarrow (1/\mu) \mathbf{B}^T \left(\mathbf{AA}^T\right) \mathbf{B}\mathbf{y}$
2. $\mathbf{d}, \mathbf{p} \leftarrow 0$
3. **repeat**
4. $\quad \mathbf{x} \leftarrow \mathbf{b} + \mathbf{p} - \mathbf{d} - \mathbf{B}^T \left(\mu\mathbf{AA}^T + \mathbf{BB}^T\right)^{-1} \mathbf{B}\mathbf{g}$
5. $\quad \mathbf{v} \leftarrow \text{soft}(\text{TV}(\mathbf{d}+\mathbf{x}, \lambda_1/\mu), \lambda_0/\mu)$
6. $\quad \mathbf{d} \leftarrow \mathbf{d} - \mathbf{p} + \mathbf{x}$
7. **until** convergence
8. $\mathbf{f} \leftarrow (\mathbf{y} - \mathbf{x}) - \mathbf{A}^{-1} \mathbf{B} (\mathbf{y} - \mathbf{x})$
9. **return** $\mathbf{f}, \mathbf{x}$

---

We have noticed that in the process of solving the LPF/CTV problem, there are more parameters for complex signal models. Such examples and more optimization algorithms are also mentioned in [50], [51]. The algorithm updates process of LPF/CTV is proposed in Algorithm 2.

## IV. SIMULATION RESULTS

### A. ECG DATABASE

The tested ECG data and BW signals come from the arrhythmia database and noise stress test database of MIT-BIH, respectively [41]. Set 20 dB, 15 dB, 10 dB, 5 dB, −5 dB and −10 dB signal-to-noise ratio (SNR) for the system respectively. First, we compare the proposed method with several TV noise reduction methods and prove that group sparseness has better performance in recovering signal details, and then

compared with the method in [34], the two methods are based on the same idea, but the recovery results are different.

### B. PERFORMANCE INDEX

We have listed two indicators for evaluating the results of the algorithm, the rooted mean square error (RMSE) and the signal-to-noise ratio improvement (SNR_imp). Among them, the RMSE calculation formula is as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(x_{obs,i} - x_{org,i}\right)^2}{n}}, \quad (55)$$

$x_{obs}$ is the signal after processing, $x_{org}$ is original input signal. As well known, the denoising effect performs better at lower RMSE and higher SNR_imp.

### C. BASELINE ESTIMATE OF REAL ECG SIGNAL

This section explains the removal of BW from real ECG signals. Fig. 3(a) shows the clean ECG signals (MIT-BIH.103), Fig. 3(b) adds the BW signals (BW comes from the noise pressure database), and Fig. 3(c) superimposes Gaussian white noises.
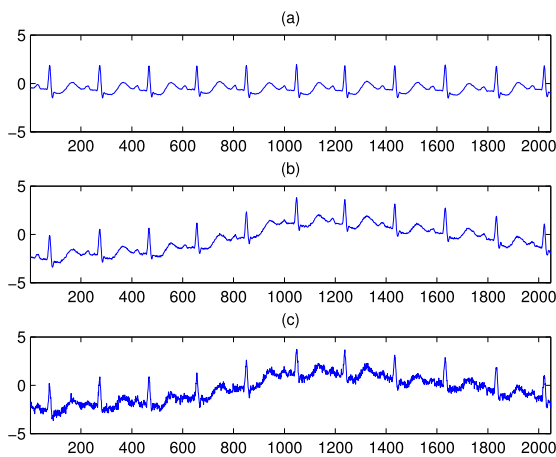


**FIGURE 3.** The example of ECG signal. (a) Original ECG signal. (b) Baseline wander with ECG signal. (c) Baseline wander and noise ECG signal.

Fig. 4 shows the results of the real ECG baseline estimation using the proposed method. In Fig. 4(a) the blue line is the real ECG signal containing baseline drift and noise, and the blue line in Fig. 4(b) is the estimation. The baseline level shown in gray is the real ECG data, and Fig. 4(c) is the noise ECG signal after removing the baseline. It is clear that after removing the influence of BW, the normal curve of signal characteristics is successfully obtained. It should be pointed out that, in the algorithm process, in order to prevent the low-pass filter from carrying too much useful information, we set the parameter $o = 1$ and $f_c = 0.007$.

### D. DENOISING OF REAL ECG SIGNAL

In order to show the denoising effect, the conclusion is divided into two parts. (1) To verify that the proposed GSTV method has a better recovery effect on the ECG signal than
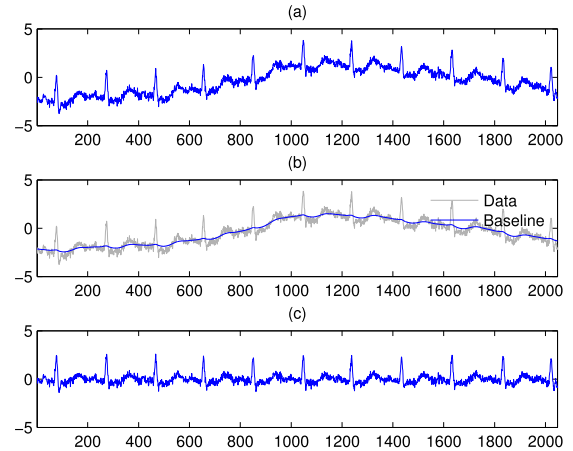


**FIGURE 4.** Baseline Wander correction. (a) The ECG signal added BW signal and noise. (b) Baseline estimation. (c) The ECG signal after BW correction.

TV, CTV. We set the noise with different SNRs to apply each algorithm. (2) In addition, the experimental method proposed by this paper is also compared with traditional algorithms.

Fig. 5 shows a schematic diagram of the denoising results of GSTV, TV, and CTV. Fig. 5(a) is the original ECG signal without noise (MIT-BIH no.103), Fig. 5(b) ECG signal with 5dB noise added, Fig. 5(c) TV denoising result Schematic diagram, Fig. 5(d) Schematic diagram of CTV denoising result and Fig. 5(e) Schematic diagram of GSTV denoising result. The red line is the denoising result, and the blue line is the original data.

In Fig. 5, it is difficult to find the difference among the denoising results of GSTV, TV and CTV without careful observation. Therefore, we made a more detailed comparison to show the difference in denoising effects. Take the 500 signal points between 260 and 760 for alignment comparison, and the result is shown in Fig. 6.

Fig. 6(a) is the detail component of noise, Fig. 6(b) is TV denoising, Fig. 6(c) is GSTV denoising, and Fig. 6(d) is CTV denoising. The black box in Fig. 6(b) shows that TV denoising will underestimate the peak information of the original ECG signal and generate stepped information at the corners of the signal waveform. Fig. 6(d), the place enclosed by the black circle shows that CTV denoising, the estimation of the ECG signal will be slightly lower. Compared with TV and CTV, GSTV considers the structured information of the signal and has a better advantage in detail recovery. In order to test the influence of the parameter $\lambda$ and group size $K$ on the denoising effect, we calculated the relationship between them, and the ordinate is displayed as the value of RMSE, as shown in Fig. 7. The Red dot is the minimum RMSE when $K = 6$ and $\lambda = 2.6$ obtained. It is worth mentioning that these solutions used to illustrate TV, CTV, and GSTV are not completely different. In addition, in order to avoid contingency, we have taken multiple signals in the MIT-BIH library for comparison. The experimental results are in Figs. 8-9.
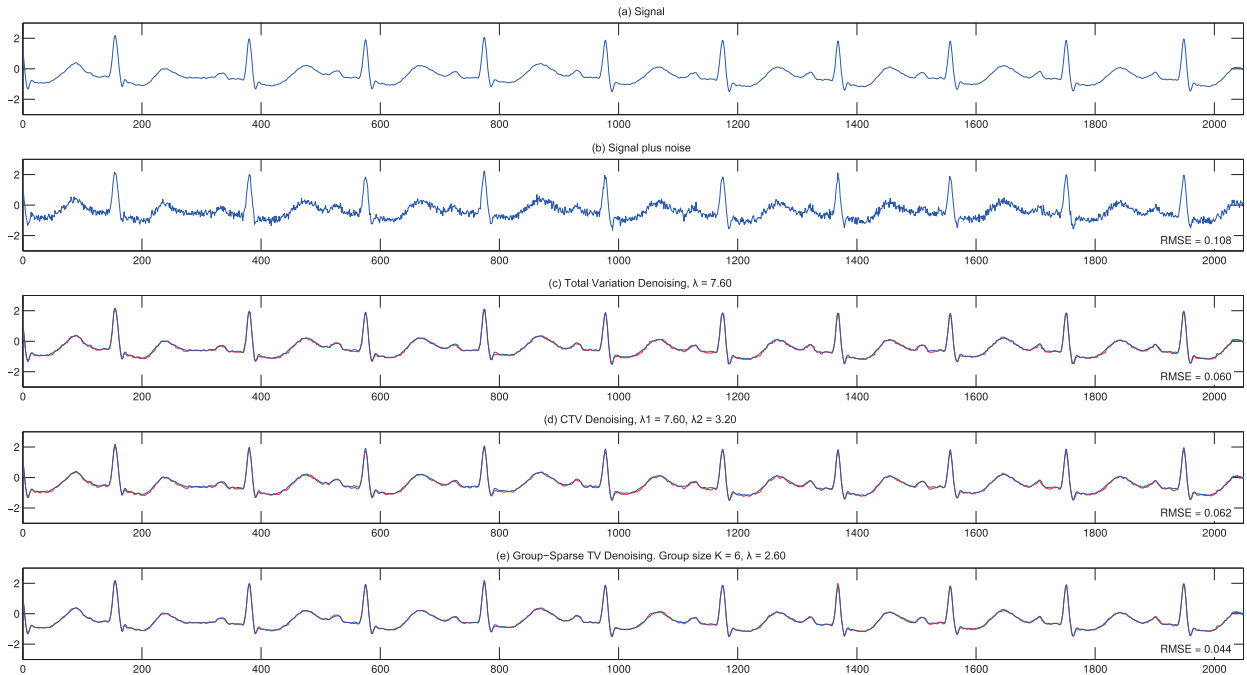
**FIGURE 5.** Denoising results (red line is the denoising result and blue line is the original signal). (a) original ECG signal without noise (MIT-BIH no.103). (b) ECG signal with 5 dB noise. (c) TV denoising result. (d) CTV denoising result. (e) GSTV denoising result.

**TABLE 1.** Comparison of the SNR_imp achieved by different methods with input SNR = 5 dB.

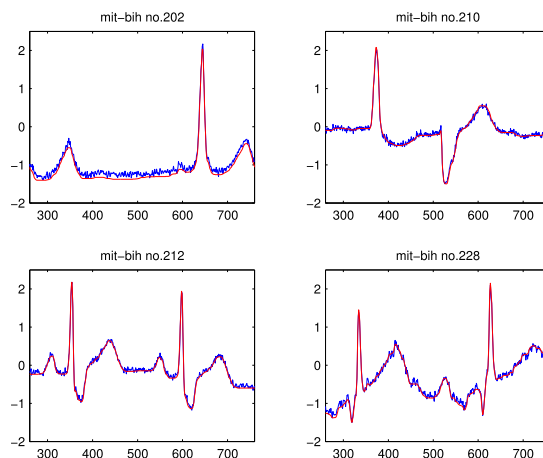| Dateset | GSTV | CTV | NLM | CEEMDAN | Dateset | GSTV | CTV | NLM | CEEMDAN |
|---|---|---|---|---|---|---|---|---|---|
| mitdb/100 | 6.814 | 7.253 | 6.843 | 6.046 | mitdb/202 | 8.005 | 8.546 | 7.704 | 7.596 |
| mitdb/101 | 6.534 | 6.927 | 6.567 | 6.029 | mitdb/203 | 5.154 | 6.325 | 4.537 | 3.448 |
| mitdb/102 | 6.256 | 6.1587 | 5.949 | 5.341 | mitdb/205 | 6.825 | 7.311 | 6.236 | 5.088 |
| mitdb/103 | 7.607 | 7.718 | 7.597 | 5.675 | mitdb/207 | 8.153 | 8.314 | 8.224 | 6.721 |
| mitdb/104 | 5.748 | 7.996 | 6.559 | 5.776 | mitdb/208 | 7.215 | 7.315 | 6.324 | 5.050 |
| mitdb/105 | 8.360 | 9.029 | 7.073 | 3.828 | mitdb/209 | 6.694 | 7.021 | 6.085 | 4.457 |
| mitdb/106 | 6.005 | 5.992 | 5.867 | 4.992 | mitdb/210 | 7.218 | 7.658 | 7.170 | 4.965 |
| mitdb/107 | 7.620 | 7.998 | 7.450 | 6.923 | mitdb/212 | 6.145 | 6.312 | 5.938 | 4.612 |
| mitdb/108 | 6.365 | 6.556 | 6.297 | 4.352 | mitdb/213 | 8.8202 | 8.6256 | 7.134 | 4.872 |
| mitdb/109 | 7.555 | 7.635 | 7.565 | 7.559 | mitdb/214 | 6.312 | 7.630 | 7.462 | 4.706 |
| mitdb/111 | 6.341 | 6.842 | 6.200 | 5.476 | mitdb/215 | 6.315 | 6.564 | 6.116 | 5.443 |
| mitdb/112 | 6.651 | 6.885 | 6.682 | 5.680 | mitdb/217 | 6.321 | 6.147 | 6.116 | 5.645 |
| mitdb/113 | 7.621 | 7.365 | 7.353 | 5.224 | mitdb/219 | 7.564 | 7.756 | 7.723 | 4.337 |
| mitdb/114 | 6.354 | 6.814 | 5.371 | 4.885 | mitdb/220 | 6.751 | 6.631 | 6.653 | 6.977 |
| mitdb/115 | 7.015 | 7.682 | 7.503 | 6.675 | mitdb/221 | 6.911 | 7.158 | 6.670 | 5.163 |
| mitdb/116 | 6.821 | 6.912 | 6.823 | 3.651 | mitdb/222 | 6.104 | 6.152 | 5.909 | 5.200 |
| mitdb/117 | 6.995 | 8.214 | 6.506 | 7.973 | mitdb/223 | 7.315 | 7.245 | 7.540 | 7.163 |
| mitdb/118 | 6.125 | 6.351 | 6.078 | 5.774 | mitdb/228 | 7.042 | 7.514 | 6.752 | 6.039 |
| mitdb/119 | 7.612 | 7.985 | 7.504 | 6.274 | mitdb/230 | 6.741 | 6.914 | 6.376 | 5.844 |
| mitdb/121 | 7.695 | 8.014 | 7.598 | 4.631 | mitdb/231 | 7.521 | 7.754 | 7.225 | 6.152 |
| mitdb/122 | 7.011 | 7.154 | 7.048 | 7.105 | mitdb/232 | 4.447 | 4.256 | 3.741 | 3.666 |
| mitdb/123 | 7.231 | 7.635 | 7.037 | 6.093 | mitdb/233 | 6.518 | 7.011 | 6.500 | 6.305 |
| mitdb/124 | 7.582 | 7.698 | 7.374 | 6.858 | mitdb/234 | 7.428 | 7.845 | 7.014 | 6.284 |
| mitdb/200 | 6.621 | 6.487 | 6.638 | 5.842 | Mean SNR_imp | 6.902 | 7.220 | 6.692 | 5.640 |
| mitdb/201 | 7.245 | 7.225 | 6.619 | 6.361 | Standard Deviation | 0.796 | 0.835 | 0.826 | 1.077 |

In Fig. 10, we compare the RMSE and SNR_imp of GSTV, CTV and TV methods at different SNRs Situation. In terms of RMSE indicators, GSTV is better than the TV and CTV methods, which means that is has a greater advantage in the fidelity of denoising signal. In terms of SNR_imp, the CTV method has more good performance. But when the SNR is

−5 dB, −10 dB, the GSTV method gradually approaches the CTV method and is better than the TV method. More information, we use a short-time Fourier transform (STFT) with 50% overlapping segments to show the spectral changes of the signal when the input SNR = 5 dB. The usage and formula derivation of STFT can be obtained in Ref. [32] Appendix B.
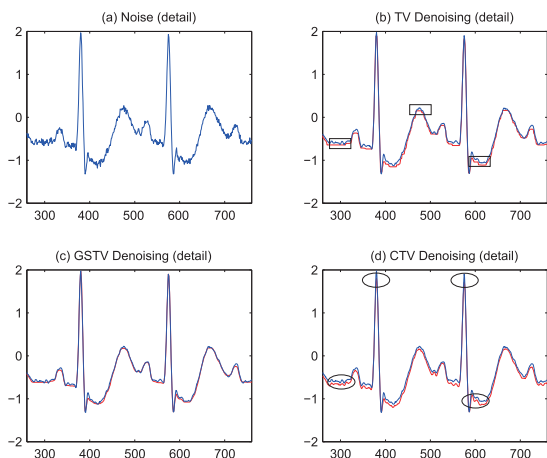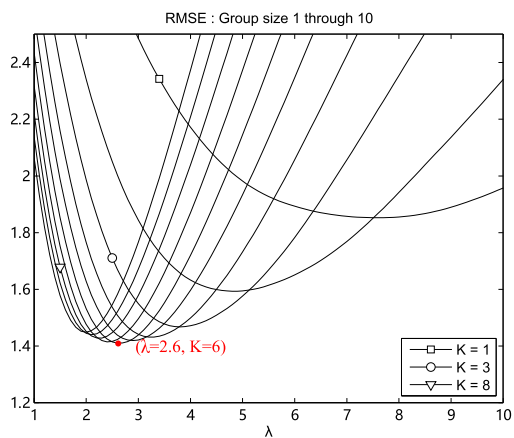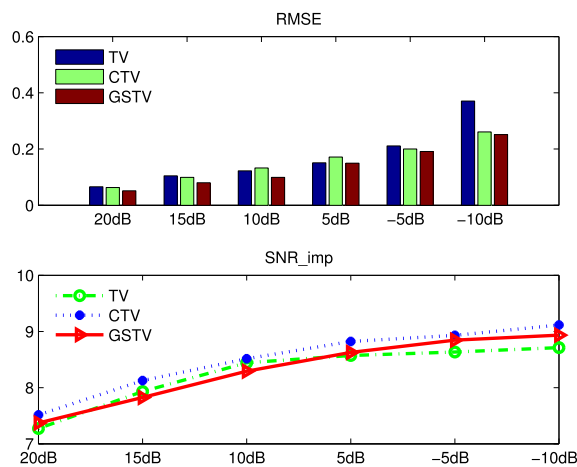
**FIGURE 6.** Detail comparison. (a) The detail component of noise. (b) The detail component of TV denoising. (c) The detail component of GSTV denoising. (d) The detail component of CTV denoising.
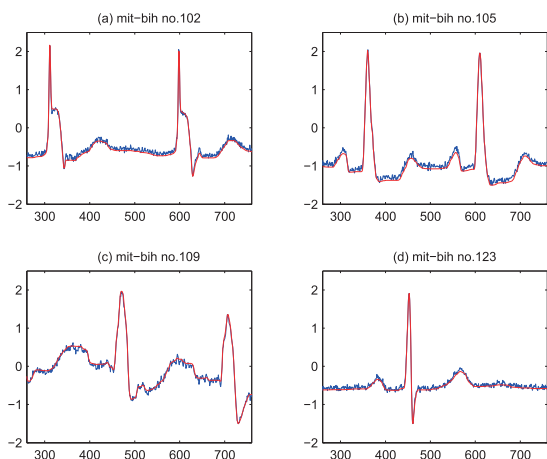


**FIGURE 7.** RMSE: Group size 1 through 10.



**FIGURE 8.** GSTV denoising detail comparison. (a) The detail component of MIT-BIH no.102. (b) The detail component of MIT-BIH no.105. (c) The detail component of MIT-BIH no.109. (d) The detail component of MIT-BIH no.123.

Fig. 11 shows the results, it can be found that after adding noise, the gray level of the signal frequency between 0.05khz and 0.2khz increases significantly, that is, the frequency range



**FIGURE 9.** GSTV denoising detail comparison. (a) The detail component of MIT-BIH no.202. (b) The detail component of MIT-BIH no.210. (c) The detail component of MIT-BIH no.212. (d) The detail component of MIT-BIH no.228.



**FIGURE 10.** Performance evaluation over RMSE and SNR_imp criteria.

of the noise. After denoising, the TV result has a certain weakening effect, but it is not ideal. CTV denoising guarantees the characteristics of 0 seconds and has a good denoising effect when the frequency is 0.05khz and 0.15khz. Compared with the existing advantages of TV and CTV, GSTV denoising has a better effect at 0.1khz. This can also prove that GSTV takes into account the structural characteristics of the signal. This method can combine the components of the signal with non-adjacent frequencies for analysis, although the sparsity of the first-order difference function of the signal is considered.

In order to fully illustrate the performance of the algorithm, the NLM [49] and CEEMDAN [52] as a reference for comparison. Table 1 lists the SNR_imp of these different methods when the input SNR = 5 dB in each piece of ECG data. Experimental results show that the SNR_imp effect of GSTV has certain advantages, though a little worse than the CTV method. In addition, we calculate the mean SNR_imp and the standard deviation of the SNR_imp of the above 48 ECG data, which is listed in Table 1. Among them, the mean improvement of GSTV is 6.902, less than the
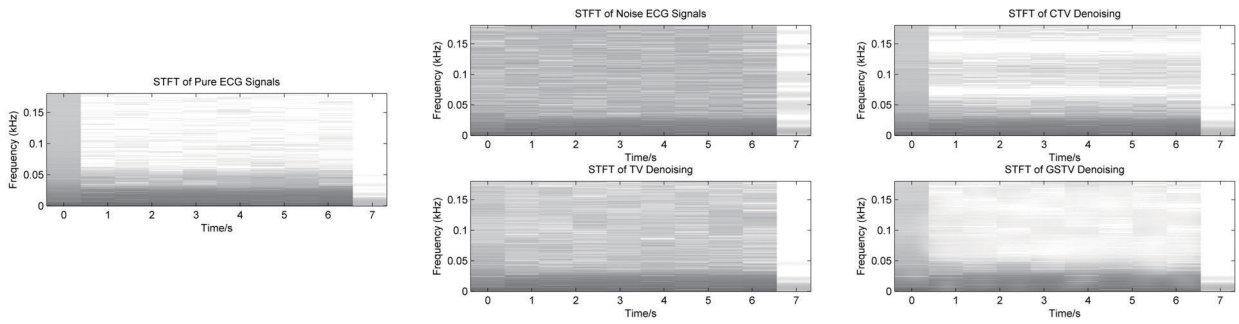
**FIGURE 11.** Comparison time-frequency properties of the TV, CTV and GSTV methods.
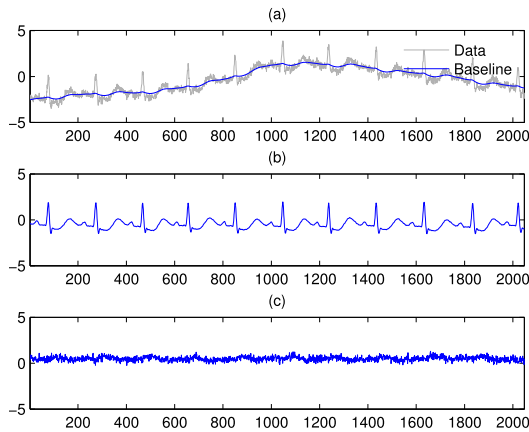


**FIGURE 12.** BW correction and denoising. (a) BW estimate and noise date. (b) Signal after denoising. (c) Residual.
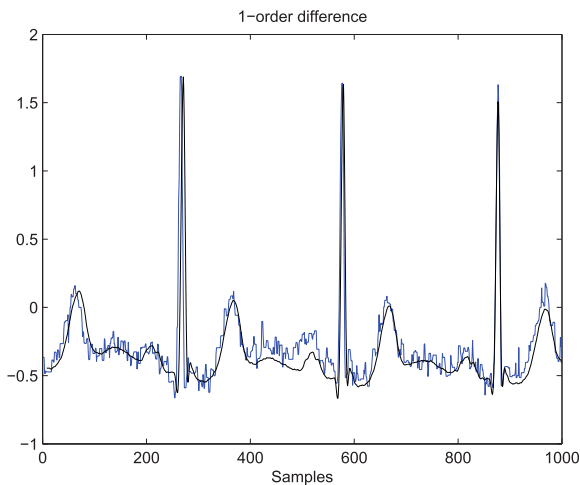


**FIGURE 13.** Comparison results between LPF/GSTV and the method in Ref. [34]. The blue line is the result of noise removal and restoration when the 1-order difference of the signal is selected. The black line is the recovery result of the LPF/GSTV method.



**FIGURE 14.** Comparison results between LPF/GSTV and the method in Ref. [34]. The blue line is the result of noise removal and restoration when the 2-order difference of the signal is selected. The black line is the recovery result of the LPF/GSTV method.

over the proposed method. For example, according to our test, the SNR_imp index of FMEMD can exceed 7dB with input SNR=5dB on data MIT-BIH no.101. The standard deviation value of the SNR_imp reflects the degree of dispersion relative to the mean value. The smaller the value, the lower the degree of dispersion. It can be seen from Table 1 that the standard deviation of the GSTV method is the smallest.

### E. COMPARISON OF HIGH-ORDER DIFFERENCE RESULTS

In this subsection, a real ECG signal with BW noises is taken as an example to illustrate the effectiveness of the algorithm. The simulation results are shown in Fig. 12. Fig. 12(a) is the estimation of the BW, Fig. 12(b) is the result after the ECG signals correction, and Fig. 12(c)is the residual. In the process of removing BW, the parameters we choose are still the original values ($o = 1$ and $f_c = 0.007$), and the residual is the noise component processed by the GSTV algorithm.

The thing worth discussing is that the LPF/GSTV method proposed in this paper is similar to the LPF joint asymmetric penalty denoising method in [34]. Inspired by the method in [34], we optimized and enriched our work. First of

denoising method of CTV, but better than NLM and CEEM-DAN. However, we highlight that the more advanced EMD methods like the fast multivariate empirical mode decomposition (FMEMD) [53] may bring performance improvement
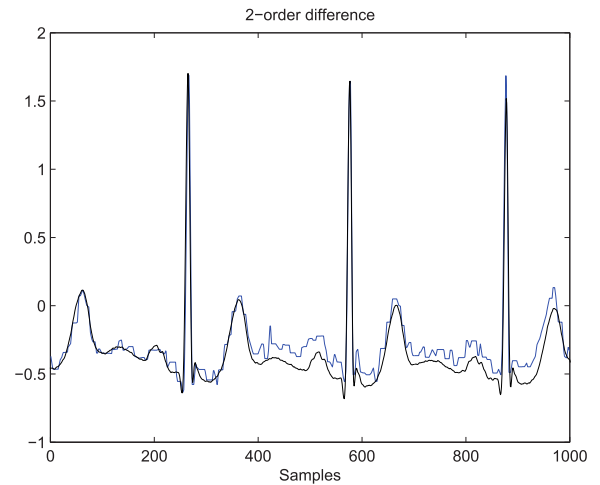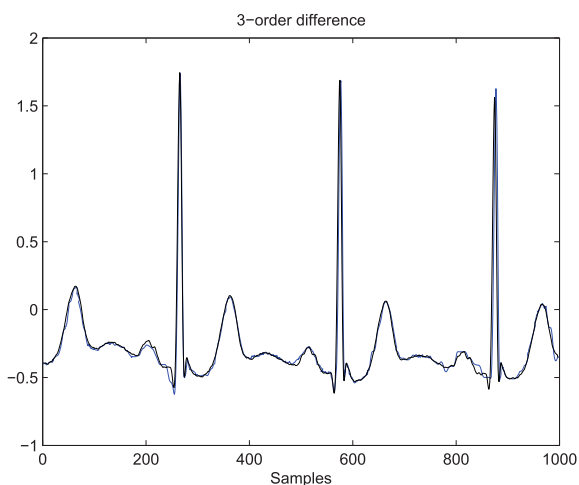
**TABLE 2.** Index (SNR_imp and RMSE ) comparison between LPF/GSTV and the method in Ref. [34] with input SNR = 5 dB.

| | | SNR_imp | RMSE |
|---|---|---|---|
| LPF/GSTV | | 8.820 | 0.015 |
| Ref. [34] | 1-order | 4.15 | 0.232 |
| | 2-order | 6.62 | 0.178 |
| | 3-order | 12.87 | 0.009 |
| | 4-order | 9.34 | 0.013 |



**FIGURE 15.** Comparison results between LPF/GSTV and the method in Ref. [34]. The blue line is the result of noise removal and restoration when the 3-order difference of the signal is selected. The black line is the recovery result of the LPF/GSTV method.
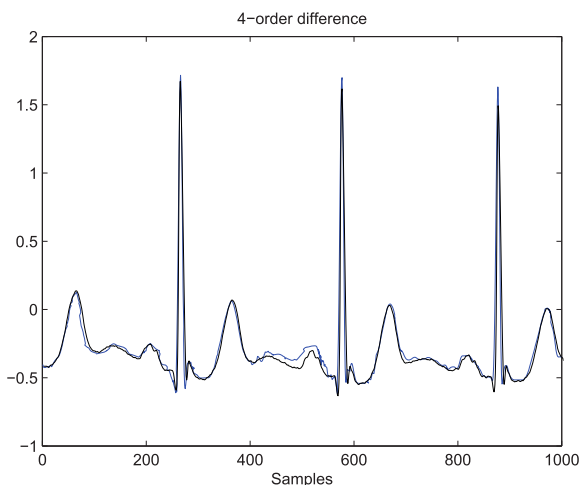


**FIGURE 16.** Comparison results between LPF/GSTV and the method in Ref. [34]. The blue line is the result of noise removal and restoration when the 4-order difference of the signal is selected. The black line is the recovery result of the LPF/GSTV method.

all, we need to point out the similarities with Ref. [34]: 1) We also combine the method of LPF to estimate the wandering of the baseline in ECG signals. 2) We all use the MM optimization algorithm to solve the optimization problem. 3) Both are based on the sparse characteristics of
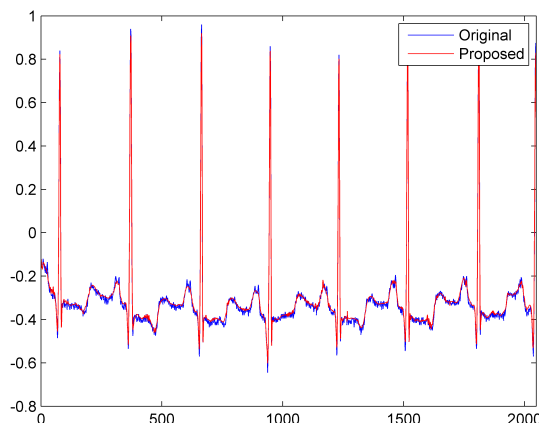


**FIGURE 17.** MIT-BIH no.100 correction and denoising.
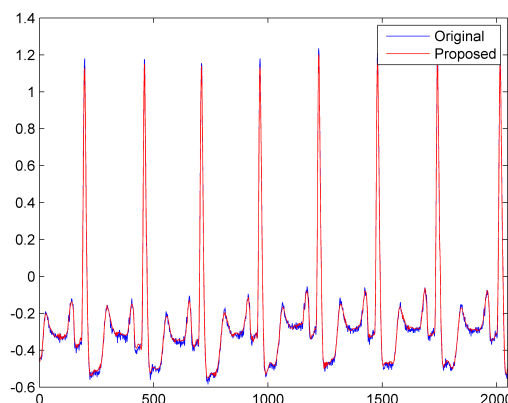


**FIGURE 18.** MIT-BIH no.105 correction and denoising.

the ECG signals. However, there are still some differences. 1) We consider the structural characteristics of the signal. In other words, we take the $K$ similar points in a piece of information as a whole, and use the TV denoising method to solve the sparse optimization. 2) When considering the third-order difference attribute of ECG information in Ref. [34], it has the best denoising result. In the whole calculation process, it is obvious that the amount of calculation is far greater than the method we proposed. For ECG signals, the best solution and recovery is when the third-order difference is true, but if it is applied to other fields, this still requires additional investigation and experimentation. In the following work, we compared the proposed LPF/GSTV method with the method in [34], and the experimental results is shown in Table 2. Figs. 13-16 shows the comparison between the proposed method and the method in [34] at different differential orders. The experimental data is MIT-BIH no.103 ECG data, and the SNR of the original noise added is 5dB. The blue line is the contrast method, and the black line is the method proposed in this paper.

It can be seen from Figs. 13-14 that when the signal is selected for low-order differential, the result of signal

recovery is completely unsatisfactory, and there is no basic information of normal signals. At this time, the parameters of SNR_imp and RMSE are meaningless. However, it is worth noting that when the signal selects the 4-order difference, the recovery effect is close to the proposed method, but not as good as the 3-order difference. In Figs. 15-16, the basic information of the signal is covered, and the denoising effect also has obvious changes. At the same time, in order to illustrate the generality of the algorithm, the denoising ability of the algorithm in the other two ECG signals (MIT-BIH no.100,105) are compared, and the result are shown in Figs. 17-18.

## V. CONCLUSION

Removing the BW and random noise in ECG signals is very important for the diagnosis of arrhythmia patients. This paper proposes an algorithm based on group sparse denoising to improve the traditional TV denoising method. In addition, we use LPF to estimate the components of BW. The group sparse characteristic of the signal is manifested in that the large value of the first-order difference function does not appear alone, and has a certain range attribute. The purpose of the proposed method is to remove the step artifacts caused by the TV denoising method. In order to illustrate the feasibility of the proposed method, we first did a unified conditional processing, and compared the LPF/TV and LPF/CTV methods as a horizontal comparison. The performance of the method is analyzed in details. We respectively give the convex cost functions of the three methods, and use the MM optimization algorithm to solve them. This method has the advantages of fast convergence and efficient calculation. In the longitudinal aspect, we compared this method with the NLM and CEEM-DAN methods to verify its performance.

Considering the high-order differential characteristics of the signal, the denoising effect is more obvious. This is also an incentive for us, considering whether the group sparsity denoising of high-order difference characteristics will have better processing results, this is a problem that we need to study and explore in our future work.

## REFERENCES

[1] M. Blanco-Velasco, B. Weng, and K. E. Barner, "ECG signal denoising and baseline wander correction based on the empirical mode decomposition," *Comput. Biol. Med.*, vol. 38, no. 1, pp. 1–13, Jan. 2008.

[2] B. Weng, B. Blanco-Velasco, and K. E. Barner, "Baseline wander correction in ECG by the empirical mode decomposition," in *Proc. IEEE 32nd Annu. Northeast Bioeng. Conf.*, Apr. 2006, pp. 135–136.

[3] D. Rui, L. I. Guo-Jun, and W. Qing, "The method research on removing baseline wander of ECG," *J. Yunnan Univ., Natural Ences Ed.*, pp. 655–660, May 2014.

[4] S. Agrawal and A. Gupta, "Fractal and EMD based removal of baseline wander and powerline interference from ECG signals," *Comput. Biol. Med.*, vol. 43, no. 11, pp. 1889–1899, Nov. 2013.

[5] K. Kærgaard, S. H. Jensen, and S. Puthusserypady, "A comprehensive performance analysis of EEMD-BLMS and DWT-NN hybrid algorithms for ECG denoising," *Biomed. Signal Process. Control*, vol. 25, pp. 178–187, Mar. 2016.

[6] N. ur Rehman and D. P. Mandic, "Filter bank property of multivariate empirical mode decomposition," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2421–2426, May 2011.

[7] Y. Xin, Y. Chen, and W. T. Hao, "ECG baseline wander correction based on mean-median filter and empirical mode decomposition," *Bio-Med. Mater. Eng.*, vol. 24, no. 1, pp. 71–365, 2014.

[8] B. Sharma and R. J. Suji, "ECG denoising using Weiner filter and adaptive least mean square algorithm," in *Proc. IEEE Int. Conf. Recent Trends Electron., Inf. Commun. Technol. (RTEICT)*, May 2016, pp. 53–57.

[9] X. Hu, S. Peng, and W.-L. Hwang, "Adaptive integral operators for signal separation," *IEEE Signal Process. Lett.*, vol. 22, no. 9, pp. 1383–1387, Sep. 2015.

[10] X. Yi, X. Hu, and S. Peng, "An operator-based and sparsity-based approach to adaptive signal separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 6186–6190.

[11] U. Biswas, A. Das, S. Debnath, and I. Oishee, "ECG signal denoising by using least-mean-square and normalised-least-mean-square algorithm based adaptive filter," in *Proc. Int. Conf. Informat., Electron. Vis. (ICIEV)*, May 2014, pp. 1–6.

[12] M. El Hanine, E. Abdelmounim, R. Haddadi, and A. Belaguid, "ElectroCardioGram signal denoising using discrete wavelet transform," in *Proc. Int. Conf. Multimedia Comput. Syst. (ICMCS)*, Apr. 2014, pp. 1065–1070.

[13] B. Arvinti, D. Toader, M. Costache, and A. Isar, "Electrocardiogram baseline wander removal using stationary wavelet approximations," in *Proc. 12th Int. Conf. Optim. Electr. Electron. Equip.*, May 2010, pp. 890–895.

[14] P. Singh, G. Pradhan, and S. Shahnawazuddin, "Denoising of ECG signal by non-local estimation of approximation coefficients in DWT," *Biocybernetics Biomed. Eng.*, vol. 37, no. 3, pp. 599–610, 2017.

[15] M. A. Awal, S. S. Mostafa, M. Ahmad, and M. A. Rashid, "An adaptive level dependent wavelet thresholding for ECG denoising," *Biocybern. Biomed. Eng.*, vol. 34, no. 4, pp. 238–249, 2014.

[16] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.

[17] R. Rubinstein, M. Zibulevsky, and M. Elad, "Double sparsity: Learning sparse dictionaries for sparse signal approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1553–1564, Mar. 2010.

[18] C. A. Metzler, A. Maleki, and R. G. Baraniuk, "From denoising to compressed sensing," *IEEE Trans. Inf. Theory*, vol. 62, no. 9, pp. 5117–5144, Sep. 2016.

[19] M. Nikolova, "An algorithm for total variation minimization and applications," *J. Math. Imag. Vis.*, vol. 20, no. 1/2, pp. 89–97, Jan. 2004.

[20] T. F. Chan, S. Osher, and J. Shen, "The digital TV filter and nonlinear denoising," *IEEE Trans. Image Process.*, vol. 10, no. 2, pp. 231–241, 2001.

[21] R. Chartrand and V. Staneva, "Total variation regularisation of images corrupted by non-Gaussian noise using a quasi-Newton method," *IET Image Process.*, vol. 2, no. 6, pp. 295–303, Dec. 2008.

[22] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[23] Z. Dou, B. Zhang, and X. Yu, "A new alternating minimization algorithm for total variation image reconstruction," in *Proc. Int. Conf. Wireless*, Apr. 2016, vol. 1, no. 3, pp. 248–272.

[24] L. A. Vese and S. J. Osher, "Image denoising and decomposition with total variation minimization and oscillatory functions," *J. Math. Imag. Vis.*, vol. 20, no. 1/2, pp. 7–18, Jan. 2004.

[25] J. L. Starck, Y. Moudden, J. Bobin, M. Elad, and D. L. Donoho, "Morphological component analysis," in *Proc. SPIE, Wavelets XI*, 2005, Art. no. 59140Q.

[26] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, "Bregman iterative algorithms for $\ell_1$-minimization with applications to compressed sensing," *SIAM J. Imag. Sci.*, vol. 1, no. 1, pp. 143–168, Jan. 2008.

[27] J. Bect, L. Blanc-Féraud, G. Aubert, and A. Chambolle, "A $\ell^1$-unified variational framework for image restoration," in *Proc. 8th Eur. Conf. Comput. Vis. Comput. Vis. (ECCV), Part IV*, Prague, Czech Republic, May 2004, pp. 1–13.

[28] J. Zhang, Z. Wei, and L. Xiao, "Fractional-order iterative regularization method for total variation based image denoising," *J. Electron. Imag.*, vol. 21, no. 4, pp. 043 005.1–043 005.9, 2012.

[29] Y. Zhou, X. Hu, Z. Tang, and A. C. Ahn, "Denoising and baseline correction of ECG signals using sparse representation," in *Proc. IEEE Workshop Signal Process. Syst. (SiPS)*, Oct. 2015, pp. 1–6.

[30] X. Ning, I. W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometric Intell. Lab. Syst.*, vol. 139, pp. 156–167, Dec. 2014.

[31] I. Selesnick, "Sparsity-assisted signal smoothing (revisited)," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 4546–4550.

[32] Z. Jin, A. Dong, M. Shu, and Y. Wang, "Sparse ECG denoising with generalized minimax concave penalty," *Sensors*, vol. 19, no. 7, p. 1718, Apr. 2019.

[33] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sep. 2017.

[34] X. Wang, Y. Zhou, M. Shu, Y. Wang, and A. Dong, "ECG baseline wander correction and denoising based on sparsity," *IEEE Access*, vol. 7, pp. 31573–31585, 2019.

[35] S. Boyd, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[36] J. Friedman, T. Hastie, H. Höfling, and R. Tibshirani, "Pathwise coordinate optimization," *Ann. Appl. Statist.*, vol. 1, no. 2, pp. 302–332, Dec. 2007.

[37] M. V. Afonso, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "An augmented lagrangian approach to linear inverse problems with compound regularization," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2010, pp. 4169–4172.

[38] J. M. Bioucas-Dias and M. A. T. Figueiredo, "An iterative algorithm for linear inverse problems with compound regularizers," in *Proc. 15th IEEE Int. Conf. Image Process.*, 2008, pp. 685–688.

[39] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[40] I. W. Selesnick, H. L. Graber, D. S. Pfeil, and R. L. Barbour, "Simultaneous low-pass filtering and total variation denoising," *IEEE Trans. Signal Process.*, vol. 62, no. 5, pp. 1109–1124, Mar. 2014.

[41] G. B. Moody, W. K. Muldrow, and R. G. Mark, "A noise stress test for arrhythmia detectors," *Comput. Cardiol.*, vol. 11, no. 3, pp. 381–384, 1984.

[42] G. Peyré and J. M. Fadili, "Group sparsity with overlapping partition functions," in *Proc. Eusipco*, Aug. 2011, pp. 303–307.

[43] R. Jenatton, J. Mairal, G. Obozinski, and F. Bach, "Proximal methods for sparse hierarchical dictionary learning," in *Proc. Int. Conf. Mach. Learn.*, Jun. 2010, pp. 21–24.

[44] P. L. Combettes and J.-C. Pesquet, "Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators," *Set-Valued Variational Anal.*, vol. 20, no. 2, pp. 307–330, Jun. 2012.

[45] C. Ng, *A Splitting Algorithm for Dual Monotone Inclusions Involving Cocoercive Operators*. New York, NY, USA: Springer-Verlag, 2013.

[46] M. A. T. Figueiredo, J. B. Dias, J. P. Oliveira, and R. D. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising," in *Proc. Int. Conf. Image Process.*, Atlanta, GA, USA, Oct. 2006, pp. 2633–2636.

[47] C. Lu, J. Feng, S. Yan, and Z. Lin, "A unified alternating direction method of multipliers by majorization minimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 3, pp. 527–541, Mar. 2018.

[48] P. L. Combettes and J. C. Pesquet, "Proximal splitting methods in signal processing," in *Proc. Fixed-Point Algorithms Inverse Problems Sci. Eng.*, vol. 49, 2011, pp. 185–212.

[49] S. Kumar, D. Panigrahy, and P. K. Sahu, "Denoising of electrocardiogram (ECG) signal by using empirical mode decomposition (EMD) with non-local mean (NLM) technique," *Biocybern. Biomed. Eng.*, vol. 38, no. 2, pp. 297–312, 2018.

[50] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imag. Vis.*, vol. 40, no. 1, pp. 120–145, May 2011.

[51] L. M. Briceño-Arias and P. L. Combettes, "A monotone+skew splitting model for composite monotone inclusions in duality," *SIAM J. Optim.*, vol. 21, no. 4, pp. 1230–1250, 2010.

[52] Y. Xu, M. Luo, T. Li, and G. Song, "ECG signal de-noising and baseline wander correction based on CEEMDAN and wavelet threshold," *Sensors*, vol. 17, no. 12, p. 2754, Nov. 2017.

[53] X. Lang, Y. Zhang, L. Xie, X. Jin, A. Horch, and H. Su, "Use of fast multivariate empirical mode decomposition for oscillation monitoring in noisy process plant," *Ind. Eng. Chem. Res.*, vol. 59, no. 25, pp. 11537–11551, Jun. 2020.

**HAO SHI** received the B.E. degree from the Qilu University of Technology. He is currently pursuing the M.E. degree with the Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. His research interests include signal processing, medical artificial intelligence, medical big data, and the medical Internet of Things.
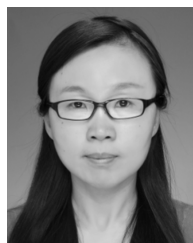


**RUIXIA LIU** received the M.S. degree from the Shaanxi University of Science and Technology, in 2004, and the Ph.D. degree in information science and engineering from the Shandong University of Science and Technology, Qingdao, China, in 2017. She is currently an Associate Researcher with the Shandong Computer Science Center (National Supercomputer Center in Jinan). Her research interests include medical information and artificial intelligence.



**CHANGFANG CHEN** received the B.S. degree in mathematics and the M.S. degree in control theory from Qufu Normal University, Qufu, China, in 2005 and 2008, respectively, and the Ph.D. degree in control theory and applications from Beihang University (BUAA), Beijing, China, in 2013. She is currently an Associate Professor with the Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. Her main research interests include nonlinear systems and digital signal processing.



**MINGLEI SHU** received the M.S. degree in power electronics and the Ph.D. degree in communication and information systems from Shandong University, Jinan, China, in 2006 and 2016, respectively. He is currently a Researcher with the Shandong Artificial Intelligence Institute, Qilu University of Technology (Shandong Academy of Sciences), where he is also the Deputy Director of the Shandong Artificial Intelligence Institute, with major research areas including artificial intelligence, the Internet of Things (IoT) medical care, and wireless sensor networks.



**YINGLONG WANG** received the M.S. degree in industrial automation and the Ph.D. degree in communication and information systems from Shandong University, Jinan, China, in 1990 and 2005, respectively. He is currently a Research Fellow with the Shandong Computer Science Center (National Supercomputer Center in Jinan). His current research interests include wireless networks, information security, and cloud computing.

• • •