IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 13, 2019, accepted February 27, 2019, date of publication March 4, 2019, date of current version March 25, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2902616

ECG Baseline Wander Correction and Denoising Based on Sparsity

XIAO WANG[1], YOU ZHOU[1], MINGLEI SHU[1], (Member, IEEE), YINGLONG WANG[1], (Member, IEEE), AND ANMING DONG[2], (Member, IEEE)

[1]Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan 250014, China
[2]School of Computer Science and Technology, Qilu University of Technology (Shandong Academy of Sciences), Jinan 250353, China

Corresponding author: Minglei Shu (smlsmile1624@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61603224 and Grant 61701269, in part by the Natural Science Foundation of Shandong Province under Grant ZR2017MF029, and in part by the Joint Research Fund for Young Scholars, Qilu University (Shandong Academy of Sciences), under Grant 2017BSHZ005.

ABSTRACT To reduce the influence of both the baseline wander (BW) and noise in the electrocardiogram (ECG) is much important for further analysis and diagnosis of heart disease. This paper presents a convex optimization method, which combines linear time-invariant filtering with sparsity for the BW correction and denoising of ECG signals. The BW signals are modeled as low-pass signals, while the ECG signals are modeled as a sequence of sparse signals and have sparse derivatives. To illustrate the positive of the ECG peaks, an asymmetric function and a symmetric function are used to punish the original ECG signals and their difference signals, respectively. The banded matrix is used to represent the optimization problem, in order to make the iterative optimization method more computationally efficient, take up the less memory, and apply to the longer data sequence. Moreover, an iterative majorization-minization algorithm is employed to guarantee the convergence of the proposed method regardless of its initialization. The proposed method is evaluated based on the ECG signals from the database of MIT-BIH Arrhythmia. The simulation results show the advantages of the proposed method compared with wavelet and median filter.

INDEX TERMS Baseline wander correction, ECG denoising, convex optimization, sparsity.

I. INTRODUCTION

Electrocardiogram (ECG) is a time-varying signal which provides some information of the heart and disease. It is widely used for the diagnosis to detect whether the heart is healthy or not. In the process of the signal acquisition, ECG signal is inevitably polluted by different kinds of interference, including baseline wander (BW) and random noise. The BW of ECG signal is mainly caused by the noise made by the breath or movement of the test people, which is a common problem and cannot be avoided. The BW has a low frequency, leading the overlap with the ST segment of ECG, which eventually influences the accuracy of the doctor judgment on the myocardial ischemia and other diseases [1]. At the same time, the spectrum of the noise is consistent with the ECG signal, making it difficult to separate the ECG signal from the noise, and the traditional filter is invalid at the receiver.

The associate editor coordinating the review of this manuscript and approving it for publication was Mohammad Zia Ur Rahman.

Considering that BW correction and denoising for ECG signal will both provide better assistance for the diagnosis, to reduce the effect of the BW and additive noise has been a hot topic in the area of ECG signal processing during the past decades. In terms of BW correction, a nonlinear filter bank was utilized in [2], leading to the stepped waveform distortion. To solve this problem, the ECG signal based on empirical mode decomposition (EMD) for BW correction was proposed in [3]. In [4], a multivariate EMD (MEMD) method for BW correction of ECG signals was presented. The ensemble EMD (EEMD) integrated empirical mode decomposition and the least mean square (LMS) algorithm was combined in [5]. A method combined fractal modeling with EMD was proposed in [6]. Considering that EMD decomposition is more complicated and the loss of waveform details is easily caused, a novel method of BW correction based on the combination of mean-median filter and EMD was proposed in [7]. In [8], a technique based on Hilbert vibration decomposition (HVD) was proposed to correct the BW. Based on the

results in [8], a method based on iterative Hilbert transform and HVD was proposed in [9], which improved the efficiency of the algorithm.

On the other hand, in terms of ECG signal denoising, wavelet algorithm is one of the most effective methods and has been researched recently in [10]–[13]. An improved Symlet wavelet threshold for ECG signal denoising was set up in [10]. Adaptive double threshold filtering (ADTF) and discrete wavelet transform were proposed in [11]. Discrete wavelet transformation and nonlocal mean (NLM) estimation was combined in [12]. A new algorithm that combines β-hill climbing metaheuristic with wavelet transform was designed in [13]. However, in these methods, the amplitudes of the R-wave and S-wave of the ECG signal will decrease and the underestimation will occur after the denoising, resulting in the loss of the characteristics and useful information of the ECG signal. As one of the new branches for the signal processing, sparse representation has been widely used in signal denoising [14]–[16], especially total variation (TV) denoising based on signal sparsity [17]–[19]. In the process of traditional TV denoising, stair-case artifacts are easily produced and the high-amplitude components are often underestimated. To improve this situation, some advanced TV denoising and algorithms were proposed [20], [21]. In [20], a denoising filter where the signal comprises a low-frequency component and a sparse or sparse derivative component was designed, using the majorization-minimization (MM) and alternating direction method of multipliers (ADMM) [22] to solve the constructed optimization problems. Based on the research in [21], the joint suppression method of BW and noise for chromatogram was designed, as well as the advanced penalty function on the basis of TV denoising. Some state-of-the-art denoising methods still have limitations. In [23], the problem of waveform unsmoothness in the sparse denoising of ECG signal is solved, but the peak underestimation is still exist. In [24], the sparse penalty function is further improved based on [23] to solve the problem of peak underestimation, however, the BW interference cannot be effectively removed.

Taking into consideration that BW correction and denoising are considered individually in most existing technologies, the researchers design the methods with combination of the both for efficiency. In [25] a method based on wavelet is proposed, but wavelet transform cannot effectively remove the smooth varying BW interference. In [26], training dictionary is used for sparse representation of given ECG signals to learn more signal details, however, the algorithm is time-consuming, and the peak underestimation is still obvious.

In order to further reduce the influence of BW interference and peak underestimation, a joint ECG BW correction and denoising based on LTI filtering and sparsity method is proposed in this paper. The main contributions are listed.

1) The sparsity of ECG signal and the derivative are analyzed. An asymmetric penalty function is utilized, which could not only promote the sparsity of the ECG signal, but also realize the precise punishment by taking into account the positive and negative peaks of the ECG signal.

2) The effect of different difference orders on the denoising of ECG signal is compared by experiments. The results show that the denoising effect with high-order difference is better than low-order difference and could completely recover the detailed features of the original ECG signal.

3) The LTI filtering and the sparse denoising algorithm are applied to the ECG signal for BW correction and denoising. The issue of the peak underestimation of the ECG signal is effectively improved, and the performance of the algorithm is systematically proposed.

This paper is organized as follows. The related techniques and mathematical methods are presented in Section II. In Section III, the optimization problem of ECG joint BW correction and denoising is proposed and the algorithms system model is given. In Section IV, an iterative algorithm for solving optimization problems is derived. Simulations for BW correction and denoising are conducted in Section V. Detail discussions of the simulation results under various algorithm are also provided. Section VI draws the conclusion.

II. PRELIMINARIES

A. NOTATION

The lower- and upper-case bold denote the vectors and matrices, respectively. The N-point signal $\mathbf{x}$ with the length of N is represented by the vector $\mathbf{x} = [x(0), \ldots, x(N-1)]^T$, where $[\cdot]^T$ is the transpose. All derivatives could be expressed as finite differences for the reason that all the works are considered in the discrete-data domain in this paper. The first-order difference matrix of size $(N-1) \times N$ is defined as

$$\mathbf{D}_1 = \begin{bmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \end{bmatrix}, \tag{1}$$

and the second-order difference matrix of size $(N-2) \times N$ can be defined as

$$\mathbf{D}_2 = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}. \tag{2}$$

Generally, the matrix of size $(N-k) \times N$ with the difference operator of order k is denoted as $\mathbf{D}_k$. $\mathbf{D}_0$ is defined as the identity matrix, i.e., $\mathbf{D}_0 = \mathbf{I}$. The $\ell 1$ and $\ell 2$ norms are defined as

$$\|\mathrm{x}\|_1 = \sum_n |x_n|, \quad \|\mathrm{x}\|_2^2 = \sum_n |x_n|^2, \tag{3}$$

respectively.

B. MAJORIZATION-MINIZATION

In the optimization, a difficult minimization problem can be replaced by a sequence of simpler ones through MM approach [27], [28]. In the following, to minimize a convex function $F(\mathbf{x})$, MM approach solves a sequence of simpler minimization problems, using the iteration

$$\mathbf{x}^{(k+1)} = \arg\min_{\mathbf{x}} G(\mathbf{x}, \mathbf{x}^{(k)}), \tag{4}$$

where $G(\mathbf{x}, \mathbf{v})$ is a convex majorizor of $F(\mathbf{x})$ that coincides with $F(\mathbf{x})$ at $\mathbf{x} = \mathbf{v}$ in each iteration. i.e., $G(\mathbf{x}, \mathbf{v}) \geq F(\mathbf{x})$ for all $\mathbf{x}$, and $G(\mathbf{v}, \mathbf{v}) = F(\mathbf{v})$. More details can be obtained from [27]. The update Eq. (4) will produce a sequence $\mathbf{x}^{(k)}$ converging to the minimizer of $F(\mathbf{x})$ after the procedure of initialization $\mathbf{x}^{(0)}$.

C. TOTAL VARIATION DENOISING

TV denosing is widely used for the sparse signal and image processing denoising, as well as other aspects, e.g., signal restoration, reconstruction and deconvolution. The method is also utilized for other signals [17], [29], [30]. It would suppress the noise effectively when the derivative of the signal is sparse, and could be regarded as a convex optimization problem containing a quadratic data fidelity.

The denoising of sparse derivative signal refers to the problem of estimating the signal $\mathbf{x}$, which has a sparse or sparse derivative. The model can be formed as

$$\mathbf{y} = \mathbf{x} + \mathbf{w}, \tag{5}$$

where $\mathbf{w}$ denotes stationary white Gaussian noise. It is well known that $\ell 1$ norm is a convex proxy for sparsity. The objective function of TV denoising contains a non-differential term and its solution has been wildly discussed [31], [32]. Assume the noise which follows stationary white Gaussian noise with variance σ^2 is added to the N-point signal $\mathbf{x}$. Considering that the appropriate data fidelity constraint is $\|\mathbf{y} - \mathbf{x}\|_2^2 \leq N\sigma^2$, the estimation of $\mathbf{x}$ was formulated as an $\ell 1$ norm based on convex optimization problem, i.e.,

$$\arg\min_{x} \ \|\mathbf{D}\mathbf{x}\|_1, \tag{6a}$$

$$\|\mathbf{y} - \mathbf{x}\|_2^2 \leq N\sigma^2. \tag{6b}$$

Using a control parameter λ, the optimization problem can converse to the unconstrained one as follows,

$$\arg\min_{\mathbf{x}} \{\frac{1}{2}\|\mathbf{y} - \mathbf{x}\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1\}. \tag{7}$$

The increase of the parameter λ could make the solution $\mathbf{x}$ closer to the piecewise constant. The same as the first-order difference method, other derivative (e.g., second or higher derivatives) approximation methods can be used for sparse derivative denoising.

D. LTI FILTERS

LTI filters are usually modeled to be recursive difference equations [33]. Considering that LTI filters are most suitable for signals limited to the known frequency band, they are basically used as the high-pass filter to cut off the lower-frequency components. One of the application is for BW of ECG signal. A linear phase filter could effectively avoided phase distortion which may alter various temporal relationships in the cardiac cycle. In this paper, a proper discrete-time filter is designed to make the proposed approach more effective and efficient. In other words, the data processing of the zero-phase, non-causal, recursive filter with finite length is described in the form of a banded matrix, which could transform to the sparse optimization to design LTI filters.

In order to avoid the unnecessary distortion, the filters $\mathbf{L}$ and $\mathbf{H}$ are taken to be zero-phase, non-causal, and recursive filters. That means the matrices $\mathbf{A}$ and $\mathbf{B}$ have the specific zero-phase property, and the shifts in peak locations would not appear in the filters. A procedure for defining such filters is given in [20]. The filter is specified by two parameters: its order $2d$ and its cutoff frequency f_c. In this paper the high-pass filter $\mathbf{H}$ is set to be the form

$$\mathbf{H} = \mathbf{B}^{-1}\mathbf{A}, \tag{8}$$

where $\mathbf{A}$ and $\mathbf{B}$ are banded convolution matrices, and both represent LTI systems. $\mathbf{H}$ represents a cascade of LTI systems.

A zero-phase non-causal high-pass filter is described by the difference equation as

$$\begin{aligned} j_1 y(n+1) + j_0 y(n) + j_1 y(n-1) \\ = -x(n+1) + 2x(n) - x(n-1), \end{aligned} \tag{9}$$

and $\mathbf{y} = \mathbf{A}^{-1}\mathbf{B}\mathbf{x}$ is defined, so the transfer function of the filter Eq. (9) is given by

$$H(z) = \frac{B(z)}{A(z)} = \frac{-z + 2 - z^{-1}}{j_1 z + j_0 + j_1 z^{-1}}. \tag{10}$$

Considering the transfer function on the higher-order high-pass filter, Eq. (10) could transform to

$$H(z) = \frac{(-z + 2 - z^{-1})^d}{(-z + 2 - z^{-1})^d + \beta(z + 2 + z^{-1})}. \tag{11}$$

The low-pass filter and the high-pass filter have the relationship shown as $\mathbf{H} = \mathbf{B}^{-1}\mathbf{A}$, so the low-pass filter $L(z)$ has the transfer function

$$L(z) = \frac{\beta(-z + 2 - z^{-1})^d}{(-z + 2 - z^{-1})^d + \beta(z + 2 + z^{-1})^d}. \tag{12}$$

E. SPARSITY OF ECG SIGNAL

A standard ECG signal is periodical and can be sectioned as several pieces according to the features of the waves, which are named as P, Q, R, S and T waves, respectively. As shown in Fig. 1, the ECG signal consists of abrupt peaks returning to a relatively flat baseline, and exhibits a form of sparsity. The low- and high- order derivatives of the original ECG signal are plotted in Fig. 2 and Fig. 3, respectively. It could be seen that the difference ECG signal is sparser than the original one, and the difference value tends to be zero when the order is

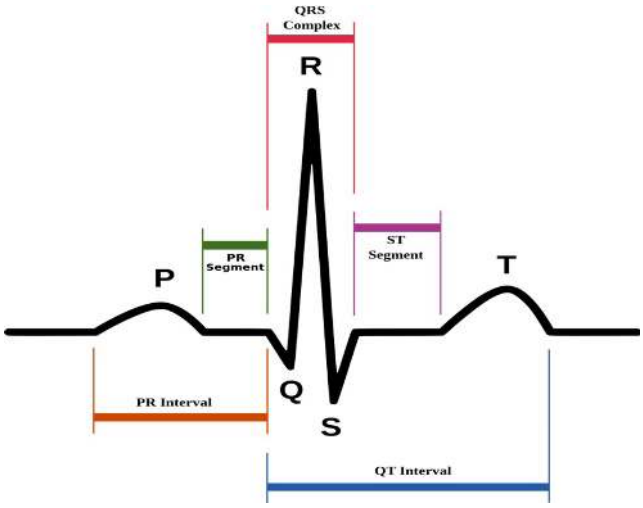


FIGURE 1. Basic pattern of ECG signal.

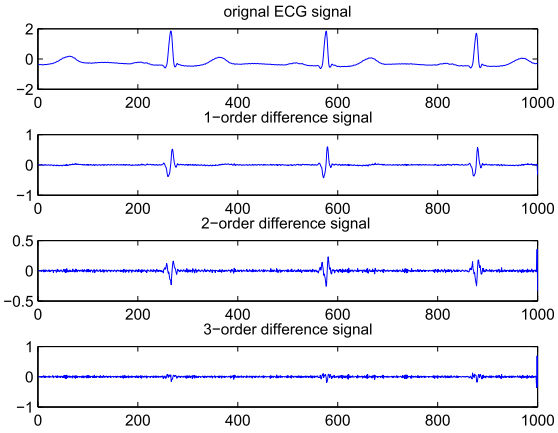


FIGURE 2. Examples for the derivatives of ECG signal with low-order difference.

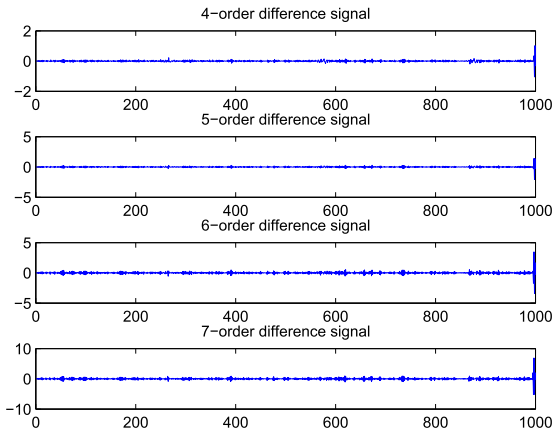


FIGURE 3. Examples for the derivatives of ECG signal with high-order difference.

higher. The difference signal will no longer change obviously when the difference order reaches to four. Therefore, the ECG signal can be regarded as a sparse one.

III. SYSTEM MODEL

A. PROBLEM FORMULATION

A mixed ECG signal consists of three parts: the original signal, BW interference, and noise signal. It is modeled as a N-point data vector

$$\mathbf{y} = \mathbf{x} + \mathbf{f} + \mathbf{n}, \tag{13}$$

where $\mathbf{x}$ is the original ECG signal with a sparse K-order derivative, $\mathbf{f}$ represents the low-pass BW signal, and $\mathbf{n}$ denotes the additive noise which follows stationary white Gaussian noise with variance σ^2.

Generally, the BW is a low-pass signal, and its frequency is typically less than 0.5 Hz. LTI filter is most suitable when the signal is approximately restricted to a known frequency band. The BW could be removed by low-pass filtering. In the following, a simplified case is considered where the low-pass signal is observed in noise alone. In this situation, the observed signal $\mathbf{y}$ is the mixture of BW signal $\mathbf{f}$ and the white Gaussian noise $\mathbf{n}$. Then $\mathbf{f}$ can be recovered approximately from $\mathbf{y}$ through a low-pass filtering, i.e.,

$$\hat{\mathbf{f}} = \mathrm{LPF}(\mathbf{y}) \approx \mathrm{LPF}(\mathbf{f} + \mathbf{n}), \tag{14}$$

where LPF represented the low-pass filtering.

In the second step, the situation that the ECG signal $\mathbf{x}$ is present in the observed signal $\mathbf{y}$ is considered. The estimation of the ECG signal $\mathbf{x}$, which denotes as $\hat{\mathbf{x}}$, can be obtained beforehand. Now the signal contains sparse signal or sparse derivative component along with noise, i.e., $\mathbf{y} = \mathbf{x} + \mathbf{f} + \mathbf{n}$. A simple optimization is utilized to estimate the $\mathbf{f}$ and $\mathbf{x}$ individually. Given the estimation $\hat{\mathbf{x}}$, $\mathbf{f}$ could be estimated as

$$\hat{\mathbf{f}} = \mathrm{LPF}(\mathbf{y} - \hat{\mathbf{x}}), \tag{15}$$

in which

$$\hat{\mathbf{f}} \approx \mathbf{f}, \quad \hat{\mathbf{x}} \approx \mathbf{x}, \tag{16}$$

are assumed. Substituting Eq. (16) in Eq. (15) gives

$$\mathbf{f} \approx \mathrm{LPF}(\mathbf{y} - \hat{\mathbf{x}}). \tag{17}$$

Therefore, the problem is to find $\hat{\mathbf{x}}$.

Substituting Eq. (13) in Eq. (17) gives

$$(\mathbf{y} - \hat{\mathbf{x}}) - \mathrm{LPF}(\mathbf{y} - \hat{\mathbf{x}}) = \mathbf{I} - \mathrm{LPF}(\mathbf{y} - \hat{\mathbf{x}}) \approx \mathbf{n}. \tag{18}$$

Note that the left-hand side of Eq. (18) constitutes a high-pass filter of $\mathbf{y} - \hat{\mathbf{x}}$. Define $\mathrm{HPF} = \mathbf{I} - \mathrm{LPF}$, Eq. (18) can be written as

$$\mathrm{HPF}(\mathbf{y} - \hat{\mathbf{x}}) \approx \mathbf{n}. \tag{19}$$

Considering $\mathbf{H}$ represents the high-pass filter matrix, the Eq. (19) can be written as

$$\mathbf{H}(\mathbf{y} - \hat{\mathbf{x}}) \approx \mathbf{n}. \tag{20}$$

On the other hand, it could be noticed that Eq. (13) is a highly underestimated equation and could be regarded as an ill-posed or N-P hard problem [34], [35]. Infinite solutions would appear for the reason that the number of unknowns is

larger than that of equations. Generally, convex optimization techniques are used to estimate transient components from observed signals. Based on TV denoising, the estimation of $\hat{\mathbf{x}}$ could be formulated as the convex optimization problem, i.e.,

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}}\{\frac{1}{2}\|\mathbf{H}(\mathbf{y}-\mathbf{x})\|_2^2 + \lambda\|\mathbf{D}\mathbf{x}\|_1\}. \tag{21}$$

After $\hat{\mathbf{x}}$ is obtained, $\mathbf{f}$ can be estimated according to Eq. (15).

B. SYMMETRIC PENALTY SPARSE DIFFERENCE MODEL

According to Eq. (13), $\mathbf{x}$ is sparse and has a sparse derivative. In sparse signal processing, the behavior of sparse signal is usually realized by using proper non-quadratic regularization terms. Under the high-order situation, the optimization problem on estimation of $\hat{\mathbf{x}}$ is

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}}\{F(\mathbf{x}) = \frac{1}{2}\|\mathbf{H}(\mathbf{y}-\mathbf{x})\|_2^2 + \sum_{i=1}^{K}\lambda_i\sum_{n=0}^{N_i-1}\phi([\mathbf{D}_i\mathbf{x}]_n)\}, \tag{22}$$

where ϕ refers to a penalty function, and the number of constraints of punishment depends on K. More discussion about the order of difference and the choice of the parameter λ_i will be given in Sec. V. Eq. (22) could promote sparsity of both $\mathbf{x}$ and its K-order difference. The high-pass filter is the same as the equation $\mathbf{H} = \mathbf{B}^{-1}\mathbf{A}$.

In TV denoising, the sparse-derivative denoising could be formulated as a minimizing problem for the $\ell 1$ norm of the derivative of $\mathbf{x}$, which subjects to the data fidelity constraint. The function $\phi_A = |x|$ may lead to the problem of the $\ell 1$ norm regularization, i.e., the function is not differentiable at zero. To address this issue, a differentiable approximation of the $\ell 1$ penalty function is proposed as

$$\phi_B(x) = |x| - \rho log(|x| + \rho). \tag{23}$$

Considering that the $\phi_B(x) = |x| - \rho log(|x| + \rho)$ degrades into the absolute value function $\phi_A(x)$ when $\rho = 0$, ρ should be set to follow the constant $\rho > 0$ and $\rho \approx 0$. On the one hand ρ should be presented to be an adequate small value to make sure the penalty function is smooth and original non-differentiable penalty function is sparse. On the other hand, ρ should be relatively large enough to avoid some problems arising from the numerical optimization algorithm. In this paper, $\rho = 10^{-6}$ is set up for the reason that the numerical issues are efficiently avoided and the effect on the optimal solution is negligible.

C. ASYMMETRIC PENALTY SPARSE DIFFERENCE MODEL

The signal $\mathbf{x}$ maybe asymmetrically sparse for some applications. For example, ECG signal has positive peaks on a relatively flat baseline, and its original signal is asymmetrically sparse. In this paper the asymmetric penalty [36] is used for

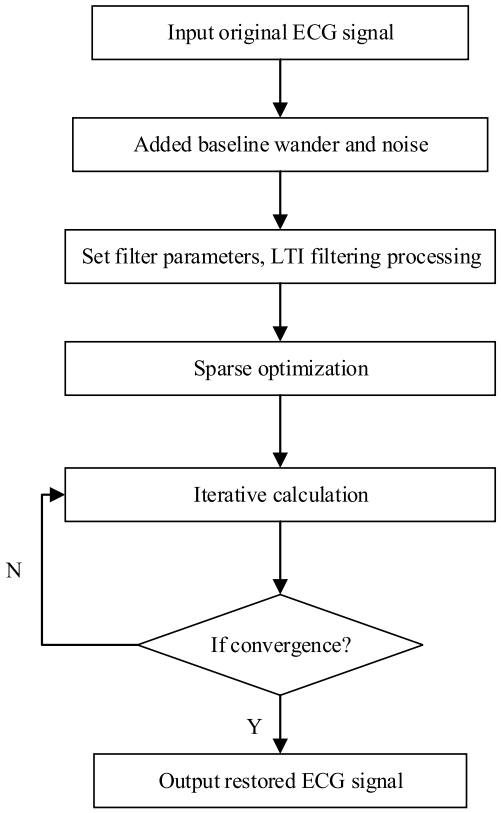


FIGURE 4. The flow-diagram of the proposed method.

punishment on positive and negative. θ is defined as

$$\theta(x;r) = \begin{cases} x, & x \geq 0 \\ -rx, & x < 0 \end{cases} \tag{24}$$

where $r > 0$ is a positive constant. It could be noticed that Eq. (24) has the same drawback as $\phi_A = |x|$, i.e., it is non-differentiable at $x = 0$. To solve this problem, a differentiable form is proposed as

$$\theta_\rho(x;r) = \begin{cases} x, & x > \rho \\ f(x), & |x| \leq \rho \\ -rx. & x < -\rho \end{cases} \tag{25}$$

Excluding the intermediate function $f(x)$, Eq. (25) will transform to the absolute value function $\phi_A(x) = |x|$ when $r = 1$. Therefore, the main problem is how to construct intermediate functions $f(x)$, and the answer will be given in Sec. IV.

IV. COMPOUND OPTIMIZATION ALGORITHM

The flow-diagram of the proposed method is shown in Fig. 4. Two algorithms for the spare optimization in the flow-diagram are designed in this section. One is based on the symmetric penalty sparse difference algorithm, and the other is a compound algorithm based on symmetric and asymmetric sparse difference penalty. MM procedure in Eq. (4) is used to derive the iterative algorithm for optimization.

A. SYMMETRIC PENALTY SPARSE DIFFERENCE ALGORITHM

According to Sec. III, it is known that ϕ is a symmetric penalty function. A majorizer $G(x, v)$ is firstly found for $\phi(x)$, i.e.,

$$g(v, v) = \phi(v), \tag{26}$$
$$g(x, v) \geq \phi(x), \tag{27}$$

Since $\phi(x)$ is a symmetric function, Eq. (27) can be regarded as an even second-order polynomial, i.e.,

$$g(x, v) = mx^2 + b. \tag{28}$$

According to Eqs. (26) and (27), it could be obtained that

$$mv^2 + b = \phi_(v), \quad 2mv = \phi'(v). \tag{29}$$

After that, the m and b can be computed as

$$m = \frac{\phi'(v)}{2v}, \quad b = \phi(v) - \frac{v}{2}\phi'(v). \tag{30}$$

Substituting Eq. (30) in Eq. (28) achieves

$$g(x, v) = \frac{\phi'(v)}{2v}x^2 + \phi(v) - \frac{v}{2}\phi'(v). \tag{31}$$

Then, it could be derived that

$$\begin{aligned}\sum_{n=0}^{N-1} g(x_n, v_n) &= \sum_n [\frac{\phi'(v)}{2v}x_n^2 + \phi(v_n) - \frac{v_n}{2}\phi'(v_n)] \\ &= \frac{1}{2}\mathbf{x}^T\frac{\phi'(v_n)}{v_n}\mathbf{x} + \sum_n[\phi(v_n) - \frac{v_n}{2}\phi'(v_n)] \\ &= \frac{1}{2}\mathbf{x}^T[\Gamma(\mathbf{v})]\mathbf{x} + c(\mathbf{v}) \\ &\geq \sum_{n=0}^{N-1}\phi(x_n),\end{aligned} \tag{32}$$

where

$$[\Gamma(v)]_{n,n} = \frac{\phi'(v_n)}{v_n}, \tag{33}$$

and

$$c(v) = \sum_n[\phi(v_n) - \frac{v_n}{2}\phi'(v_n)]. \tag{34}$$

Based on Eq. (32), it can be further obtained that

$$\begin{aligned}&\sum_{i=0}^{K}\lambda_i\sum_{n=0}^{N-1} g([\mathbf{D}_i\mathbf{x}]_n, [\mathbf{D}_i\mathbf{v}]_n) \\ &= \sum_{i=0}^{K}[\frac{\lambda_i}{2}(\mathbf{D}_i\mathbf{x})^T\frac{\phi'([\mathbf{D}_i\mathbf{v}]_n)}{[\mathbf{D}_i\mathbf{v}]_n}\mathbf{D}_i\mathbf{x} \\ &\quad + \sum_n[\phi([\mathbf{D}_i\mathbf{v}]_n) - \frac{(\mathbf{D}_i\mathbf{v})_n}{2}\phi'(\mathbf{D}_i\mathbf{v})_n] \\ &= \sum_{i=0}^{K}[\frac{\lambda_i}{2}(\mathbf{D}_i\mathbf{x})^T[\Lambda(\mathbf{D}_i\mathbf{v})](\mathbf{D}_i\mathbf{x}) + c_i(\mathbf{v})] \\ &\geq \sum_{i=0}^{K}\lambda_i\sum_{n=0}^{N_i-1}\phi([\mathbf{D}_i\mathbf{x}]_n),\end{aligned} \tag{35}$$

where

$$[\Lambda(\mathbf{D}_i)]_{n,n} = \frac{\phi'([\mathbf{D}_i\mathbf{v}]_n)}{[\mathbf{D}_i\mathbf{v}]_n}, \tag{36}$$

and

$$c_i(\mathbf{v}) = \sum_n[\phi([\mathbf{D}_i\mathbf{v}]_n) - \frac{(\mathbf{D}_i\mathbf{v})_n}{2}\phi'(\mathbf{D}_i\mathbf{v})_n]. \tag{37}$$

Base on Eq. (35), the majorizer of $F(\mathbf{x})$ based on MM is given by

$$\begin{aligned}G(\mathbf{x}, \mathbf{v}) = &\frac{1}{2}\|\mathbf{H}(\mathbf{y} - \mathbf{x})\|_2^2 \\ &+ \sum_{i=0}^{K}[\frac{\lambda_i}{2}(\mathbf{D}_i\mathbf{x})^T[\Lambda(\mathbf{D}_i\mathbf{v})](\mathbf{D}_i\mathbf{x})] + c(\mathbf{v}).\end{aligned} \tag{38}$$

Minimizing $G(\mathbf{x},\mathbf{v})$ with respect to $\mathbf{x}$ yields

$$\mathbf{x} = (\mathbf{H}^T\mathbf{H} + \sum_{i=0}^{K}\lambda_i\mathbf{D}_i^T[\Lambda(\mathbf{D}_i\mathbf{v})]\mathbf{D}_i)^{-1}\mathbf{H}^T\mathbf{H}\mathbf{y}. \tag{39}$$

Combining Eqs. (8) and (39), $\mathbf{x}$ can be written as

$$\begin{aligned}\mathbf{x} &= \mathbf{A}(\mathbf{B}^T\mathbf{B} + \mathbf{A}^T(\sum_{i=0}^{K}\lambda_i\mathbf{D}_i^T[\Lambda(\mathbf{D}_i\mathbf{v})\mathbf{D}_i)\mathbf{A})^{-1}\mathbf{B}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{y} \\ &= \mathbf{A}\mathbf{E}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{y},\end{aligned} \tag{40}$$

where

$$\mathbf{E} = \mathbf{B}^T\mathbf{B}\mathbf{A} + \mathbf{A}^T(\sum_{i=0}^{K}\lambda_i\mathbf{D}_i^T[\Lambda(\mathbf{D}_i\mathbf{v})]\mathbf{D}_i)\mathbf{A}. \tag{41}$$

The complete algorithm is proposed in Algorithm 1.

Algorithm 1 Symmetric Penalty Sparse Difference Algorithm

1: Input: y, A, B, λ_i, $i = 0, \ldots, K$
2: $\mathbf{b} = \mathbf{B}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{y}$
3: $\mathbf{x} = \mathbf{y}$ (initialization)
Repeat
4: $[\Lambda_i]_{n,n} = \frac{\phi'([\mathbf{D}_i\mathbf{x}]_n)}{[\mathbf{D}_i\mathbf{x}]_n}$, $i = 0, \ldots, K$
5: $\mathbf{I} = \sum_{i=0}^{K}\lambda_i\mathbf{D}_i\Lambda_i\mathbf{D}_i$
6: $\mathbf{E} = \mathbf{B}^T\mathbf{B} + \mathbf{A}^T\mathbf{I}\mathbf{A}$
7: $\mathbf{x}^{(k+1)} = \mathbf{A}[\mathbf{E}^{(k)}]^{-1}\mathbf{b}$
If the convergence conditions are met, output $\mathbf{x}$, otherwise, $k = k + 1$, and go to step 4-7.
8: $\mathbf{H} = \mathbf{B}\mathbf{A}^{-1}$
$\mathbf{f} = \mathbf{y} - \mathbf{x} - \mathbf{H}(\mathbf{y} - \mathbf{x})$
Output: $\mathbf{x}$, $\mathbf{f}$

B. COMPOUND PENALTY SPARSE DIFFERENCE ALGORITHM

For the consideration of the positive of the ECG signal peaks, asymmetric penalty is used for the original signal, and symmetric penalty is used for the difference signal. In this way, not only the detail part of ECG signal could be better represented, but also the sparsity of ECG signal could be enhanced and more real ECG signal could be restored. Eq. (22) can be further written as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}}\{F(\mathbf{x}) = \frac{1}{2}\|\mathbf{H}(\mathbf{y}-\mathbf{x})\|_2^2 + \lambda_0 \sum_{n=0}^{N-1} \theta_\rho(x_n; r) + \sum_{i=1}^{K} \lambda_i \sum_{n=0}^{N_i-1} \phi([\mathbf{D}_i\mathbf{x}]_n)\}, \quad (42)$$

where $\phi = \phi_B$ for the optimal solution $\mathbf{x}$. To construct the intermediate function $\mathbf{x}$ in Eq. (25), a quadratic equation is introduced as

$$g(x, v) = ax^2 + bx + c. \quad (43)$$

According to the MM algorithm

$$g(x, v) = \theta(x, r), \quad g'(v, v) = \theta'(v; r), \quad (44)$$
$$g(s, v) = \theta(s, r), \quad g'(s, v) = \theta'(s; r). \quad (45)$$

Note that a, b, c and s are all functions of v. Solving for them gives

$$a = \frac{1+r}{4|v|}, \quad b = \frac{1-r}{2}, \; c = \frac{(1+r)|v|}{4}, \; s = -v. \quad (46)$$

Substituting Eq. (46) in Eq. (43), the majorizer for $\theta(x, r)$ can be obtained as

$$g(x, v) = \frac{1+r}{4|v|}x^2 + \frac{1-r}{2}x + \frac{(1+r)|v|}{4}. \quad (47)$$

Furthermore, to address 'devide-by-zero' numerical issue when v approaches zero in Eq. (47), $\theta_\rho(x; r)$ is defined to be the second order polynomial with $v = \rho$ in a neighborhood of $x = 0$, i.e.,

$$\theta_\rho(x; r) = \begin{cases} x, & x > \rho \\ \frac{1+r}{4\rho}x^2 + \frac{1-r}{2}x + \frac{(1+r)\rho}{4}, & |x| \le \rho \\ -rx. & x < -\rho \end{cases} \quad (48)$$

Assume that $g(x, v)$ is the majorizer of the asymmetric and differentiable function $\theta_\rho(x; r)$, then define $f(x) = \frac{1+r}{4\rho}x^2 + \frac{1-r}{2}x + \frac{(1+r)\rho}{4}$, $\|x\| \le \rho$. For $\theta_\rho(x; r)$ to be a majorizer of $g(x, v)$, it can be shown as

$$g(x, v) = \begin{cases} \frac{1+r}{4v}x^2 + \frac{1-r}{2}x + \frac{(1+r)v}{4} & \ge f(x), \\ & v > \rho \\ -\frac{1+r}{4v}x^2 + \frac{1-r}{2}x - \frac{(1+r)v}{4} & \ge f(x). \\ & v < -\rho \end{cases} \quad (49)$$

Then, it could be derived that

$$g(x, v) - f(x) = (\frac{1+r}{4v} - \frac{1+r}{4\rho})x^2 + (v-\rho)\frac{1+r}{4} = \frac{(1+r)(v-\rho)(v\rho - x^2)}{4v\rho} > 0 \quad (50)$$

when $v > \rho$, and

$$g(x, v) - f(x) = (-\frac{1+r}{4v} - \frac{1+r}{4\rho})x^2 + (v+\rho)\frac{1+r}{4} = -\frac{(1+r)(v+\rho)(v\rho + x^2)}{4v\rho} > 0 \quad (51)$$

when $v < -\rho$. Therefore, in the domain $[-\rho, \rho]$, $\theta_\rho(x; r)$ itself is used as majorizer and is found to be

$$g(x, v) = \begin{cases} \frac{1+r}{4|v|}x^2 + \frac{1-r}{2}x + \frac{(1+r)|v|}{4}, & |v| > \rho \\ \frac{1+r}{4\rho}x^2 + \frac{1-r}{2}x + \frac{(1+r)\rho}{4}. & |v| \le \rho \end{cases} \quad (52)$$

It can be obtained that

$$\sum_{n=0}^{N-1} g(x_n, v_n) = \mathbf{x}^T[\Gamma(\mathbf{v})]\mathbf{x} + \mathbf{b}^T\mathbf{x} + c(\mathbf{v}) \ge \sum_{n=0}^{N-1} \theta_\rho(x_n; r), \quad (53)$$

where $\Gamma(\mathbf{v})$ is a diagonal matrix and expressed as

$$[\Gamma(\mathbf{v})]_{n,n} = \begin{cases} \frac{(1+r)}{4|v_n|}, & |v_n| \ge \rho \\ \frac{(1+r)}{4\rho}, & |v_n| \le \rho \end{cases} \quad (54)$$

and

$$\mathbf{b} = \frac{1-r}{2}. \quad (55)$$

Combining Eq. (35) with Eq. (53), the majorizer for $F(\mathbf{x})$ is given by

$$G(\mathbf{x}, \mathbf{v}) = \frac{1}{2}\|\mathbf{H}(\mathbf{y}-\mathbf{x})\|_2^2 + \lambda_0\mathbf{x}^T[\Gamma(\mathbf{v})]\mathbf{x} + \lambda_0\mathbf{b}^T\mathbf{x} + \sum_{i=1}^{k}[\frac{\lambda_i}{2}(\mathbf{D}_i\mathbf{x})^T] + c(\mathbf{v}). \quad (56)$$

Minimizing $G(\mathbf{x}, \mathbf{v})$ with respect to x yields

$$\mathbf{x} = \mathbf{H}^T\mathbf{H} + 2\lambda_0\Gamma(\mathbf{v}) + \sum_{i=1}^{K}\lambda_i(\mathbf{D}_i)^T[\Lambda(\mathbf{D}_i\mathbf{v})\mathbf{D}_i]^{-1}(\mathbf{H}^T\mathbf{H}\mathbf{y} - \lambda_0\mathbf{b}). \quad (57)$$

$\mathbf{H}$ has the expression that $\mathbf{H} = \mathbf{B}^{-1}\mathbf{A}$, hence, Eq. (57) could be written as

$$\mathbf{x} = \mathbf{A}(\mathbf{B}^T\mathbf{B} + \mathbf{A}^T\mathbf{K}\mathbf{A})^{-1}(\mathbf{B}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{y} - \lambda_0\mathbf{A}^T\mathbf{b}) = \mathbf{A}\mathbf{P}^{-1}(\mathbf{B}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{y} - \lambda_0\mathbf{A}^T\mathbf{b}), \quad (58)$$

Algorithm 2 Compound Penalty Sparse Difference Algorithm

Input: y, $r \geq 0$, A, B, λ_i, $i = 0, \ldots, K$
$\mathbf{b} = \frac{1-r}{2}$
$\mathbf{z} = \mathbf{B}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{y} - \lambda_0\mathbf{A}^T\mathbf{b}$
$\mathbf{x} = \mathbf{y}$ (initialization)
Repeat
$[\Lambda_i]_{n,n} = \frac{\phi'([\mathbf{D}_i\mathbf{x}]_n)}{[\mathbf{D}_i\mathbf{x}]_n}$, $i = 0, \ldots, K$
$\mathbf{P} = \mathbf{B}^T\mathbf{B} + \mathbf{A}^T\mathbf{K}\mathbf{A}$
$\mathbf{K} = 2\lambda_0\Gamma + \sum_{i=1}^{K} \lambda_i(\mathbf{D}_i)^T\Lambda_i\mathbf{D}_i$
$\mathbf{x} = \mathbf{A}\mathbf{P}^{-1}\mathbf{z}$
Until convergence
$\mathbf{H} = \mathbf{B}\mathbf{A}^{-1}$
$\mathbf{f} = \mathbf{y} - \mathbf{x} - \mathbf{H}(\mathbf{y} - \mathbf{x})$
Output: $\mathbf{x}$, $\mathbf{f}$

where

$$\mathbf{P} = \mathbf{B}^T\mathbf{B} + \mathbf{A}^T\mathbf{K}\mathbf{A}, \tag{59}$$

and

$$\mathbf{K} = 2\lambda_0\Gamma(\mathbf{v}) + \sum_{i=1}^{K} \lambda_i(\mathbf{D}_i)^T[\Lambda(\mathbf{D}_i\mathbf{v})]\mathbf{D}_i. \tag{60}$$

$\mathbf{P}$ and $\mathbf{K}$ are the banded matrices. The MM iteration can be implemented using the fast solver for the banded systems, shown as

$$\mathbf{K} = 2\lambda_0\Gamma(\mathbf{x}^{(k)}) + \sum_{i=1}^{K} \lambda_i(\mathbf{D}_i)^T[\Lambda(\mathbf{D}_i\mathbf{x}^{(k)})]\mathbf{D}_i, \tag{61}$$

$$\mathbf{P}^{(k)} = \mathbf{B}^T\mathbf{B} + \mathbf{A}^T\mathbf{K}^{(k)}\mathbf{A}, \tag{62}$$

$$\mathbf{x}^{(k+1)} = \mathbf{A}\mathbf{P}^{-1}(\mathbf{B}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{y} - \lambda_0\mathbf{A}^T\mathbf{b}). \tag{63}$$

Furthermore, the matrix-vector multiplications are also computationally efficient for the reason that all matrices are banded.

Finally, the update equations constitute the Algorithm 2, which in this paper is listed as *Compound Penalty Sparse Difference Algorithm*.

V. SIMULATION RESULTS

The experimental environment is Windows 10 operating system, MATLAB R2014A, and the machine is configured with Intel(R) Core i7-6700 CPU 3.40 GHz processor and 8 GB RAM.

A. ECG DATABASE

The ECG signals used for the test are taken from MIT-BIH Arrhythmia Database [37], while the BW signals are taken from MIT-BIH Noise Stress Test Database [38], and all the BW signals are obtained from the real physical active of volunteers. Set 0 dB, 1.25 dB and 5 dB signal noise radio (SNR) for the system, respectively. The proposed method was compared with the wavelet and median filter methods for ECG signals. Also, the white Gaussian noise with variance σ^2 is added on the ECG signals.

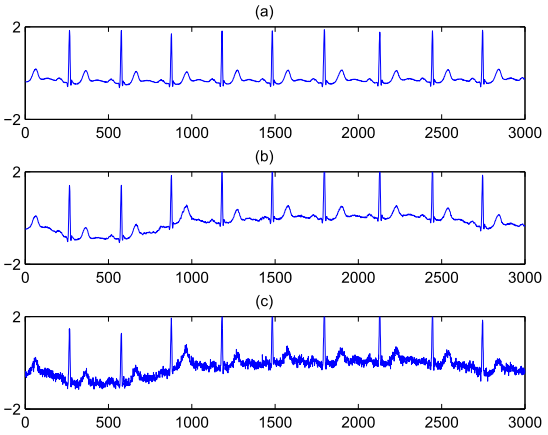


FIGURE 5. The example of ECG signal. (a) Original ECG signal. (b) Baseline wander ECG signal. (c) Baseline wander and noise ECG signal.

B. PERFORMANCE INDEX

The performance of proposed method is evaluated based on the calculated SNR and mean square error (MSE). The SNR and MSE can be represented as follows,

$$\text{SNR} = 10\log\frac{\sum_{n=0}^{N-1}[x(n)]^2}{\sum_{n=0}^{N-1}[x(n) - \hat{x}(n)]^2}, \tag{64}$$

$$\text{MSE} = \frac{1}{N}\sum_{n=0}^{N-1}[x(n) - \hat{x}(n)]^2. \tag{65}$$

It is well known that the denoising effect performs better at high SNR and low MSE.

C. BASELINE CORRECTION OF REAL ECG SIGNAL

The example in this subsection illustrates the use of the correction of BW in real ECG signal. Fig. 5 shows the example of ECG signal. Fig. 5 (a) represents the raw ECG signal (MIT-BIH record No.103), Fig. 5 (b) represents the ECG signal with BW, while Fig. 5 (c) is the ECG signal with both BW and white Gaussian noise. To clarify the use of the propose algorithm, set $d = 1$ with a cut-off frequency of $f_c = 0.009$. The ECG signal is modeled as a signal with three-order sparse derivatives, and the reason will be given in the last subsection.

Fig. 6 illustrates the effect of BW elimination with different methods, the blue line is original clean ECG signal, while the line of ECG signal with BW correction using the proposed algorithm, the wavelet and median filter are red, green and black, respectively. It is clearly shown that after the processing of the proposed method, the ECG signal successfully removes the BW and gets its normal steady curve and original characteristics. Compared with other two algorithms, the proposed method has the better performance on the BW correction, and the restored signal is basically consistent with the original ECG signal. Moreover, the waveform using the proposed method is more smooth, and the low-frequency noise in the BW could be efficiently removed.

Table 1 illustrates the comparison of the SNR and MSE of wavelet, median filter, and the proposed method for BW

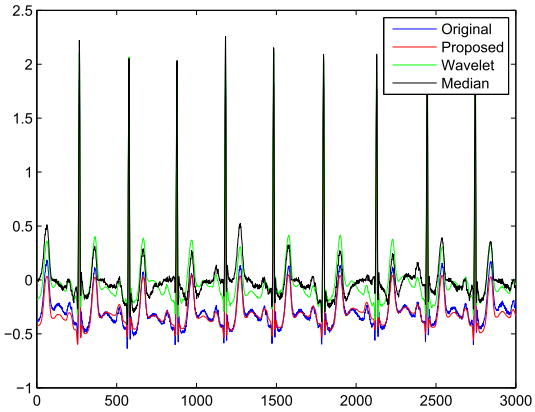


FIGURE 6. Compare with other methods of ECG baseline correction.

TABLE 1. Experimental results for BW correction with different SNRs and MSEs.

	SNR=0 dB		SNR=1.25 dB		SNR=5 dB	
	SNR/dB	MSE	SNR/dB	MSE	SNR/dB	MSE
Proposed method	12.53	0.009	15.88	0.006	17.87	0.003
Wavelet	3.90	0.085	4.02	0.083	4.08	0.082
Median	3.89	0.208	4.45	0.184	5.14	0.084

correction. The results show that after BW noise signals are added to the ECG signal with SNRs of 0 dB, 1.25 dB, and 5 dB, respectively, the proposed method provides superior performance than wavelet and median filter with different SNRs and MSEs. For example, the experiment results for MIT-BIH record no.103 show that for SNR of 1.25 dB, the output SNRs of wavelet and median filter are 4.02 dB and 4.45 dB, while the proposed method gives an output SNR of 15.88 dB. At the same time, the MSE of proposed method is 0.006, which is lower than other methods.

D. DENOISING OF REAL ECG SIGNAL

To show the denoising effect, the proposed scheme is compared with the wavelet algorithm which has the best denoising effect among the traditional algorithms. The real ECG signal with additive white Gaussian noise is used for comparison.

The denoising performance of wavelet is shown in Fig. 7. It is clearly shown that the wavelet algorithm achieves great denoising performance. However, it could be found from Fig. 7 (c) and Fig. 7 (d) that the denoised signal has slight wave oscillation, and obviously underestimates the amplitudes of ECG peaks. It is noticed that the SNR of the denoised signal is 10.33 dB, and MSE is 0.019.

Fig. 8 shows the denoising performance of the proposed algorithm. It is shown in Fig. 8 (c) and (d) that compared with wavelet, the proposed method could effectively reduce the noise and avoid the wave oscillation and overestimation of the amplitudes of ECG peaks. The SNR is 16.46 dB and MSE is 0.004, which could estimate the amplitudes of ECG peaks more accurately.

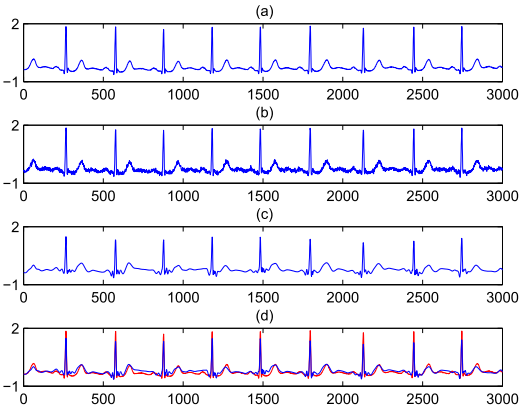


FIGURE 7. Denoised ECG signal using wavelet algorithm (Red line is original ECG signal, and blue line is denoised ECG signal in Fig. 7 (d).) (a) Original ECG signal. (b) Noise ECG signal. (c) Denoised ECG signal. (d) Compare with original ECG signal.

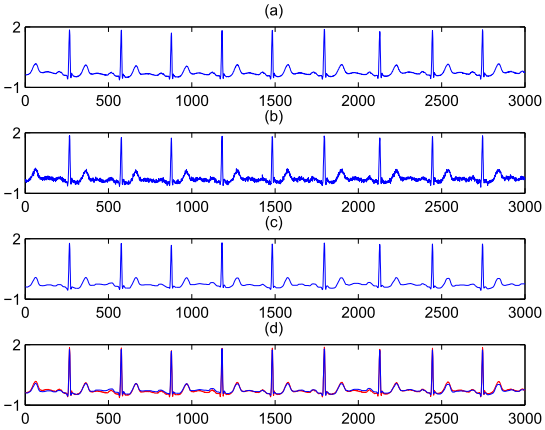


FIGURE 8. Denoised ECG signal using the proposed algorithm (Red line is original ECG signal, and blue line is denoised ECG signal in Fig. 8 (d).) (a) Original ECG signal. (b) Noise ECG signal. (c) Denoised ECG signal. (d) Compare with original ECG signal.

E. JOINT BW CORRECTION AND DENOISING OF REAL ECG SIGNAL

In this subsection, the example is applied to illustrate the performance of the proposed algorithm for joint BW correction and denoising.

The simulation results are shown in Fig. 9. Set the parameters $r = 1$ and $f_c = 0.009$. As shown in Fig. 9 (b), the baseline is well estimated. Fig. 9 (c) illustrates that the proposed algorithm achieves great performance for BW correction and denoising, and Fig. 9 (d) shows the estimation of residual constitute, i.e., the noise of the ECG signal.

The median filter and wavelet are used for comparison with the proposed algorithm, which is shown in Figs. 10 - 12. The same degree of BW is added to explore the different denoising performance in different SNRs. It can be observed that the proposed algorithm could effectively isolate and preserve the P-waves, QRS-complex and T-waves in the denoised ECG signal. The wavelet algorithm has a good denoising performance, but it cannot well correct the BW. The median filter could effectively eliminate the BW, however, the effect

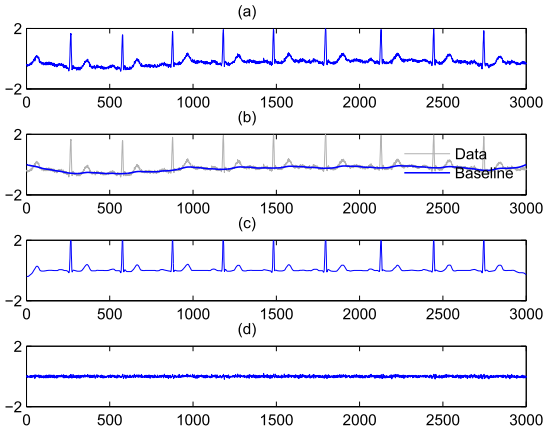


FIGURE 9. Joint baseline wander correction and denoising. (a) The ECG signal added BW signal and noise. (b) Baseline, estimation (r = 1, fc = 0.009, d = 1). (c) The ECG signal of BW correction and denoising. (d) Residual.

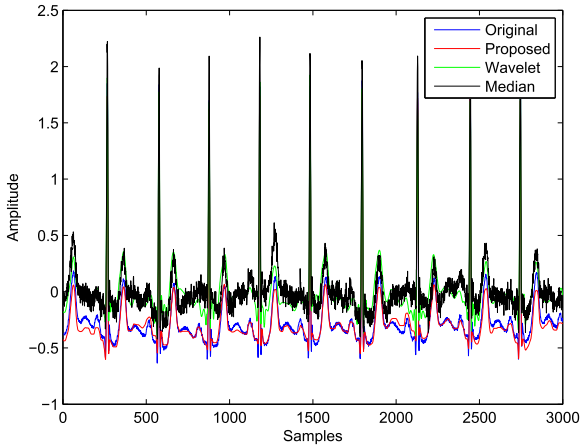


FIGURE 10. ECG signal added noise with SNR = 5 dB.

on signal denoising is much poor. Compared with other two algorithms, the proposed algorithm achieves a better performance of BW correction and denoising.

Furthermore, in order to prove the universality and validity of the proposed method, another two samples (MIH-BIH no.105, MIT-BIH no. 213) with morphological variations are used for joint BW correction and denoising. The simulations are shown in Fig. 13 and Fig. 14. The performance comparison between SNR and MSE of different methods for the three signals is shown in Table 2. It can be seen that the proposed method achieves good performance under different morphological variations of ECG signals. The wavelet and median methods could only provide minor improvement in the SNR and MSE.

F. DIFFERENCE ORDER

According to Eq. (42), the estimation results of ECG signal $\mathbf{x}$ are related to the value of difference order λ. Considering that the regularization parameter λ is critical for the reconstruction of ECG signal, the proper difference order for ECG signal is

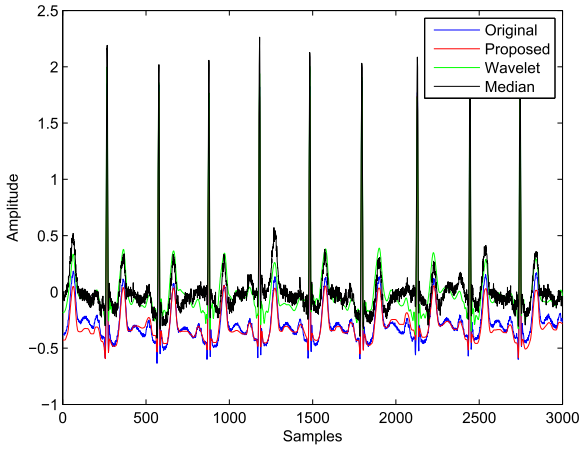


FIGURE 11. ECG signal added noise with SNR = 10 dB.

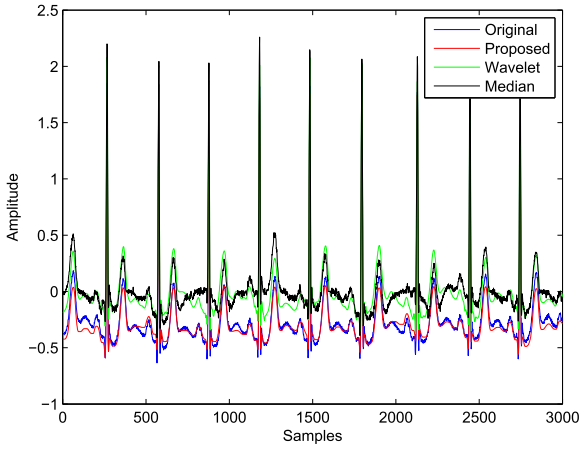


FIGURE 12. ECG signal added noise with SNR = 15 dB.

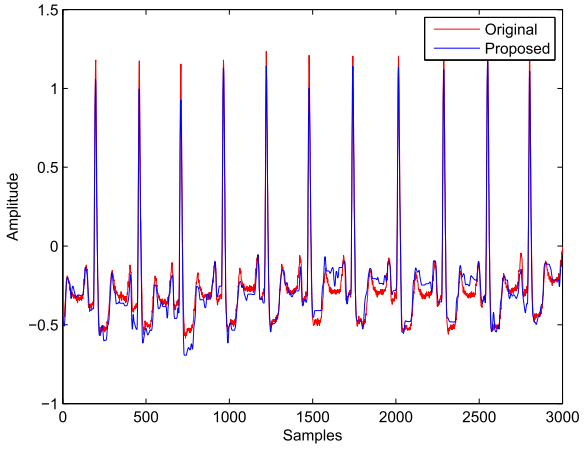


FIGURE 13. MIT-BIH no. 105 joint BW correction and denoising.

discussed in this subsection. The example is considered under the situation that $f_c = 0.009$, $d = 1$ and $r = 1$. A mix ECG signal is shown in Fig. 15, which is a real ECG signal added the BW signal and white Gaussian noise. Figs. 16 - 19 provide the reconstructions of ECG signal under different orders.

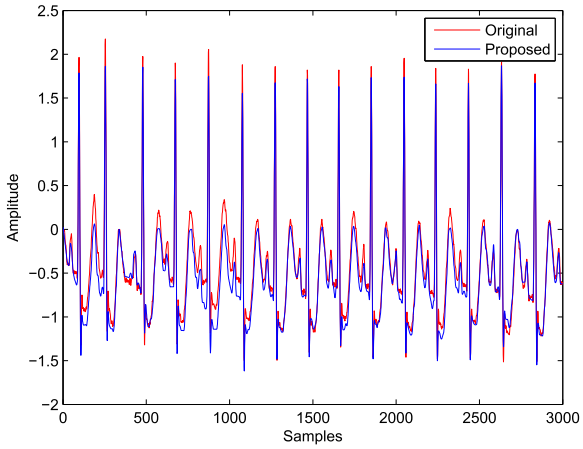


FIGURE 14. MIT-BIH no. 213 joint BW correction and denoising.

TABLE 2. Experiment results of output SNR and MSE in different input SNRs.

MIT-BIH no.			103	105	213
Proposed method	5dB	SNR	15.94	15.84	13.36
		MSE	0.003	0.004	0.022
	10dB	SNR	16.56	17.65	13.69
		MSE	0.002	0.003	0.022
	15dB	SNR	16.82	17.89	14.68
		MSE	0.002	0.002	0.023
Wavelet	5dB	SNR	4.85	10.38	7.49
		MSE	0.086	0.018	0.127
	10dB	SNR	4.98	13.39	8.12
		MSE	0.084	0.008	0.110
	15dB	SNR	5.14	14.91	8.22
		MSE	0.082	0.007	0.107
Median	5dB	SNR	3.85	5.11	3.63
		MSE	0.159	0.158	0.563
	10dB	SNR	3.97	5.37	3.68
		MSE	0.158	0.152	0.561
	15dB	SNR	4.06	5.61	3.84
		MSE	0.158	0.150	0.546

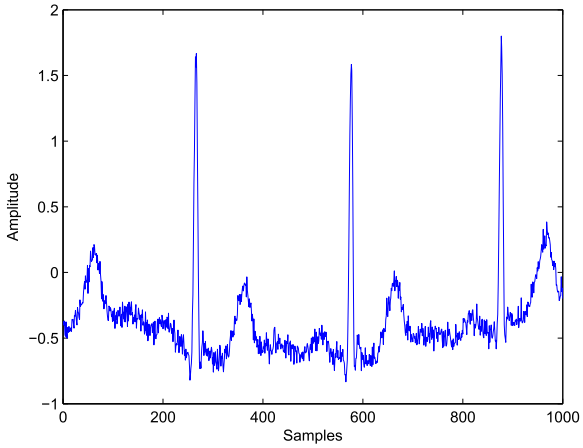


FIGURE 15. Mixed ECG signal.

As shown in Fig. 16, when 1-order difference is chosen, the BW in the mixed ECG signal is corrected and the noise is reduced effectively. However, it could be observed that the reconstructed ECG signal is not smooth enough and has many stair-case artifacts, which cannot retain the detailed features

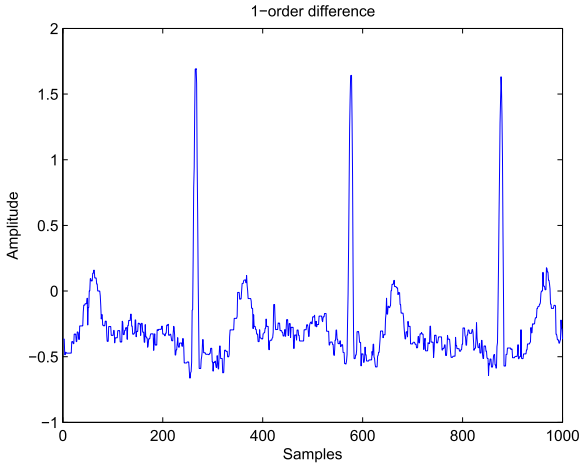


FIGURE 16. Reconstructed ECG signal with 1-order difference.

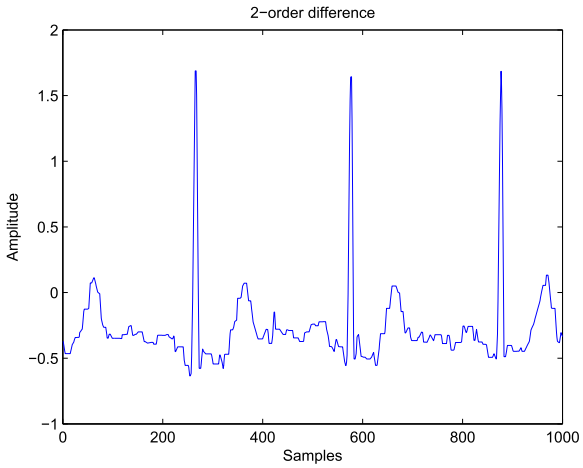


FIGURE 17. Reconstructed ECG signal with 2-order difference.

of the original signal. The SNR of the reconstructed ECG signal is 14.49 dB, and the regularization parameter λ is set as follows: $\lambda_0 = 0.6$, $\lambda_1 = 7$.

As shown in Fig. 17, when 2-order difference is chosen, the noise is further reduced. There are still stair-case artifacts and sawtooth step waveforms, but the number of those is significantly reduced compared with the 1-order difference case, and the original signal is roughly restored with the preservation of more details. The SNR of the reconstructed ECG signal is 15.21 dB, and the regularization parameter λ is set as follows: $\lambda_0 = 0.6$, $\lambda_1 = 7$, $\lambda_2 = 7$.

As shown in Fig. 18, when 3-order difference is chosen, the noise is continuously reduced. The waveform of the reconstructed signal becomes smoother, and it is hard to visibly find the existence of stair-case artifacts and serrated stepped waveforms. The waveform of reconstructed signal is closer to that of the original signal, and the details is more accurate. The SNR of the reconstructed ECG signal is 16.78 dB, and the regularization parameter λ is set as follows: $\lambda_0 = 0.6$, $\lambda_1 = 7$, $\lambda_2 = 7$, $\lambda_3 = 20$.

As shown in Fig. 19, when 4-order difference is chosen, there is no significant change in the reconstructed ECG signal

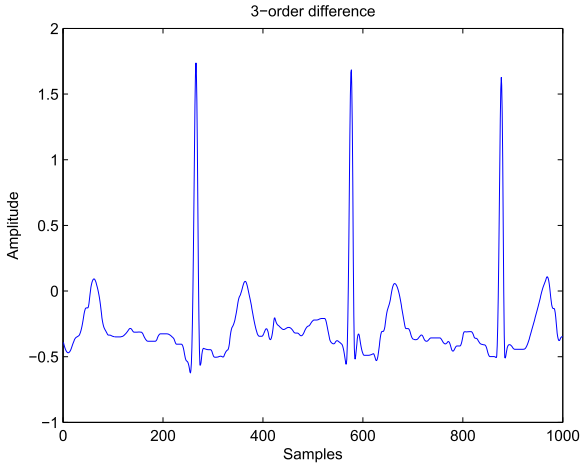


FIGURE 18. Reconstructed ECG signal with 3-order difference.

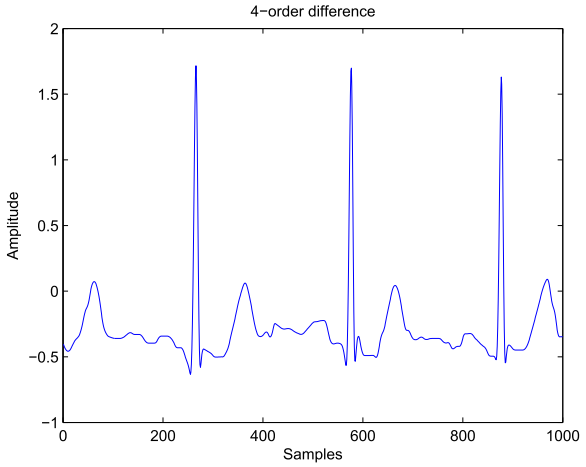


FIGURE 19. Reconstructed ECG signal with 4-order difference.

compared to the 3-order difference case. Moreover, the SNR of the reconstructed ECG signal is 16.57 dB, which is lower than that of the 3-order difference case, showing that the performance is not as good as 3-order difference case. As a result, the difference with 4-order or more is not recommended and 3-order is most proper for the example.

VI. CONCLUSION

In this paper, a convex optimization method for BW correction and denoising of ECG signals is proposed, combining LTI filtering with sparse denoising. Wavelet and median filter are applied for comparison. Simulation results show that the proposed algorithm has a better performance for BW correction of ECG signal. For the aspect of denoising, the proposed method achieves a good noise reduction effect and effectively solves the problem of low peak estimation of ECG signal existing in the traditional method. In addition, the simulation experiments of ECG signal combined with joint BW correction and denoising are also carried out. The results show that the proposed method could effectively reduce the noise and correct the BW in the ECG signal. The waveform of ECG signal is smooth and more details are preserved. Moreover, the advantages of the proposed method is shown by means of SNR and MSE. Finally, the difference order of ECG signal is discussed, and the conclusion that the 3-order difference is suitable under the situation of the setting parameter is obtained. It could be noticed that there are some limitations with proposed method, i.e., parameter could only be selected based on experience at present. The adaptive parameters for any input ECG signals will be considered to further improve the computational efficiency in future work.

REFERENCES

[1] J. A. van Alsté, W. Van Eck, and O. E. Herrmann, "ECG baseline wander reduction using linear phase filters," *Comput. Biomed. Res.*, vol. 19, no. 5, pp. 417–427, 1986.
[2] J. M. Łęski and N. Henzel, "ECG baseline wander and powerline interference reduction using nonlinear filter bank," *Signal Process.*, vol. 85, no. 4, pp. 781–793, Apr. 2005.
[3] 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. 2005.
[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, Jul. 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, no. 11, pp. 178–187, Mar. 2016.
[6] N. U. 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. 365–371, 2014.
[8] H. Sharma and K. K. Sharma, "Baseline wander removal of ECG signals using Hilbert vibration decomposition," *Electron. Lett.*, vol. 51, no. 6, pp. 447–449, Mar. 2015.
[9] H. Sharma and K. K. Sharma, "ECG-derived respiration based on iterated Hilbert transform and Hilbert vibration decomposition," *Australas. Phys. Eng. Sci. Med.*, vol. 41, no. 2, pp. 429–443, Apr. 2018.
[10] M. A. Awal, S. S. Mostafa, M. Ahmad, and M. A. Rashid, "An adaptive level dependent wavelet thresholding for ECG denoising," *Biocybernetics Biomed. Eng.*, vol. 34, no. 4, pp. 238–249, 2014.
[11] W. Jenkal *et al.*, "An efficient algorithm of ECG signal denoising using the adaptive dual threshold filter and the discrete wavelet transform," *Biocybernetics Biomed. Eng.*, vol. 36, no. 3, pp. 499–508, 2016.
[12] 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.
[13] Z. A. A. Alyasseri, A. T. Khader, M. A. Al-Betar, and M. A. Awadallah, "Hybridizing β-hill climbing with wavelet transform for denoising ECG signals," *Inf. Sci.*, vol. 429, pp. 229–246, Mar. 2018.
[14] 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.
[15] 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.
[16] 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.
[17] K. Bredies, K. Kunisch, and T. Pock, "Total generalized variation," *SIAM J. Imag. Sci*, vol. 3, no. 3, pp. 492–526, 2010.
[18] 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, Feb. 2001.
[19] 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, 1992.
[20] 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.

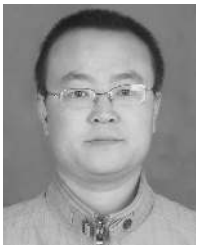
[21] X. Ning, I. W. Selesnick, and L. Duval, "Chromatogram baseline estimation and denoising using sparsity (BEADS)," *Chemometrics Intell. Lab. Syst.*, vol. 139, pp. 156–167, Dec. 2014.
[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[23] I. Selesnick, "Sparsity-assisted signal smoothing (revisited)," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2017, pp. 4546–4550.
[24] I. Selesnick, "Sparse regularization via convex analysis," *IEEE Trans. Signal Process.*, vol. 65, no. 17, pp. 4481–4494, Sep. 2017.
[25] M. El Hanine, E. Abdelmounim, R. Haddadi, and A. Belaguid, "Electro-CardioGram signal denoising using discrete wavelet transform," *Comput. Technol. Appl.*, vol. 5, pp. 98–104, Sep. 2014.
[26] 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.
[27] M. A. T. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization–minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.
[28] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, 2004.
[29] S. H. Lee and M. G. Kang, "Total variation-based image noise reduction with generalized fidelity function," *IEEE Signal Process. Lett.*, vol. 14, no. 11, pp. 832–835, Nov. 2007.
[30] Y. Hu and M. Jacob, "Higher degree total variation (HDTV) regularization for image recovery," *IEEE Trans. Image Process.*, vol. 21, no. 5, pp. 2559–2571, May 2012.
[31] I. W. Selesnick and I. Bayram, "Sparse signal estimation by maximally sparse convex optimization," *IEEE Trans. Image Process.*, vol. 62, no. 5, pp. 1078–1092, Mar. 2013.
[32] I. Selesnick, "Total variation denoising via the Moreau envelope," *IEEE Signal Process. Lett.*, vol. 24, no. 2, pp. 216–220, Feb. 2017.
[33] T. W. Parks and C. S. Burrus, *Digital Filter Design*. Hoboken, NJ, USA: Wiley, 1987.
[34] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[35] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.
[36] P. H. C. Eilers, "Unimodal smoothing," *J. Chemometrics*, vol. 19, nos. 5–7, pp. 317–328, Jan. 2005.
[37] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH Arrhythmia Database," *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May 2001.
[38] G. B. Moody, W. E. Muldrow, and R. G. Mark, "A noise stress test for arrhythmia detectors," *Comput. Cardiol.*, vol. 11, pp. 381–384, 1984.

XIAO WANG received the B.E. degree in automation from the North China Institute of Science and Technology. He is currently pursuing the M.E. degree in 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.

YOU ZHOU received the B.E. degree in electronic and information engineering from Qufu Normal University, China, in 2010, and the Ph.D. degree in communication and information systems from Shandong University, China, in 2018. He did the research with Oregon State University, US, as a Visiting Scholar, from 2016 to 2017. He is currently 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 spatial modulation, multiple-input multiple-output wireless communication, and cooperative communication.

MINGLEI SHU received the B.S. degree in automation, the M.S. degree in power electronics, and the Ph.D. degree in communication and information systems from Shandong University, China, in 2003, 2006, and 2017, respectively. He is currently a Research Fellow with Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. He is also the Head of the Information Medicine Team of the Shandong Computer Science Center (National Supercomputer Center in Jinan), the Executive Director of the Sino-Australian Joint Laboratory of International Health Technology, the Vice President of the Medical and Health Branch of Shandong Internet of Things Association, the Vice President of the śInternet+ ś Alliance, the Executive Director of the Telemedicine and Information Technology Branch of China Medical Equipment Association, and an Executive Director of the śInternet + ś Medical Professional Committee of China Health Information and Big Date Association. His research interests include medical artificial intelligence, medical big data and the medical Internet of Things, wireless sensor networks, wireless body area networks, and information security.

YINGLONG WANG received the B.S. degree in electronic technology and the M.S. degree in industrial automation from the Shandong University of Technology, Jinan, China, in 1987 and 1990, respectively, and the Ph.D. degree in communication and information systems from Shandong University, Jinan, in 2005. He is currently a Research Fellow with the Shandong Computer Science Center (National Supercomputer Center in Jinan), Qilu University of Technology (Shandong Academy of Sciences), Jinan. He is also the Director of the Sino-Australian Joint Laboratory of International Health Technology, the Vice Chairman of the Shandong Science and Technology Association, the President of the Shandong Internet of Things Association, a Member of the Shandong Information Expert Group, a member of the Shandong Information Expert Consultation Committee, the Vice Chairman of the Shandong Computer Society, and the Vice Chairman of the Shandong Information Standardization Technology Committee. His current research interests include medical artificial intelligence, high-performance computing, wireless sensor networks, information security, and cloud computing.



ANMING DONG (S'11–M'17) received the B.E. degree in electronic information science and technology from Liaocheng University, Liaocheng, China, in 2004, the M.E. degree in communications and information systems from Lanzhou University, Lanzhou, China, in 2007, and the Ph.D. degree in communications and information systems from Shandong University, Jinan, China, in 2016. He is currently with the School of Information, Qilu University of Technology (Shandong Academy of Sciences), Jinan, China. Since 2018, he has been with the Shandong Computer Science Center (National Supercomputer Center in Jinan), as a Cooperative Researcher. He has been a Reviewer of a number of IEEE journals and conferences. His main research interests include transceiver design for multiple-input multiple-output wireless communications, advanced signal processing techniques, and interference management in multiuser multiple-input multiple-output systems. He was a recipient of the Excellent Doctoral Dissertation Awards of Shandong Province and Shandong University, in 2017.