

Article

Pruning-Based Sparse Recovery for Electrocardiogram Reconstruction from Compressed Measurements

Jaeseok Lee, Kyungsoo Kim and Ji-Woong Choi *

Department of Information & Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology (DGIST), Daegu 771-813, Korea; jayslee@dgist.ac.kr (J.L.); ssi09@dgist.ac.kr (K.K.)

* Correspondence: jwchoi@dgist.ac.kr; Tel.: +82-53-785-6311; Fax: +82-53-785-6309

Academic Editor: Panicos Kyriacou

Received: 24 October 2016; Accepted: 3 January 2017; Published: 7 January 2017

Abstract: Due to the necessity of the low-power implementation of newly-developed electrocardiogram (ECG) sensors, exact ECG data reconstruction from the compressed measurements has received much attention in recent years. Our interest lies in improving the compression ratio (CR), as well as the ECG reconstruction performance of the sparse signal recovery. To this end, we propose a sparse signal reconstruction method by pruning-based tree search, which attempts to choose the globally-optimal solution by minimizing the cost function. In order to achieve low complexity for the real-time implementation, we employ a novel pruning strategy to avoid exhaustive tree search. Through the restricted isometry property (RIP)-based analysis, we show that the exact recovery condition of our approach is more relaxed than any of the existing methods. Through the simulations, we demonstrate that the proposed approach outperforms the existing sparse recovery methods for ECG reconstruction.

Keywords: biomedical signal processing; electrocardiogram; compressed sensing; sparse signal recovery; tree pruning

1. Introduction

It is well known that electrocardiogram (ECG) sensors enable effective medical diagnosis for heart problems, such as arrhythmia and myocardial infarction, in everyday life [1–3]. In this regard, implanted ECG-based pacemakers and wearable ECG monitoring devices were developed to detect critical problems in the cardiovascular system [4]. Meanwhile, recently-developed electrocardiogram (ECG) sensors in everyday life require stable and long time capability for developing wearable devices in ambulatory environments [5,6]. Due to the growing demand of smart wearable devices, the major issue for recent ECG sensors is to achieve efficient management of large quantities of real-time biosignals in ambulatory environments [7]. As a means of ECG signal processing implemented with low power and small data storage, one of the promising paradigms that has received much attention recently is the compressed sensing (CS)-based signal compression and reconstruction [8–24]. The well-known finding of the CS-based data reconstruction is that signals can be recovered from far fewer measurements than traditional schemes whenever the signal is sparse (signals with a very small number of nonzero coefficients, that is, \mathbf{s} is a K -sparse signal if $\|\mathbf{s}\|_0 = K \ll \dim(\mathbf{s})$), which can be exactly reconstructed from underdetermined measurement $\mathbf{y} = \Phi\mathbf{s}$ and the sensing mechanism satisfies the restricted isometry property (RIP), which is given as follows.

Definition 1. The sensing matrix $\Phi \in \mathbb{R}^{M \times N}$ ($M \ll N$) is said to satisfy the RIP of order K if there exists a restricted isometry constant (RIC) $\delta_K \in (0, 1)$, such that:

$$\sqrt{1 - \delta_K} \|\mathbf{s}\|_2 \leq \|\Phi \mathbf{s}\|_2 \leq \sqrt{1 + \delta_K} \|\mathbf{s}\|_2$$

for any K -sparse vector $\mathbf{s} \in \mathbb{R}^N$.

Additional benefits of the CS-based ECG processing are (1) the computationally-efficient data compression and (2) the guarantee of exact reconstruction from far fewer measurements than conventional methods. While the computational efficiency of data compression can be easily demonstrated since it requires only simple linear matrix multiplication (see Definition 1), our main interest lies in the reduction in the number of required measurements ensuring exact sparse signal recovery from compressed ECG data. A popular method for identifying the sparsest signal \mathbf{s} from the measurements $\mathbf{y} = \Phi \mathbf{s}$ is to formulate the ℓ_0 -minimization problem as:

$$\min_{\mathbf{s}} \|\mathbf{s}\|_0 \quad \text{subject to } \mathbf{y} = \Phi \mathbf{s} \quad (1)$$

where $\Phi \in \mathbb{R}^{M \times N}$ ($M \ll N$) is often called the sensing matrix. Since Equation (1) is known to be NP-hard [9], ℓ_1 -relaxation methods, such as basis pursuit (BP) [9], BP denoising (BPDN) [10] (or Lasso [11]) and the Dantzig selector [12], were introduced. Other than ℓ_1 -relaxation methods, further reduction in complexity can be achieved by the greedy approaches. To be specific, greedy algorithms attempt to identify the support $T = \{j | 1 \leq j \leq N, s_j \neq 0\}$ (index set of nonzero entries of \mathbf{s}) in an iterative fashion, returning a sequence of estimates of the sparse input vector. However, although the greedy algorithm, such as orthogonal matching pursuit (OMP) [21], enables computationally-efficient implementation, its performance in general is not quite satisfactory, in particular in the presence of noise. This is mainly due to the fact that stepwise identification of the support elements might lead to a myopic decision in each iteration. Moreover, such a criterion does not provide any further chances to correct the mistake of selecting incorrect (i.e., j , such that $s_j = 0$) indices once selected [21–23] (see Figure 1a).

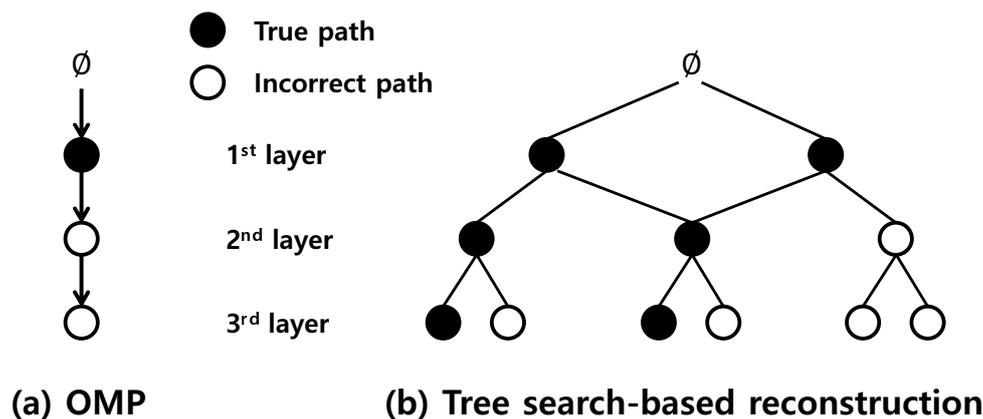


Figure 1. Illustration of conventional orthogonal matching pursuit (OMP) and recovery based on tree search where the true paths contain only elements in the support T , and incorrect paths consist of at least one element from T^c . While OMP does not provide any further chances after it selects an incorrect index in the second layer (a), multiple path investigation in (b) enables a reduction in the misdetection of the support element.

The aim of this work is to introduce a new sparse signal recovery scheme that overcomes such drawbacks of conventional methods and achieves effective ECG reconstruction. By employing the tree

search with an aggressive pruning strategy, our method achieves not only accurate ECG reconstruction, but also real-time implementation suitable for ambulatory environments. The main benefit of the tree search is that it enables multiple candidate investigations for identifying the support $T = \{j \mid 1 \leq j \leq N, s_j \neq 0\}$ (see Figure 1b). That is, since the tree search examines the reliability of multiple index sets simultaneously, it improves the reconstruction performance by reducing the misdetection, as well as the false alarm probabilities (the misdetection and the false alarm probabilities in this manuscript denote the probabilities of the support index not being selected and the incorrect index ($j \in T^c$) being identified as the support, respectively). In fact, many of the previous works in the literature focused on recovering sparse signals using the tree structure [16–20]. For instance, tree search-based orthogonal matching pursuit (TB-OMP) constructs the search tree by spreading multiple branches for each path [16], and its modified version was introduced in [17]. In fast Bayesian matching pursuit (FBMP), a fixed number of paths with the best posterior probabilities survives in each layer [18]. Further, the multipath matching pursuit (MMP) [19] attempts to select multiple branches ($L \geq 2$) by choosing maximally-correlated indices with the residual, and the combined method of A^* search [25,26] and orthogonal matching pursuit (OMP) [21] was introduced as a stage-wise residual minimization employing an effective pruning strategy [20].

Our approach, referred to as tree pruning-based matching pursuit (TPMP), provides further improvement by exploiting the full dictionary information with aggressive pruning strategies for each path. To be specific, the proposed TPMP considerably reduces the computational burden of the brute-force tree search, yet achieves excellent recovery performance by jointly implementing two pruning criteria, that is (1) the pre-scanning and (2) the pruning-based tree search. In the pre-scanning stage, we greedily choose a small number of promising column indices of the sensing matrix. If we denote the set of column indices obtained in the pre-scanning stage as Θ , then we set $K \leq |\Theta| \ll N$ where $|\Theta|$ is the cardinality of Θ . Once the pre-scanning is finished, the search tree is initialized by spreading the paths using only elements of Θ , so that the number of total possible paths in the tree is reduced from $\binom{N}{K}$ to $\binom{|\Theta|}{K}$, where $\binom{a}{b} = \frac{a!}{b!(a-b)!}$. For additional alleviation of the computational burden, TPMP employs a pruning strategy for removing unpromising paths from the tree. Similar to sphere decoding (SD) or list sphere decoding (LSD) with probabilistic pruning criteria [27–30], the pruning strategy is based on computing the cost function by greedily estimating the further path. Instead of obtaining the probabilistic characteristics of each path as previous works, our method exploits a full-blown candidate with cardinality K , which is constructed by combining the current path and greedily estimating further indices considering the complete dictionary information. By doing so, TPMP reduces the possibility of not selecting the support element (as well as selecting the incorrect index) subject to the constraint of sparsity level K . In addition, we demonstrate that this can rather reduce the running time complexity by shutting down the search in the early stage of the search tree.

While the preliminary version of this work was presented for an arbitrary system in [31], we show that the proposed method is highly suitable for ECG processing with some modifications and performs close to the best possible estimator (the estimator referred to as the oracle least squares (LS) estimator where the support information is given) [32]. To be specific, we reduced the cardinality of Θ for constructing smaller number of paths in the search tree for real-time implementation. In addition, we modified the cost function computation to maintain high reconstruction accuracy since investigating a smaller number of paths might degrade the performance. In order to achieve further reduction in complexity, we also employ a new stopping criterion with marginal performance loss by limiting the minimum pruning threshold. Moreover, compared to [31], we demonstrate that such modifications not only reduce the search complexity, but also improve the exact recovery condition (ERC) bound. From numerical simulations, we show that our proposed method outperforms the existing methods with practical complexity and provides additional flexibility for hardware implementation.

The rest of this manuscript is organized as follows. In Section 2, we briefly provide our setup for compressing and reconstructing ECG and then propose the TPMP algorithm. In Section 3, we analyze

the exact recovery condition under which TPMP identifies the support accurately. In Section 4, we provide the numerical performance of the proposed method and then conclude in Section 5.

2. Tree Search-Based ECG Reconstruction

In this section, we introduce a low-power ECG reconstruction method based on the tree search where the system model is provided in Figure 2. We first introduce an existing ECG compression procedure following the compressed sensing-based system architecture (Section 2.1) and then discuss our proposed method for reconstructing the ECG data from compressed measurements (Section 2.2).

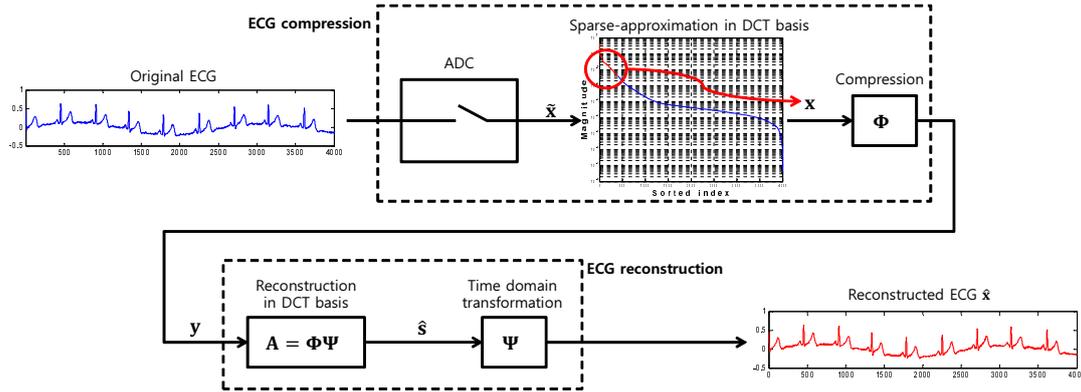


Figure 2. Basic structure of ECG compression and reconstruction. Note that the reconstruction is based on the discrete cosine transform (DCT) basis ($\hat{\mathbf{s}}$), while compression is performed in the time domain ($\mathbf{y} = \Phi\mathbf{x}$).

2.1. ECG Compression

The digitized signal $\tilde{\mathbf{x}}$ of the original ECG is approximated into \mathbf{x} by selecting only K dominant elements of $\tilde{\mathbf{s}} = \Psi^{-1}\tilde{\mathbf{x}}$ where Ψ is the $N \times N$ discrete cosine transform (DCT) basis matrix. In other words, $\tilde{\mathbf{s}}$ can be approximated as a K -sparse signal \mathbf{s} with negligible information loss as:

$$\tilde{\mathbf{x}} = \Psi\tilde{\mathbf{s}} \approx \mathbf{x} = \Psi\mathbf{s}. \quad (2)$$

After that, \mathbf{x} is compressed into $\mathbf{y} \in \mathbb{R}^{M \times N}$ ($M \ll N$) as:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{v} = \Phi\Psi\mathbf{s} + \mathbf{v} = \mathbf{A}\mathbf{s} + \mathbf{v} = \sum_{j \in T} a_j s_j + \mathbf{v} \quad (3)$$

where $\Phi \in \mathbb{R}^{M \times N}$ is the sensing matrix (or compression matrix), $\mathbf{A} = \Phi\Psi$, a_j and s_j are the j -th column of \mathbf{A} and the j -th entry of \mathbf{s} , respectively, and $\mathbf{v} \in \mathbb{R}^M$ is the additive noise (while the sparse structure in the DCT domain is still preserved, the noise \mathbf{v} denotes the measurement distortion of \mathbf{y} after the compression procedure or during the transmission process). Note that since $|T| = K \ll N$, CS-based compression offers the linear superposition of K elements of \mathbf{s} and, thus, enables its implementation with a substantially small number of digital architectures. From the measurement reduction perspective, it is worthwhile noting that the support information at the compression stage cannot be jointly provided to the reconstruction part. That is, for T to be given at the reconstruction stage, the information amount to be delivered increases from $\dim(\mathbf{y}) = M$ to $\dim(\mathbf{y}) + |T| = M + K$ ($\dim(\mathbf{y})$ is the dimension of \mathbf{y}), which is against our intention. Furthermore, since the compression is based on the approximated signal \mathbf{s} with a sparse structure in the DCT basis, the sensing matrix should then obey the restricted isometry property (RIP) given as:

$$\sqrt{1 - \delta_K} \|\mathbf{s}\|_2 \leq \|\Phi\Psi\mathbf{s}\|_2 = \|\mathbf{A}\mathbf{s}\|_2 \leq \sqrt{1 + \delta_K} \|\mathbf{s}\|_2.$$

Considering such property, one of the good choices for the sensing matrix Φ is a random matrix, since such a matrix is said to obey the RIP with high probability [33].

2.2. ECG Reconstruction via Tree Search

In order to achieve low complexity ECG reconstruction with improved recovery accuracy, two key pruning criteria of our method are the pre-scanning and the pruning-based tree search. In the pre-scanning stage, we greedily choose the columns of $\mathbf{A} = \Phi\Psi$ that are highly likely to be associated with nonzero entries of the sparse vector. In other words, the pre-scanning reduces the index set to be investigated from $\Omega = \{1, 2, \dots, N\}$ to a small subset Θ of Ω (i.e., $\Theta \subset \Omega$ and $|\Theta| \ll N$). Then, the tree search is performed by using only the elements of Θ (see Figure 3). While any existing sparse recovery algorithm can be used to obtain Θ , we use a simple method for complexity reduction by choosing only K indices in this work. That is, we select K column indices of $\mathbf{A} = \Phi\Psi$ corresponding to the columns with maximum correlation in magnitude with \mathbf{y} as:

$$\Theta = \arg \max_{|I_K|=K} \|\mathbf{A}'_{I_K} \mathbf{y}\|_2 \tag{4}$$

where I_K is an arbitrary subset of Ω with cardinality K and \mathbf{A}_{I_K} is the submatrix of \mathbf{A} containing the columns associated with the indices in I_K . Note that since Θ is constructed by simply choosing K indices corresponding to maximally correlated columns with \mathbf{y} , the computational burden in the first stage is nearly negligible.

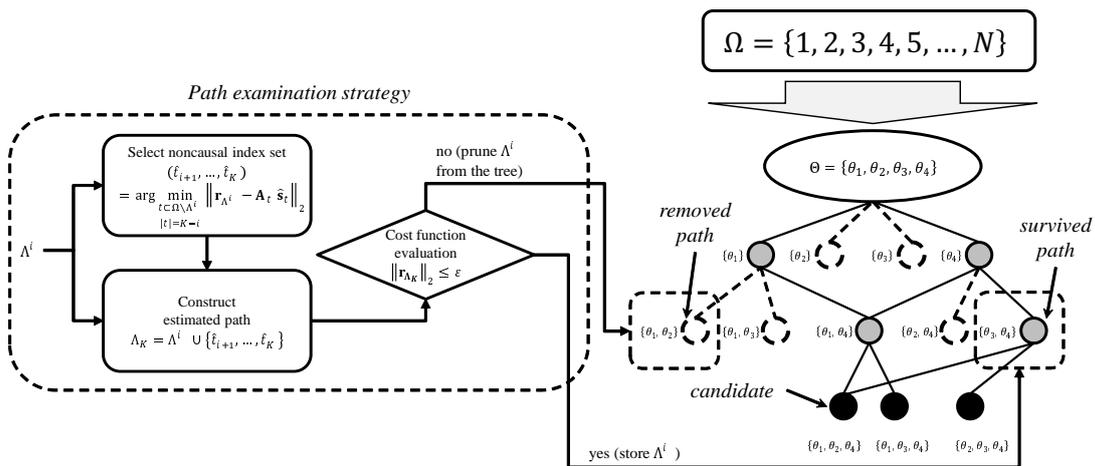


Figure 3. Illustration of the proposed method when $|\Theta| = 4$ and $K = 3$. The branches of each node consist of only the elements in Θ , and the paths with large cost functions are removed from the search.

Once the pre-scanning is finished, a pruning-based tree search is performed to select the index set that minimizes the cost function. In this stage, an aggressive pruning strategy is employed to remove the paths with a small possibility of being the support (index set of nonzero entries). As shown in Figure 3, only the paths that are not removed in the i -th layer spread branches in the $(i + 1)$ -th layer. The pruning strategy is based on removing paths with a larger cost function than the pruning threshold ϵ , since such paths have little hope to be the support. In the beginning of the search, the initial pruning threshold ϵ can be determined as any positive number, since the cost function of the support T is zero ($\|\mathbf{r}_T\|_2 = 0$) and thus the true path can survive as long as T is found at least once. In order to compute the cost function $J(\Lambda^i) = \|\mathbf{y} - \mathbf{A}_{\Lambda^i} \hat{\mathbf{s}}_{\Lambda^i}\|_2$ of the path $\Lambda^i = \{t_1, t_2, \dots, t_i\}$, we obtain the temporarily required indices following the current path, the so-called posterior indices. By doing so, proposed TPMP greedily obtains the remaining part of each path Λ^i and estimates its residual in magnitude at the end of the search (i.e., bottom layer). To this end, the posterior index set $\{\hat{t}_{i+1}, \hat{t}_{i+2}, \dots, \hat{t}_K\}$ of each path is temporarily chosen where \hat{t}_ℓ ($i + 1 \leq \ell \leq K$) are highly likely to be the support among

the elements of $\Theta \setminus \Lambda^i$. In fact, a similar concept of estimating the residual magnitude when the search is completed was proposed in [20]. While [20] presented three cost models to directly estimate the residual in magnitude (for example, when using the multiplicative cost model, the estimated residual in magnitude at the bottom layer is determined by the multiplication of a constant α and residual magnitude of the current path, i.e., $\alpha \|\mathbf{r}_{\Lambda^i}\|_2$), we focus on obtaining actual child node of each path. This problem is yet another problem of reconstructing the $(K - i)$ -sparse signal, and in fact, the proper choice of the algorithm enables sufficient reconstruction accuracy with practical computational complexity (in our numerical simulations, we used the subspace pursuit (SP) algorithm). For instance, one can attempt to find $\{\hat{t}_{i+1}, \dots, \hat{t}_K\}$ minimizing the residual in magnitude:

$$\{\hat{t}_{i+1}, \dots, \hat{t}_K\} = \arg \min_{\substack{I \subset \Theta \setminus \Lambda^i, \\ |I|=K-i}} \|\mathbf{r}_{\Lambda^i} - \mathbf{A}_I \hat{\mathbf{s}}_I\|_2 \quad (5)$$

where $\mathbf{r}_{\Lambda^i} = \mathbf{y} - \mathbf{A}_{\Lambda^i} \hat{\mathbf{s}}_{\Lambda^i}$, $\hat{\mathbf{s}}_{\Lambda^i} = \mathbf{A}_{\Lambda^i}^\dagger \mathbf{r}_{\Lambda^i}$ and $\mathbf{A}_{\Lambda^i}^\dagger = (\mathbf{A}_{\Lambda^i}' \mathbf{A}_{\Lambda^i})^{-1} \mathbf{A}_{\Lambda^i}'$. To be specific, the posterior indices $\{\hat{t}_{i+1}, \dots, \hat{t}_K\}$ in Equation (5) can be obtained by MMP [19], where this choice is to pursue accurate estimation of the cost function. On the other hand, other greedy methods, such as orthogonal matching pursuit (OMP) [21] or subspace pursuit (SP) [23], can be also used for simpler hardware implementations.

After the posterior index set is obtained, the cost function of Λ^i is then computed using $\Lambda_K = \Lambda^i \cup \{\hat{t}_{i+1}, \dots, \hat{t}_K\}$ (note that the cost function is computed by using the candidate with cardinality K). That is, if the ℓ_2 -norm of the residual is greater than the threshold ϵ ($\|\mathbf{r}_{\Lambda_K}\|_2 > \epsilon$), then the path is removed and whenever the search of a layer is finished and $\|\mathbf{r}_{\Lambda_K}\|_2$ is replaced as the newly updated ϵ . The construction of the posterior index set by using the existing greedy method might be computationally burdensome if a nontrivial number of paths exist in the tree. Therefore, we attempt to additionally alleviate such search complexity by employing the stopping criteria, which constrains the minimum residual in magnitude by $cE[\|\mathbf{v}\|_2^2] = cN\sigma^2$ for some non-negative constant $0 \leq c \leq 1$. In fact, although we assumed c to satisfy $0 \leq c \leq 1$ since $\|\mathbf{r}_T\|_2 = \|\mathbf{P}_T^\perp \mathbf{v}\|_2 \leq \|\mathbf{v}\|_2$, if a larger error tolerance is acceptable, c can be assumed to be any proper positive constant larger than 1. In the noise-free scenario ($\mathbf{v} = \mathbf{0}$), the initial pruning threshold ϵ can be determined as any positive number. This is because since $\|\mathbf{r}_T\|_2 = 0$ for noiseless \mathbf{y} , any positive ϵ is larger than the cost function of the support T (set of nonzero elements of the sparse vector) and thus the true path can survive as long as Λ_K is obtained as T at least once. Therefore, if we set $c = 0$ and whenever any path satisfying $\mathbf{r}_{\Lambda_K} = \mathbf{0}$ is found, then we regard Λ_K as the support and immediately shut down the search. On the other hand, in the noisy scenario ($\mathbf{v} \neq \mathbf{0}$), we assume a positive c ($0 < c$), since $\|\mathbf{r}_T\|_2 = \|\mathbf{P}_T^\perp \mathbf{v}\|_2 > 0$. Note that from the accurate reconstruction perspective, too aggressive pruning should be avoided, and thus, small c should be assumed, and vice versa for complexity reduction. Through performance guarantee analyses in Section 3 and numerical simulations in Section 4, we demonstrate that this stopping criteria not only improves the recovery performance as well as the condition bound, but also achieves substantial reduction in search complexity. We summarize the proposed TPMP algorithm in Table 1.

Table 1. The TPMP algorithm.

Input: Measurement \mathbf{y} , sensing matrix \mathbf{A} , sparsity K , $0 < c < 1$	
Output: Reconstructed ECG $\hat{\mathbf{x}}$	
Initialization: $i := 0, W^0 := \emptyset, \epsilon = \infty$	
$\Theta = \arg \max_{ I_k =K} \ \mathbf{A}'_{I_k} \mathbf{y}\ _2$	(pre-scanning)
while $i < K$ do	
$i := i + 1, W^i := \emptyset, \epsilon_{i+1} := \epsilon_i$	
for $l = 1$ to $ W^{i-1} $ do	
$\theta := \Theta \setminus \Lambda^{i-1}(l)$	
for $j = 1$ to $ \theta $ do	
$\Lambda^i := \Lambda^{i-1}(l) \cup \{t_i(j)\}$	(update j -th path)
if $\Lambda^i \notin W^i$ then	(check the duplicated path)
Obtain $\{\hat{t}_{i+1}, \dots, \hat{t}_K\}$	(posterior index set construction)
$\Lambda_K = \Lambda^i \cup \{\hat{t}_{i+1}, \dots, \hat{t}_K\}$	
$\mathbf{r}_{\Lambda_K} = \mathbf{P}_{\Lambda_K}^\perp \mathbf{y}$	
if $\ \mathbf{r}_{\Lambda_K}\ _2 \leq \epsilon_i$ then	(pruning decision)
$W^i := W^i \cup \Lambda^i, I^* := \Lambda_K$	
if $\ \mathbf{r}_{I^*}\ _2^2 \leq cN\sigma^2$ then	(search termination)
$j = \theta + 1, l = W^{i-1} + 1, i = K + 1$	
end if	
if $\ \mathbf{r}_{I^*}\ _2 \leq \epsilon_{i+1}$ then	(update pruning threshold)
$\epsilon_{i+1} := \ \mathbf{r}_{I^*}\ _2$	
end if	
end if	
end for	
end while	
$\hat{\mathbf{s}}^* = \mathbf{A}'_{I^*} \mathbf{y}$	
return $\hat{\mathbf{x}}^* = \mathbf{\Psi} \hat{\mathbf{s}}^*$	(ECG reconstruction)

3. Recovery Bound for Exact Reconstruction

In this section, we provide the sufficient condition under which TPMP accurately reconstructs the K -sparse signal \mathbf{s} . In our analysis, we assume that the posterior index set of each path is constructed by MMP to show how maximally our bound can be relaxed.

The following lemmas are useful for our analysis.

Lemma 1. (Lemma 3 in [8]): *If the sensing matrix \mathbf{A} satisfies the RIP of both orders K_1 and K_2 , then $\delta_{K_1} < \delta_{K_2}$ for any $K_1 < K_2$.*

Lemma 2. (Consequences of RIP [8,22]): *If $0 < \delta_{|I|} < 1$ exists for $I \subset \Omega$, then for any vector $\mathbf{x} \in \mathbb{R}^{|\Omega|}$,*

$$\begin{aligned} (1 - \delta_{|I|}) \|\mathbf{s}\|_2 &\leq \|\mathbf{A}'_I \mathbf{A}_I \mathbf{s}\|_2 \leq (1 + \delta_{|I|}) \|\mathbf{s}\|_2, \\ \frac{1}{1 + \delta_{|I|}} \|\mathbf{s}\|_2 &\leq \|(\mathbf{A}'_I \mathbf{A}_I)^{-1} \mathbf{s}\|_2 \leq \frac{1}{1 - \delta_{|I|}} \|\mathbf{s}\|_2. \end{aligned}$$

Lemma 3. (Lemma 2.1 in [34]): *Let $I_1, I_2 \subset \Omega$ and $I_1 \cap I_2 = \emptyset$. If $0 < \delta_{|I_1|+|I_2|} < 1$ exists, then:*

$$\|\mathbf{A}'_{I_1} \mathbf{A}_{I_2} \mathbf{s}\|_2 \leq \delta_{|I_1|+|I_2|} \|\mathbf{s}\|_2.$$

Lemma 4. *For $M \times N$ matrix \mathbf{A} , $\|\mathbf{A}\|_2$ is bounded as:*

$$\|\mathbf{A}\|_2 \leq \sqrt{1 + \delta_{\min(M,N)}} \quad (6)$$

Definition 2. (Residual definition in [35]): For the index set Λ^i with cardinality $|\Lambda^i| = i$, the residual \mathbf{r}_{Λ^i} is rewritten as:

$$\begin{aligned}\mathbf{r}_{\Lambda^i} &= \mathbf{P}_{\Lambda^i}^\perp \mathbf{y} = \mathbf{y} - \mathbf{A}_{\Lambda^i} \hat{\mathbf{s}}_{\Lambda^i} \\ &= \mathbf{y} - \mathbf{A}_{\Lambda^i} (\mathbf{A}'_{\Lambda^i} \mathbf{A}_{\Lambda^i})^{-1} \mathbf{A}'_{\Lambda^i} \mathbf{y} \\ &= \mathbf{A}_{T \setminus \Lambda^i} \mathbf{s}_{T \setminus \Lambda^i} - \mathbf{A}_{\Lambda^i} (\mathbf{A}'_{\Lambda^i} \mathbf{A}_{\Lambda^i})^{-1} \mathbf{A}'_{\Lambda^i} \mathbf{A}_{T \setminus \Lambda^i} \mathbf{s}_{T \setminus \Lambda^i} \\ &= \mathbf{A}_{T \setminus \Lambda^i} \mathbf{s}_{T \setminus \Lambda^i} - \mathbf{A}_{\Lambda^i} \mathbf{z}_{\Lambda^i} \\ &= \mathbf{A}_{T \cup \Lambda^i} \bar{\mathbf{s}}_{T \cup \Lambda^i}\end{aligned}\quad (7)$$

where $\mathbf{P}_{\Lambda^i}^\perp = \mathbf{I} - \mathbf{A}_{\Lambda^i} (\mathbf{A}'_{\Lambda^i} \mathbf{A}_{\Lambda^i})^{-1} \mathbf{A}'_{\Lambda^i}$ is the projection matrix onto the orthogonal complement of $\text{span}(\Lambda^i)$, $\mathbf{z}_{\Lambda^i} = (\mathbf{A}'_{\Lambda^i} \mathbf{A}_{\Lambda^i})^{-1} \mathbf{A}'_{\Lambda^i} \mathbf{A}_{T \setminus \Lambda^i} \mathbf{s}_{T \setminus \Lambda^i}$ and $\bar{\mathbf{s}}_{T \cup \Lambda^i} = [\mathbf{s}_{T \setminus \Lambda^i} - \mathbf{z}_{\Lambda^i}]'$.

3.1. Exact Recovery from Noiseless Measurements

TPMP is guaranteed to exactly reconstruct \mathbf{s} if the following two conditions are jointly satisfied:

- (3-1) (Theorem 1) At least one support index should be found in the pre-scanning (i.e., $T \cap \Theta \neq \emptyset$).
- (3-2) (Theorem 2) At least one true path $\Lambda^i \subset T$ has to survive the pruning strategy in each layer.

If (3-1) holds, then at least one branch in each layer of the tree is the support element. Therefore, whenever there is at least one true path in current layer that is not removed by the pruning strategy, (3-1) enables the true path to proceed further. Along with (3-1), the additional condition that ensures the survival of the true path is necessary for exact support identification, which is given as (3-2). In our analysis, both (3-1) and (3-2) are guaranteed under the results in Theorems 1 and 2, respectively, and these theorems jointly provide the overall sufficient condition for exact recovery in Theorem 3.

First, we obtain the sufficient condition for (3-1). Let κ be the largest correlation in magnitude between \mathbf{y} and the columns associated with correct indices ($j \in T$) and ζ be the K -th largest correlation in magnitude between \mathbf{y} and the columns corresponding to incorrect indices ($j \in T^c$). That is,

$$\kappa = \max_{j \in T} |a'_j \mathbf{y}|, \quad \zeta = \min_{j \in I_K} |a'_j \mathbf{y}|$$

where a_j is the j -th column of \mathbf{A} and $I_K = \arg \max_{|I|=K, I \subset T^c} \|\mathbf{A}'_I \mathbf{y}\|_2$. The following lemma provides the lower and the upper bounds of κ and ζ , respectively.

Lemma 5. κ and ζ satisfy:

$$\kappa \geq \frac{1 - \delta_{2K}}{\sqrt{K}} \|\mathbf{s}_T\|_2, \quad \zeta \leq \frac{\delta_{2K}}{\sqrt{K}} \|\mathbf{s}_T\|_2. \quad (8)$$

Proof of Lemma 5. See Appendix A. \square

Using Lemma 5, one can obtain the sufficient condition for (3-1).

Theorem 1 (Sufficient condition for (3-1)). *At least one support element is found in the pre-scanning stage under:*

$$\delta_{2K} < 0.5. \quad (9)$$

Proof of Theorem 1. In order to choose at least one correct index in the pre-scanning stage, we should have $\kappa > \zeta$. From Lemma 5, we can easily obtain the desired result. \square

Next, the condition (3-2) is guaranteed if the posterior indices of a true path always contain the support elements, that is $\{\hat{t}_{i+1}, \dots, \hat{t}_K\} = \Omega \setminus \Lambda^i$ where $\Lambda^i \subset T$. This is because $\|\mathbf{r}_T\|_2 = 0$, and thus, the condition (3-2) always holds for any positive pruning threshold ϵ as:

$$\|\mathbf{r}_{\Lambda^i \cup \{\hat{t}_{i+1}, \dots, \hat{t}_K\}}\|_2 = \|\mathbf{r}_T\|_2 < \epsilon. \tag{10}$$

As mentioned, the problem to choose the posterior index set for a given true path $\Lambda^i \subset T$ is equivalent to the problem of reconstructing the $(K - i)$ -sparse signal from the measurement \mathbf{r}_{Λ^i} . Before we proceed, we provide useful definitions in our analysis. Let Y_l be the combination of Λ^i and $\{\hat{t}_{i+1}, \dots, \hat{t}_{i+l}\}$ where $1 \leq l \leq K - i$ ($Y_l = \Lambda^i \cup \{\hat{t}_{i+1}, \dots, \hat{t}_{i+l}\}$). Next, let λ^l be the largest correlation in magnitude between the residual \mathbf{r}_{Y_l} and columns associated with correct indices and γ^l be the $(K - i)$ -th largest correlation in magnitude between \mathbf{r}_{Y_l} and columns associated with incorrect indices. That is,

$$\lambda^l = \max_{j \in T \setminus Y_l} |a'_j \mathbf{r}_{Y_l}|, \quad \gamma^l = \min_{j \in D_i} |a'_j \mathbf{r}_{Y_l}|$$

where $D_i = \arg \max_{|D|=K-i, D \subset \Omega \setminus T} \|\mathbf{A}'_D \mathbf{r}_{Y_l}\|_2$. In the following lemma, we provide the lower bound of λ^l and the upper bound of γ^l .

Lemma 6. *If $Y_l \subset T$, then:*

$$\lambda^l \geq \frac{1 - \delta_K}{\sqrt{K-1}} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2, \quad \gamma^l \leq \frac{\delta_{2K-1}}{\sqrt{K-1}} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2. \tag{11}$$

where $\bar{\mathbf{s}}_{T \cup Y_l} = \left[\mathbf{s}_{T \setminus Y_l} - (\mathbf{A}'_{Y_l} \mathbf{A}_{Y_l})^{-1} \mathbf{A}'_{Y_l} \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l} \right]^T$.

Proof of Lemma 6. See Appendix B. \square

The following theorem provides the sufficient condition for (3-2).

Theorem 2 (Sufficient condition for (3-2)). *The posterior index set of a true path $\Lambda^i \subset T$ consists only of correct ones under:*

$$\delta_{2K-1} < 0.5. \tag{12}$$

for any $1 \leq i \leq K - 1$.

Proof of Theorem 2. The element of posterior index set \hat{t}_{i+l} becomes $\hat{t}_{i+l} \in T$ for any $1 \leq l \leq K - i$ if the inequality $\lambda^l > \gamma^l$ is satisfied. That is, from Lemma 6, we have:

$$\delta_K + \delta_{2K-1} < 1.$$

From Lemma 1, this inequality can be rewritten as $2\delta_{2K-1} < 1$, which is the desired result. \square

The overall recovery condition of TPMP can be obtained by combining Theorems 1 and 2.

Theorem 3 (Recovery condition of TPMP). *TPMP exactly identifies the support of any K -sparse signal \mathbf{s} from $\mathbf{y} = \mathbf{A}\mathbf{s}$ under:*

$$\delta_{2K} < 0.5. \tag{13}$$

Proof of Theorem 3. The condition (Equation (13)) is obtained by choosing a stricter condition between Theorems 1 and 2. \square

It is worthwhile to note that TPMP provides a more relaxed recovery condition than conventional greedy algorithms, such as OMP ($\delta_{K+1} < \frac{1}{\sqrt{K+1}}$) [21], SP ($\delta_{3K} < 0.165$) [23], gOMP ($\delta_{K^2} < \frac{\sqrt{L}}{\sqrt{L+\sqrt{K}}}$) [35] and MMP ($\delta_{2K} < 0.33$). In addition, our condition achieves the state-of-the-art recovery bound for the greedy algorithm, which was presented in [24].

3.2. Reconstruction from Noisy Measurements

We also consider reconstructing ECG when the compressed signal \mathbf{y} is distorted by noise. In this scenario, the measurement \mathbf{y} is defined as:

$$\mathbf{y} = \Phi\mathbf{x} + \mathbf{v} = \mathbf{A}\mathbf{s} + \mathbf{v}$$

where \mathbf{v} is an additive noise vector. Using the new expression of \mathbf{y} containing \mathbf{v} , we analyze the condition of TPMP to accurately identify the support by following the main architecture of the proofs for the noiseless scenario. Two requirements of TPMP to identify the support are (1) at least one support element should be chosen in the pre-scanning process (i.e., $T \cap \Omega \neq \emptyset$) (Theorem 4) and (2) true path ($\Lambda^i \subset T$) should survive the pruning strategy (Theorem 7). It is worth noting that while the pre-scanning condition (Theorem 4) is similar to that in the previous section, the search tree condition (Theorem 7) should satisfy the additional requirement compared to the result in Theorem 2. In the noiseless scenario (i.e., $\mathbf{v} = \mathbf{0}$), the support T always survives the pruning strategy whenever it is detected once since $\|\mathbf{r}_T\|_2 = 0$ is the minimum residual in magnitude. On the other hand, in the presence of noise (i.e., $\mathbf{v} \neq \mathbf{0}$), the additional guarantee of the support having the minimum residual in magnitude is required, that is,

$$\arg \min_{\Lambda_K} \|\mathbf{r}_{\Lambda_K}\|_2 = T$$

should hold to ensure the search tree condition.

Before we proceed, we provide a useful lemma in our analysis.

Lemma 7. For any $Y_l \subset T$, $\bar{\mathbf{s}}_{T \cup Y_l} = [\mathbf{s}_{T \setminus Y_l} - \mathbf{z}_{Y_l}]'$ satisfies:

$$\|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 \geq \frac{\sqrt{2(1 + \delta_K^2)}}{1 + \delta_K} \|\mathbf{s}_{T \setminus Y_l}\|_2. \tag{14}$$

Proof of Lemma 7. Since $\mathbf{z}_{\Lambda^i} = (\mathbf{A}'_{\Lambda^i} \mathbf{A}_{\Lambda^i})^{-1} \mathbf{A}'_{\Lambda^i} \mathbf{A}_{T \setminus \Lambda^i} \mathbf{s}_{T \setminus \Lambda^i}$ (see Lemma 2), $\|\bar{\mathbf{s}}_{T \cup Y_l}\|_2^2$ is:

$$\|\bar{\mathbf{s}}_{T \cup Y_l}\|_2^2 = \left\| \begin{bmatrix} \mathbf{s}_{T \setminus Y_l} \\ -(\mathbf{A}'_{Y_l} \mathbf{A}_{Y_l})^{-1} \mathbf{A}'_{Y_l} \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l} \end{bmatrix} \right\|_2^2 \tag{15}$$

$$= \|\mathbf{s}_{T \setminus Y_l}\|_2^2 + \|(\mathbf{A}'_{Y_l} \mathbf{A}_{Y_l})^{-1} \mathbf{A}'_{Y_l} \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l}\|_2^2. \tag{16}$$

From Lemma 2 and Definition 1, we then have:

$$\begin{aligned} & \|(\mathbf{A}'_{Y_l} \mathbf{A}_{Y_l})^{-1} \mathbf{A}'_{Y_l} \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l}\|_2^2 \\ & \geq \frac{1}{(1 + \delta_{|Y_l|})^2} \|\mathbf{A}_{Y_l} \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l}\|_2^2 \\ & \geq \frac{1 - \delta_{Y_l}}{(1 + \delta_{|Y_l|})^2} \|\mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l}\|_2^2 \\ & \geq \frac{(1 - \delta_{Y_l})(1 - \delta_{T \setminus Y_l})}{(1 + \delta_{|Y_l|})^2} \|\mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l}\|_2^2 \\ & \geq \frac{(1 - \delta_K)^2}{(1 + \delta_K)^2} \|\mathbf{s}_{T \setminus Y_l}\|_2^2. \end{aligned} \tag{17}$$

From Equations (16) and (17), we obtain the lower bound of $\|\bar{\mathbf{s}}_{T \cup Y_i}\|_2$ as:

$$\begin{aligned} \|\bar{\mathbf{s}}_{T \cup Y_i}\|_2^2 &\geq \|\mathbf{s}_{T \setminus Y_i}\|_2^2 + \frac{(1 - \delta_K)^2}{(1 + \delta_K)^2} \|\mathbf{s}_{T \setminus Y_i}\|_2^2 \\ &\geq \frac{2(1 + \delta_K^2)}{(1 + \delta_K)^2} \|\mathbf{s}_{T \setminus \Lambda^i}\|_2^2 \end{aligned} \quad (18)$$

which is the desired result. \square

We first analyze the condition ensuring at least one support element is chosen by pre-scanning from noisy measurements. Let ρ be the largest correlation in magnitude between \mathbf{y} , and the columns associated with correct indices and η be the K -th largest correlation in magnitude between \mathbf{y} and the columns associated with incorrect indices. That is,

$$\rho = \max_{j \in T} |a'_j \mathbf{y}|, \quad \eta = \min_{j \in I_K} |a'_j \mathbf{y}|$$

where $I_K = \arg \max_{|I|=K, I \subset T^c} \|\mathbf{A}'_I \mathbf{y}\|_2$.

In the following lemmas, we provide the lower bound of ρ and the upper bound of η .

Lemma 8. ρ satisfies:

$$\rho \geq \frac{1}{\sqrt{K}} \left[(1 - \delta_K) \|\mathbf{s}_T\|_2 - \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \right] \quad (19)$$

and η satisfies:

$$\eta \leq \frac{1}{\sqrt{K}} \left[\delta_{2K} \|\mathbf{s}_T\|_2 + \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \right]. \quad (20)$$

Proof of Lemma 8. See Appendix C. \square

The following theorem provides the condition ensuring that at least one support element is identified by the pre-scanning.

Theorem 4 (Pre-scanning condition). *At least one element in the support is found in the pre-scanning stage if the nonzero entries of the original sparse signal \mathbf{s} satisfy:*

$$\min_{j \in T} |s_j| > \frac{2\sqrt{1 + \delta_K}}{\sqrt{K}(1 - \delta_K - \delta_{2K})} \|\mathbf{v}\|_2. \quad (21)$$

Proof of Theorem 4. It is clear that at least one support element ($j \in T$) is chosen in the pre-scanning if:

$$\rho > \eta. \quad (22)$$

From Lemma 8, Equation (22) can be rewritten as:

$$\frac{1}{\sqrt{K}} \left[(1 - \delta_K) \|\mathbf{s}_T\|_2 - \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \right] > \frac{1}{\sqrt{K}} \left[\delta_{2K} \|\mathbf{s}_T\|_2 + \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \right] \quad (23)$$

and thus, we have:

$$\|\mathbf{s}_T\|_2 > \frac{2\sqrt{1 + \delta_K}}{1 - \delta_K - \delta_{2K}} \|\mathbf{v}\|_2. \quad (24)$$

Since $\|\mathbf{s}_T\|_2 \geq \sqrt{K} \min_{j \in T} |s_j|$, we obtain the desired result as:

$$\min_{j \in T} |s_j| > \frac{2\sqrt{1+\delta_K}}{\sqrt{K}(1-\delta_K-\delta_{2K})} \|\mathbf{v}\|_2. \quad (25)$$

□

Next, we analyze the sufficient condition ensuring that the true path is not removed from the search tree. This requirement holds if (1) the posterior indices $\{\hat{t}_{i+1}, \dots, \hat{t}_K\}$ of any true path $\Lambda^i \subset T$ satisfy $\{\hat{t}_{i+1}, \dots, \hat{t}_K\} = T \setminus \Lambda^i$ and (2) the corresponding $\Lambda_K = \Lambda^i \cup \{\hat{t}_{i+1}, \dots, \hat{t}_K\} = T$ satisfies $\|\mathbf{r}_{\Lambda_K}\|_2 < \epsilon$. For obtaining the condition ensuring $\{\hat{t}_{i+1}, \dots, \hat{t}_K\} = T \setminus \Lambda^i$ for any $\Lambda^i \subset T$, let β^l be the largest correlation in magnitude between correct indices and \mathbf{r}_{Y_l} and α^l be the $(K-i)$ -th largest correlation in magnitude between \mathbf{r}_{Y_l} and columns associated with incorrect indices. That is,

$$\begin{aligned} \beta^l &= \arg \max_{j \in T \setminus Y_l} |a'_j \mathbf{r}_{Y_l}|, \\ \alpha^l &= \arg \min_{j \in D_i} |a'_j \mathbf{r}_{Y_l}| \end{aligned}$$

where $D_i = \arg \max_{|D|=K-i, D \subset \Omega \setminus T} \|\mathbf{A}'_D \mathbf{r}_{Y_l}\|_2$. The following lemma provides the lower bound of β^l and the upper bound of α^l .

Lemma 9. For any $Y_l \subset T$, β^l and α^l satisfy:

$$\beta^l > \frac{1}{\sqrt{K-1}} \left[(1-\delta_K) \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 - \sqrt{1+\delta_{K-1}} \|\mathbf{v}\|_2 \right] \quad (26)$$

and:

$$\alpha^l < \frac{1}{\sqrt{K-1}} \left[\delta_{2K-1} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \sqrt{1+\delta_{K-1}} \|\mathbf{v}\|_2 \right], \quad (27)$$

respectively.

Proof of Lemma 9. See Appendix D. □

The guaranteed condition of the posterior indices to satisfy $\{\hat{t}_{i+1}, \dots, \hat{t}_K\} = T \setminus \Lambda^i$ can be identified by combining Lemmas 7 and 9.

Theorem 5. For any $\Lambda^i \subset T$, the posterior indices satisfy $\{\hat{t}_{i+1}, \dots, \hat{t}_K\} = T \setminus \Lambda^i$ under:

$$\min_{j \in T} |s_j| > \frac{(1+\delta_K) \sqrt{2(1+\delta_{K-1})}}{(1-\delta_K-\delta_{2K-1}) \sqrt{1+\delta_K^2}} \|\mathbf{v}\|_2. \quad (28)$$

Proof of Theorem 5. Similar to Theorem 2, one can notice that the posterior indices of the true path Λ^i contain only true indices for any $1 \leq l \leq K-i$ if:

$$\beta^l > \alpha^l, \quad (29)$$

which can be rewritten by using Lemma 9 as:

$$\frac{1}{\sqrt{K-1}} \left[(1-\delta_K) \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 - \sqrt{1+\delta_{K-1}} \|\mathbf{v}\|_2 \right] > \frac{1}{\sqrt{K-1}} \left[\delta_{2K-1} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \sqrt{1+\delta_{K-1}} \|\mathbf{v}\|_2 \right] \quad (30)$$

where $Y_l \subset T$. After some manipulations, we have:

$$\|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 > \frac{2\sqrt{1 + \delta_{K-1}}}{1 - \delta_K - \delta_{2K-1}} \|\mathbf{v}\|_2. \quad (31)$$

Recall Equation (14) from Lemma 7 that:

$$\|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 \geq \frac{\sqrt{2(1 + \delta_K^2)}}{1 + \delta_K} \|\mathbf{s}_{T \setminus Y_l}\|_2$$

and since $\|\mathbf{s}_{T \cup Y_l}\|_2 \geq \min_{j \in T} |s_j|$, we get the desired result. \square

Next, we provide the guaranteed condition under which the residual of the support satisfies:

$$\|\mathbf{r}_T\|_2 \leq \epsilon \quad (32)$$

for any positive pruning threshold ϵ . Recall that the pruning threshold is updated by the smallest residual in magnitude among all Λ_K found in each layer of the tree. Therefore, as long as Equation (32) holds and $\Lambda_K = T$ is found at least once, T has a smaller residual in magnitude than any possible pruning threshold ϵ and, thus, cannot be removed from the search tree.

Theorem 6. *The support has the minimum residual in magnitude among all possible Λ_K ($|\Lambda_K| = K$) if:*

$$\min_{j \in T} |s_j| > \frac{\sqrt{2}(1 + \delta_K) \|\mathbf{v}\|_2}{\sqrt{(1 - \delta_{2K})(1 + \delta_K^2)}}. \quad (33)$$

Proof of Theorem 6. See Appendix E. \square

If Theorems 5 and 6 jointly hold, then the condition that the true path Λ^i is not removed can be guaranteed as follows.

Theorem 7 (Search tree condition). *The true path $\Lambda^i \subset T$ survives the pruning strategy for any i under:*

$$\min_{j \in T} |s_j| > \max(\mu, \omega) \|\mathbf{v}\|_2 \quad (34)$$

where $\mu = \frac{(1 + \delta_K) \sqrt{2(1 + \delta_{K-1})}}{(1 - \delta_K - \delta_{2K-1}) \sqrt{1 + \delta_K^2}}$ and $\omega = \frac{\sqrt{2}(1 + \delta_K)}{\sqrt{(1 - \delta_{2K})(1 + \delta_K^2)}}$.

Proof of Theorem 7. Immediate from Theorems 5 and 6. \square

By combining the results from Theorems 4 and 7, we obtain the sufficient condition of exact support identification from noisy measurements.

Theorem 8 (Exact support identification of TPMP). *The TPMP algorithm accurately identifies the support from the noisy measurement $\mathbf{y} = \mathbf{A}\mathbf{s} + \mathbf{v}$ under:*

$$\min_{j \in T} |s_j| > \max(v, \mu, \omega) \|\mathbf{v}\|_2 = \gamma \|\mathbf{v}\|_2 \quad (35)$$

where $v = \frac{2\sqrt{1 + \delta_K}}{\sqrt{K}(1 - \delta_K - \delta_{2K})}$, $\mu = \frac{(1 + \delta_K) \sqrt{2(1 + \delta_{K-1})}}{(1 - \delta_K - \delta_{2K-1}) \sqrt{1 + \delta_K^2}}$ and $\omega = \frac{\sqrt{2}(1 + \delta_K)}{\sqrt{(1 - \delta_{2K})(1 + \delta_K^2)}}$.

Proof of Theorem 8. Immediate from Theorems 4 and 7. \square

Note that the sufficient condition given in Equation (35) infers that the signal-to-noise ratio (SNR) of the sparse signal should be higher than the constant γ . If Equation (35) holds, the support T can be exactly identified, and the signal reconstruction is based on the columns of \mathbf{A} associated with T . In this sense, the system is equivalent to the overdetermined system ($\mathbf{y} = \mathbf{A}_T \mathbf{s}_T + \mathbf{v}$) and achieves identical performance to the best possible estimator referred to as the oracle estimator $\hat{\mathbf{s}}_T = \mathbf{A}_T^\dagger \mathbf{y}$.

4. Simulation and Discussion

In this section, we evaluate the numerical ECG recovery performance of the proposed TPMP algorithm and the existing sparse recovery algorithms. The simulation is based on the discrete cosine transform (DCT) basis matrix $\mathbf{\Psi} \in N \times N$, the random Bernoulli sensing matrix $\mathbf{\Phi} \in M \times N$ ($M \ll N$) where each element of $\mathbf{\Phi}$ is either $\pm \frac{1}{\sqrt{M}}$, and the measurement is distorted by an additive noise vector (according to the results in [19], we set the signal-to-noise ratio (SNR) as 40 dB when $\mathbf{v} \neq \mathbf{0}$) $\mathbf{v} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. In the simulation, we check the reconstruction performance by performing at least 5000 independent trials for each number of measurements M , which is directly related to the compression ratio (CR) defined as $CR = \frac{N-M}{N} \times 100$ (%). In addition, we used two measures for the performance evaluation: (1) the exact recovery ratio (ERR), which is the probability of the exact identification of the support of \mathbf{s} ($T = \{j | s_j \neq 0\}$) and (2) the percentage root-mean-square difference (PRD), which is defined as:

$$PRD = \frac{\|\tilde{\mathbf{x}} - \hat{\mathbf{x}}\|_2}{\|\tilde{\mathbf{x}} - E[\tilde{\mathbf{x}}]\|_2}$$

where $\tilde{\mathbf{x}}$ is the digitized signal of original ECG and $\hat{\mathbf{x}}$ is the reconstructed ECG. We exploited six ECG samples from the European ST-TDatabase in Physionet [36], and a randomly chosen window of 1000 continuative signal samples among all the data is used for each trial with sparsity level $K = 100$. The samples are measured from distinct patients including people with normal status, left circumflex artery (LCA) or right coronary artery (RCA) diseases (see Figure 4). Each record is two hours in duration and contains two signals, each sampled at 250 samples per second with 12-bit resolution over a nominal 20-mV input range. The sample values were rescaled after digitization with reference to calibration signals in the original analog recordings, in order to obtain a uniform scale of 200 ADC units per mV for all signals, and each of the signal files is 5,400,000 bytes long. All algorithms under test are coded by MATLAB software and run by a personal computer with an Intel Core i5 processor and Microsoft Windows 7.

- Oracle estimator [32]
- Linear MMSE
- Basis pursuit (BP) [9]
- Orthogonal matching pursuit (OMP) [21]
- Subspace pursuit (SP) [23]
- Multipath matching pursuit (MMP) [19] with $L = 2$
- Depth-first multipath matching pursuit (MMP-DF) [19] with $L = 2$: MMP with reduced complexity
- TPMP ($c = 0, 1$)

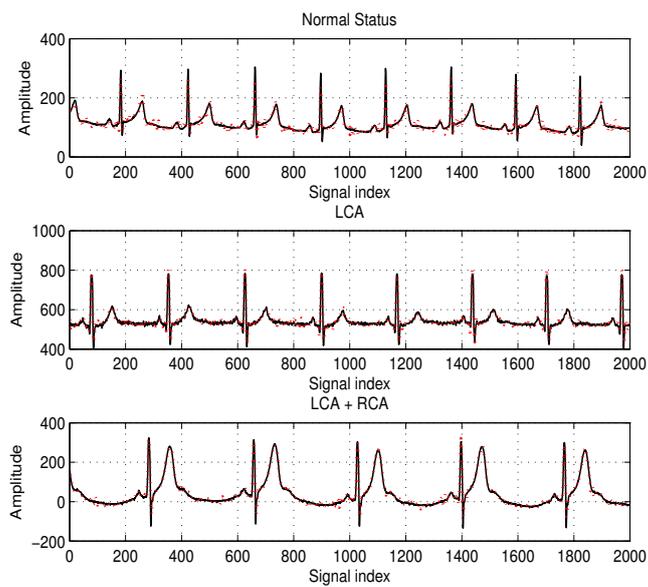


Figure 4. Examples of the tested ECG samples in our simulations and the reconstructed signal for each sample signal by tree pruning-based matching pursuit (TPMP) at compression ratio (CR) = 65, where the solid and dotted lines denote the original ECG (\mathbf{x}) and the reconstructed ECG ($\hat{\mathbf{x}} = \Psi\hat{\mathbf{s}}$), respectively.

Before the discussion, it is clear that higher CR (or smaller M) degrades the reconstruction performance (see Figure 5), and thus, we demonstrate the effectiveness of the proposed method by showing that TPMP requires minimum M for accurately identifying the support.

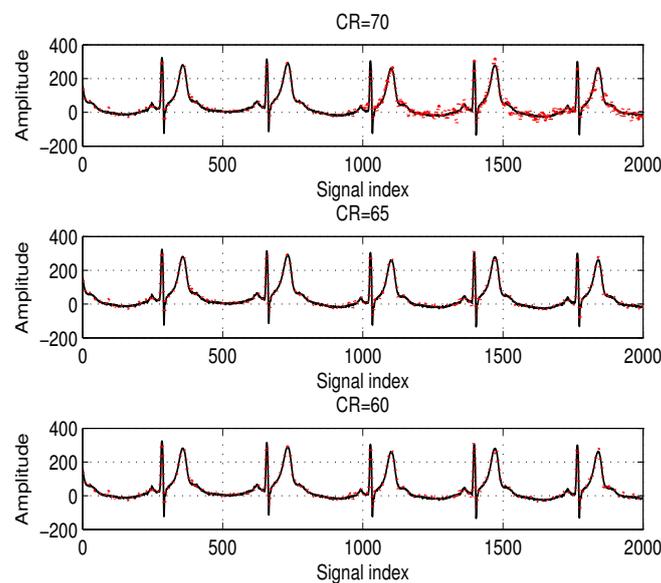


Figure 5. ECG reconstruction using the proposed method when $N = 1000$ and $K = 100$ where the solid and dotted lines denote the original ECG (\mathbf{x}) and the reconstructed ECG ($\hat{\mathbf{x}} = \Psi\hat{\mathbf{s}}$), respectively.

First, we evaluate the ECG reconstruction performance from noiseless measurements ($\mathbf{y} = \Phi\mathbf{x}$). Figure 6 provides the ERR performance when $K = 100$ (i.e., $\frac{K}{N} = 10\%$) as a function of M . Overall, we observe that TPMP performs better than existing algorithms, in particular for small M . In particular, it is shown that while ERR of TPMP drops moderately with the decrement of M , that of other conventional algorithms drops sharply and fails to provide reliable recovery performance. In Figure 7, we plot the PRD performance of the sparse recovery methods. Note that since the exact support information is given to the oracle estimator, it can be regarded to have the lower bound of PRD (since $\tilde{\mathbf{x}}$ is approximated to \mathbf{x} , PRD determined by the approximation error $\|\tilde{\mathbf{x}} - \mathbf{x}\|_2$ is the lower bound of PRD). Due to the multiple candidate investigation, we observe that TPMP reaches the optimal performance with minimum M among tested algorithms. For instance, while SP requires at least $M = 385$ measurements for optimal performance, TPMP requires only $M = 325$. In addition, while MMP provides lower PRD for very small M , the PRD of TPMP reaches the best possible performance with smaller M . This demonstrates that TPMP not only outperforms conventional methods in reconstruction accuracy, but also enables reduction in data storage. Figure 8 shows the running time complexity. Since TPMP performs the tree search, it is no wonder that the running time complexity of TPMP can be large. Interestingly, we observe that the complexity of TPMP becomes similar to that of conventional greedy algorithms in the large M regime. This is because $\|\mathbf{r}_T\|_2 = 0$, and thus, whenever any path satisfying $\|\mathbf{r}_{\Lambda_K}\|_2 = 0$ is found, we regard Λ_K to be the support and immediately stop the search. In this sense, the support can be identified in the early layer of the search.

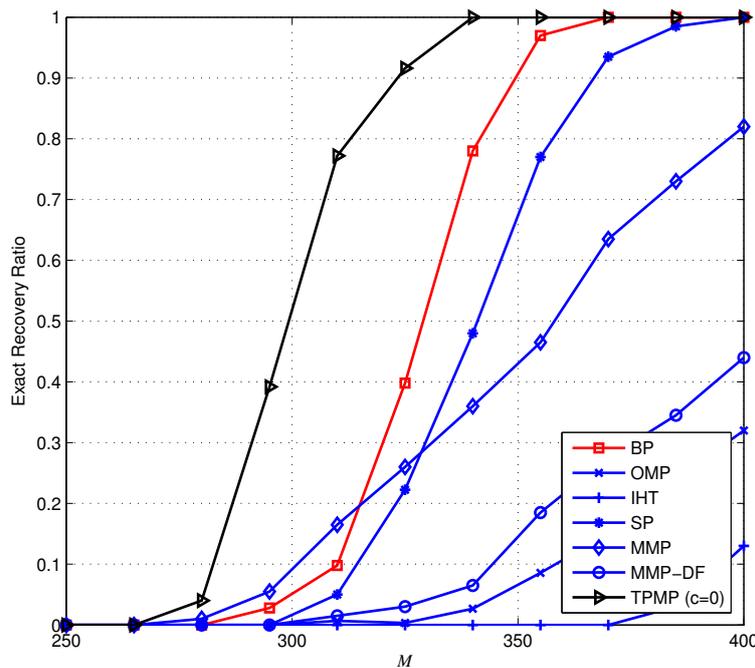


Figure 6. Exact recovery ratio (ERR) performance of the sparse signal recovery methods when $N = 1000$ and $K = 100$.

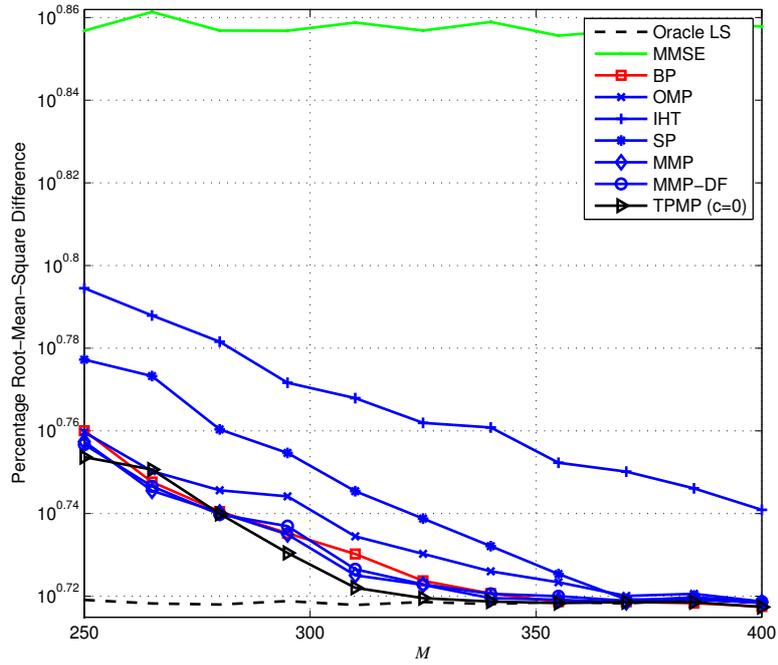


Figure 7. Percentage root-mean-square difference (PRD) performance of the sparse signal recovery methods when $N = 1000$ and $K = 100$.

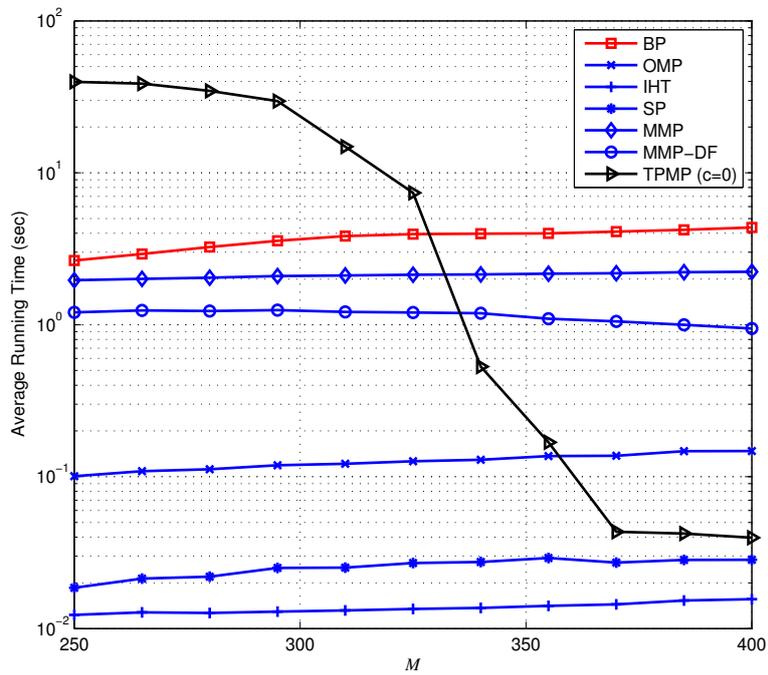


Figure 8. Complexity of the sparse signal recovery methods when $N = 1000$ and $K = 100$.

Next, we provide the recovery performance in the presence of noise, that is when the measurement is defined as $\mathbf{y} = \Phi\mathbf{x} + \mathbf{v}$. Recall that in the noiseless scenario, the search could be finished in the early stage whenever any Λ_K satisfying $\|\mathbf{r}_{\Lambda_K}\|_2 = 0$ is found with $c = 0$. However, this is no longer valid in the presence of noise since $\|\mathbf{r}_T\|_2 = \|\mathbf{P}_T^\perp \mathbf{v}\|_2 \neq 0$, and thus, we assume positive c in the noisy setting. Note that this stopping criterion does not affect the recovery condition since if Theorem 8 holds, then $\|\mathbf{r}_T\|_2$ still has the minimum residual in magnitude, and thus, T is selected as the support whether $\|\mathbf{r}_T\|_2^2 > cN\sigma^2$ or not. Therefore, one can notice that c is used only to shut down the search earlier than the original TPMP, and thus proper choice of c is required for the minimum tradeoff between the numerical performance and the complexity. In Figure 9, we plot the PRD performance of the sparse recovery algorithms. Similar to the noiseless scenario, we observe that the proposed TPMP algorithm outperforms conventional methods. In particular, the PRDs of TPMP with both $c = 0$ and $c = 1$ are smaller than that of MMP. To be specific, we observe in Figure 9 that TPMP performs closest to the lower bound of PRD (PRD of oracle LS) among all of the tested algorithms. In order to demonstrate the validity of real-time implementation, Figure 10 provides the average running time complexity of the sparse recovery algorithms as a function of M . Similar to the results in Figure 8, the running time complexity of TPMP is the highest due to the tree search, especially for small M . Nevertheless, the computational burden of TPMP can be substantially reduced by limiting the minimum pruning threshold determined by c . In particular, if $c = 1$, significant complexity reduction over the original TPMP ($c = 0$) is achieved, and it performs with similar complexity as OMP. In addition, since TPMP with $c = 1$ performs similar to MMP with smaller complexity than MMP and MMP-DF, one can notice that TPMP provides a better tradeoff between the performance and the complexity than MMP.

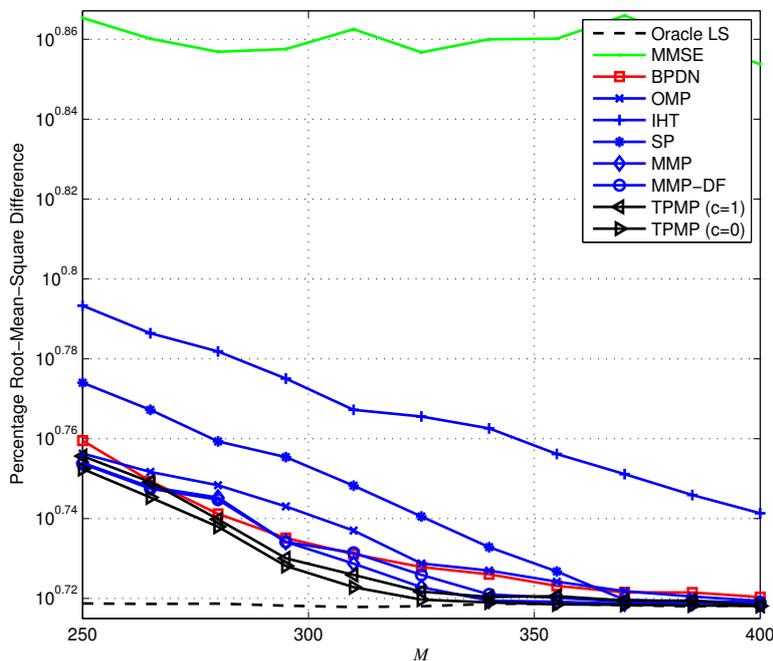


Figure 9. PRD performance of the sparse signal reconstruction from noisy measurements.

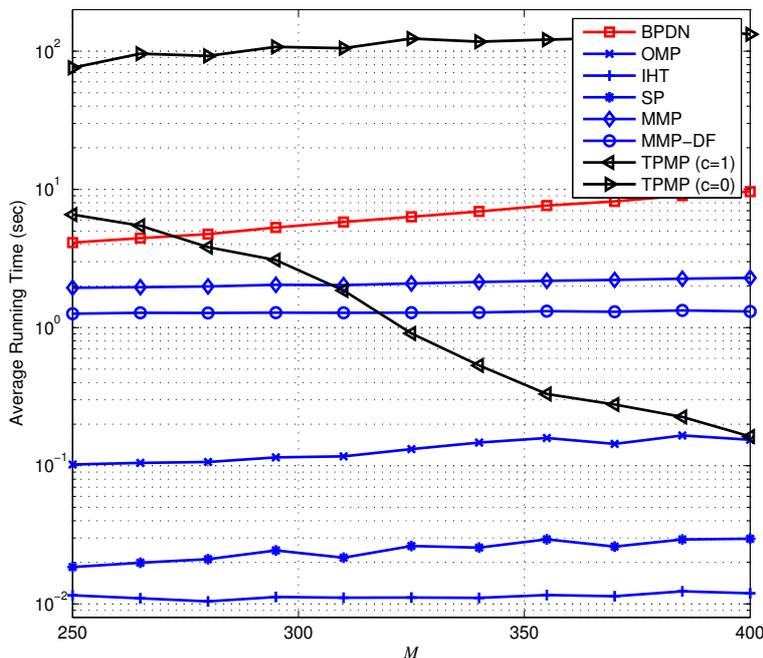


Figure 10. Average running time complexity of the sparse signal reconstruction from noisy measurements.

5. Conclusions

In this work, we proposed an effective ECG reconstruction method referred to as tree pruning-based matching pursuit (TPMP). In order to improve the accuracy of ECG recovery for large CR (or small M), the TPMP algorithm performs the tree search and investigates multiple promising candidates. Further, the complexity overhead caused by the tree search is reduced by the pruning strategy. In our analysis, we analyzed the sufficient condition of TPMP for exactly identifying the support, which provides an improved recovery bound compared to the existing methods. In addition, our numerical results discussed in this work demonstrate that TPMP provides improved performance with competitive complexity with conventional algorithms.

Acknowledgments: This work was partly supported by the Daegu Gyeongbuk Institute of Science and Technology (DGIST) R & D Program of the Ministry of Science, ICT and Future Planning (16-BD-0404), convergence technology development program for bionic arm through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No. 2016M3C1B2912987), and the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2015R1A2A2A01008218).

Author Contributions: Jaeseok Lee and Ji-Woong Choi contributed to the algorithm design, recovery bound analysis and simulation. Kyungsoo Kim contributed to the background knowledge of electrocardiogram processing and the intuition for proper mathematical modeling.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Lemma 5

From the definition of κ , we have:

$$\begin{aligned}
 \kappa &= \max_{j \in T} |a'_j \mathbf{y}| \\
 &= \|\mathbf{A}'_T \mathbf{y}\|_\infty \geq \frac{1}{\sqrt{K}} \|\mathbf{A}'_T \mathbf{y}\|_2
 \end{aligned}
 \tag{A1}$$

where (A1) is from the inequality $\|\mathbf{u}\|_\infty \geq \frac{1}{\sqrt{\|\mathbf{u}\|_0}} \|\mathbf{u}\|_2$ for any vector \mathbf{u} . Using Definition 1, we further have:

$$\begin{aligned} \kappa &\geq \frac{1}{\sqrt{K}} \|\mathbf{A}'_T \mathbf{y}\|_2 = \frac{1}{\sqrt{K}} \|\mathbf{A}'_T \mathbf{A} \mathbf{s}\|_2 \\ &= \frac{1}{\sqrt{K}} \|\mathbf{A}'_T \mathbf{A}_T \mathbf{s}_T\|_2 \\ &\geq \frac{1-\delta_K}{\sqrt{K}} \|\mathbf{s}_T\|_2 \\ &> \frac{1-\delta_{2K}}{\sqrt{K}} \|\mathbf{s}_T\|_2 \end{aligned} \quad (\text{A2})$$

where (A2) is from Lemma 2 and (A2) is from Lemma 1.

Next, from the definition of ζ , we have:

$$\sqrt{K} \zeta \leq \sqrt{\sum_{j \in I_K} |a'_j \mathbf{y}|^2} = \|\mathbf{A}'_{I_K} \mathbf{y}\|_2 \quad (\text{A3})$$

where $I_K = \arg \max_{|I|=K, I \subset T^c} \|\mathbf{A}'_I \mathbf{y}\|_2$. Since I_K and T are disjoint ($I_K \subset T^c$) and $|I_K| = K$, we have:

$$\|\mathbf{A}'_{I_K} \mathbf{y}\|_2 = \|\mathbf{A}'_{I_K} \mathbf{A}_T \mathbf{s}_T\|_2 \leq \delta_{2K} \|\mathbf{s}_T\|_2 \quad (\text{A4})$$

where (A4) is from Lemma 3. From (A3) and (A4), we get the desired result.

Appendix B. Proof of Lemma 6

For any $Y_l = \Lambda^i \cup \{\hat{t}_{i+1}, \dots, \hat{t}_{i+l}\} \subset T$ and $\lambda^l = \max_{j \in T \setminus Y_l} |a'_j \mathbf{r}_{Y_l}|$, we have:

$$\begin{aligned} \max_{j \in T \setminus Y_l} |a'_j \mathbf{r}_{Y_l}| &= \|\mathbf{A}'_{T \setminus Y_l} \mathbf{r}_{Y_l}\|_\infty \\ &\geq \frac{1}{\sqrt{|T \setminus Y_l|}} \|\mathbf{A}'_{T \setminus Y_l} \mathbf{r}_{Y_l}\|_2 \\ &= \frac{1}{\sqrt{K-i-l}} \|\mathbf{A}'_T \mathbf{A}_{T \cup Y_l} \bar{\mathbf{s}}_{T \cup Y_l}\|_2 \end{aligned} \quad (\text{B1})$$

where $\mathbf{P}_{Y_l}^\perp = \mathbf{I} - \mathbf{A}_{Y_l} (\mathbf{A}'_{Y_l} \mathbf{A}_{Y_l})^{-1} \mathbf{A}'_{Y_l}$, and (B1) is from Definition 2. Since $T \cup Y_l = T$ ($Y_l \subset T$), (B1) can be rewritten from Lemma 2 as:

$$\begin{aligned} \max_{j \in T \setminus Y_l} |a'_j \mathbf{r}_{Y_l}| &\geq \frac{1}{\sqrt{K-i-l}} \|\mathbf{A}'_T \mathbf{A}_{T \cup Y_l} \bar{\mathbf{s}}_{T \cup Y_l}\|_2 \\ &= \frac{1}{\sqrt{K-i-l}} \|\mathbf{A}'_T \mathbf{A}_T \bar{\mathbf{s}}_{T \cup Y_l}\|_2 \\ &\geq \frac{1-\delta_{|T|}}{\sqrt{K-i-l}} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 \\ &\geq \frac{1-\delta_{|T|}}{\sqrt{K-i}} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 \\ &> \frac{1-\delta_K}{\sqrt{K}} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 \end{aligned} \quad (\text{B2})$$

where (B2) is from Lemma 1.

Next, From the definition of $\gamma^l = \min_{j \in D_K} |a'_j \mathbf{r}_{Y_l}|$, we have:

$$\sqrt{|D_K|} \gamma^l \leq \sqrt{\sum_{j \in D_K} |a'_j \mathbf{r}_{Y_l}|^2} = \|\mathbf{A}'_{D_K} \mathbf{r}_{Y_l}\|_2 \quad (\text{B3})$$

where $D_K = \arg \max_{|D|=K-i, D \subset \Omega \setminus T} \|\mathbf{A}'_D \mathbf{r}_{Y_i}\|_2$ and $Y_i \subset T$. Using the triangle inequality, we have:

$$\|\mathbf{A}'_{D_K} \mathbf{r}_{Y_i}\|_2 = \|\mathbf{A}'_{D_K} \mathbf{A}_{T \cup Y_i} \bar{\mathbf{s}}_{T \cup Y_i}\|_2 \quad (\text{B4})$$

$$= \|\mathbf{A}'_{D_K} \mathbf{A}_T \bar{\mathbf{s}}_{T \cup Y_i}\|_2 \quad (\text{B5})$$

$$\leq \delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_i}\|_2 \quad (\text{B6})$$

$$< \delta_{2K} \|\bar{\mathbf{s}}_{T \cup Y_i}\|_2 \quad (\text{B7})$$

where (B4) is from Definition 2, (B5) is because $T \cup Y_i = T$, (B6) is from Lemma 3 and (B7) is from Lemma 1.

Appendix C. Proof of Lemma 8

From the definition of ρ , we have:

$$\begin{aligned} \rho &= \max_{j \in T} |a'_j \mathbf{y}| = \|\mathbf{A}'_T \mathbf{y}\|_\infty \\ &\geq \frac{1}{\sqrt{|T|}} \|\mathbf{A}'_T \mathbf{y}\|_2 \\ &= \frac{1}{\sqrt{K}} \|\mathbf{A}'_T (\mathbf{A}_T \mathbf{s}_T + \mathbf{v})\|_2 \\ &\geq \frac{1}{\sqrt{K}} (\|\mathbf{A}'_T \mathbf{A}_T \mathbf{s}_T\|_2 - \|\mathbf{A}'_T \mathbf{v}\|_2). \end{aligned} \quad (\text{C1})$$

From Lemma 2 and Definition 1, we have:

$$\|\mathbf{A}'_T \mathbf{A}_T \mathbf{s}_T\|_2 \geq (1 - \delta_K) \|\mathbf{s}_T\|_2 \quad (\text{C2})$$

and:

$$\|\mathbf{A}'_T \mathbf{v}\|_2 \leq \sqrt{1 + \delta_K} \|\mathbf{v}\|_2, \quad (\text{C3})$$

respectively. Using (C2) and (C3), ρ is lower bounded as:

$$\rho \geq \frac{1}{\sqrt{K}} \left[(1 - \delta_K) \|\mathbf{s}_T\|_2 - \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \right], \quad (\text{C4})$$

which is the desired result.

From the definition of η , we have:

$$\sqrt{K} \eta \leq \sqrt{\sum_{j \in I_K} |a'_j \mathbf{y}|^2} = \|\mathbf{A}'_{I_K} \mathbf{y}\|_2 \quad (\text{C5})$$

where $I_K = \arg \max_{|I|=K, I \subset T^c} \|\mathbf{A}'_I \mathbf{y}\|_2$. Using the triangle inequality, we have:

$$\begin{aligned} \|\mathbf{A}'_{I_K} \mathbf{y}\|_2 &= \|\mathbf{A}'_{I_K} (\mathbf{A}_T \mathbf{s}_T + \mathbf{v})\|_2 \\ &\leq \|\mathbf{A}'_{I_K} \mathbf{A}_T \mathbf{s}_T\|_2 + \|\mathbf{A}'_{I_K} \mathbf{v}\|_2. \end{aligned} \quad (\text{C6})$$

Since I_K and T are disjoint ($I_K \subset T^c$), we further have:

$$\|\mathbf{A}'_{I_K} \mathbf{A}_T \mathbf{s}_T\|_2 \leq \delta_{2K} \|\mathbf{s}_T\|_2 \quad (\text{C7})$$

from Lemma 3 and:

$$\|\mathbf{A}'_{I_k} \mathbf{v}\|_2 \leq \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \quad (\text{C8})$$

from Definition 1. Using (C7) and (C8), we have:

$$\|\mathbf{A}'_{I_k} \mathbf{y}\|_2 \leq \delta_{2K} \|\mathbf{s}_T\|_2 + \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \quad (\text{C9})$$

and since $\|\mathbf{A}'_{I_k} \mathbf{y}\|_2 \geq \sqrt{K} \eta$, we have:

$$\eta \leq \frac{1}{\sqrt{K}} \left[\delta_{2K} \|\mathbf{s}_T\|_2 + \sqrt{1 + \delta_K} \|\mathbf{v}\|_2 \right], \quad (\text{C10})$$

which is the desired result.

Appendix D. Proof of Lemma 9

Since $Y_l \subset T$ and $\beta^l = \max_{j \in T \setminus Y_l} |a'_j \mathbf{r}_{Y_l}|$, we have:

$$\begin{aligned} \max_{j \in T \setminus Y_l} |a'_j \mathbf{r}_{Y_l}| &= \|\mathbf{A}'_{T \setminus Y_l} \mathbf{r}_{Y_l}\|_\infty \\ &\geq \frac{1}{\sqrt{|T \setminus Y_l|}} \|\mathbf{A}'_{T \setminus Y_l} \mathbf{P}_{Y_l}^\perp \mathbf{y}\|_2 \\ &= \frac{1}{\sqrt{|T \setminus Y_l|}} \|\mathbf{A}'_{T \setminus Y_l} \mathbf{P}_{Y_l}^\perp (\mathbf{A}_T \mathbf{s}_T + \mathbf{v})\|_2 \\ &= \frac{1}{\sqrt{K-i-l}} \|\mathbf{A}'_T \mathbf{P}_{Y_l}^\perp \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l} + \mathbf{A}'_{T \setminus Y_l} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\geq \frac{1}{\sqrt{K-i-l}} [\|\mathbf{A}'_T \mathbf{P}_{Y_l}^\perp \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l}\|_2 - \|\mathbf{A}'_{T \setminus Y_l} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2] \\ &\geq \frac{1}{\sqrt{K-i-l}} [\|\mathbf{A}'_T \mathbf{A}_T \bar{\mathbf{s}}_{T \cup Y_l}\|_2 - \|\mathbf{A}'_{T \setminus Y_l} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2] \end{aligned} \quad (\text{D1})$$

where (D2) is from Definition 2. From (B2) in Appendix B and Lemma 4, we have:

$$\|\mathbf{A}'_T \mathbf{A}_T \bar{\mathbf{s}}_{T \cup Y_l}\|_2 > (1 - \delta_K) \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 \quad (\text{D2})$$

and:

$$\begin{aligned} \|\mathbf{A}'_{T \setminus Y_l} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 &\leq \|\mathbf{A}'_{T \setminus Y_l}\|_2 \|\mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\leq \sqrt{1 + \delta_{|T \setminus Y_l|}} \|\mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\leq \sqrt{1 + \delta_{|T \setminus Y_l|}} \|\mathbf{v}\|_2 \\ &\leq \sqrt{1 + \delta_{K-i-l}} \|\mathbf{v}\|_2, \end{aligned} \quad (\text{D3})$$

respectively. From (D2), (D2) and (D4), we have:

$$\begin{aligned} \beta^l &> \frac{1}{\sqrt{K-i-l}} \left[\|\mathbf{A}'_T \mathbf{A}_T \bar{\mathbf{s}}_{T \cup Y_l}\|_2 - \|\mathbf{A}'_{T \setminus Y_l} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \right] \\ &> \frac{1}{\sqrt{K-i-l}} \left[(1 - \delta_K) \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 - \sqrt{1 + \delta_{K-i-l}} \|\mathbf{v}\|_2 \right] \\ &> \frac{1}{\sqrt{K-1}} \left[(1 - \delta_K) \|\bar{\mathbf{s}}_{T \cup Y}\|_2 - \sqrt{1 + \delta_{K-1}} \|\mathbf{v}\|_2 \right]. \end{aligned} \quad (\text{D4})$$

Next, from the definition of $\alpha^l = \min_{j \in D_K} |a'_j \mathbf{r}_{Y_l}|$, we have:

$$\sqrt{|D_K|} \alpha^l \leq \sqrt{\sum_{j \in D_K} |a'_j \mathbf{r}_{Y_l}|^2} = \|\mathbf{A}'_{D_K} \mathbf{r}_{Y_l}\|_2 \quad (\text{D5})$$

where $D_K = \arg \max_{|D|=K-i, D \subset \Omega \setminus T'} \|\mathbf{A}'_D \mathbf{r}_{Y_l}\|_2$. Using the triangle inequality, (D5) can be rewritten as:

$$\begin{aligned} \|\mathbf{A}'_{D_K} \mathbf{r}_{Y_l}\|_2 &= \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{y}\|_2 \\ &= \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp (\mathbf{A}_T \mathbf{s}_T + \mathbf{v})\|_2 \\ &= \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l} + \mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\leq \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{A}_{T \setminus Y_l} \mathbf{s}_{T \setminus Y_l}\|_2 + \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &= \|\mathbf{A}'_{D_K} \mathbf{A}_{T \cup Y_l} \bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2. \end{aligned} \quad (\text{D6})$$

where (D7) is from Definition 2. From (B6) in Appendix B and Lemma 4, (D7) is upper bounded as:

$$\begin{aligned} \|\mathbf{A}'_{D_K} \mathbf{r}_{Y_l}\|_2 &\leq \|\mathbf{A}'_{D_K} \mathbf{A}_{T \cup Y_l} \bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\leq \delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \|\mathbf{A}'_{D_K} \mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\leq \delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \|\mathbf{A}'_{D_K}\|_2 \|\mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\leq \delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \sqrt{1 + \delta_{|D_K|}} \|\mathbf{P}_{Y_l}^\perp \mathbf{v}\|_2 \\ &\leq \delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \sqrt{1 + \delta_{|D|}} \|\mathbf{v}\|_2 \\ &\leq \delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \sqrt{1 + \delta_{K-i}} \|\mathbf{v}\|_2. \end{aligned} \quad (\text{D7})$$

Then by combining (D5) and (D8), we get the desired result as:

$$\begin{aligned} \alpha^l &\leq \frac{1}{\sqrt{|D_K|}} [\delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \sqrt{1 + \delta_{K-i}} \|\mathbf{v}\|_2] \\ &< \frac{1}{\sqrt{K-i}} [\delta_{2K-i} \|\bar{\mathbf{s}}_{T \cup Y_l}\|_2 + \sqrt{1 + \delta_{K-i}} \|\mathbf{v}\|_2]. \end{aligned} \quad (\text{D8})$$

Since $1 \leq i \leq K-1$, we obtain the desired result.

Appendix E. Proof of Theorem 6

If the support has the minimum residual in magnitude among all possible candidates with cardinality K , then:

$$\|\mathbf{r}_T\|_2 < \|\mathbf{r}_{\Lambda_K}\|_2$$

for any $\Lambda_K \neq T$. Note that the above inequality always holds if the upper bound of $\|\mathbf{r}_T\|_2$ is smaller than the lower bound of $\|\mathbf{r}_{\Lambda_K}\|_2$. In this regard, we first obtain the upper bound of $\|\mathbf{r}_T\|_2$ as:

$$\begin{aligned} \|\mathbf{r}_T\|_2 &= \|\mathbf{P}_T^\perp \mathbf{y}\|_2 = \|\mathbf{P}_T^\perp (\mathbf{A}_T \mathbf{s}_T + \mathbf{v})\|_2 \\ &= \|\mathbf{P}_T^\perp \mathbf{v}\|_2 \leq \|\mathbf{v}\|_2 \end{aligned} \quad (\text{E1})$$

where (E1) is because $\mathbf{P}_T^\perp \mathbf{A}_T \mathbf{x}_T = \mathbf{0}$.

Next, we obtain the lower bound of $\|\mathbf{r}_{\Lambda_K}\|_2$ when $\Lambda_K \neq T$. For any $\Lambda_K \neq T$, we have:

$$\begin{aligned} \|\mathbf{r}_{\Lambda_K}\|_2 &= \|\mathbf{P}_{\Lambda_K}^\perp \mathbf{y}\|_2 = \|\mathbf{P}_{\Lambda_K}^\perp (\mathbf{A} \mathbf{s} + \mathbf{v})\|_2 \\ &= \|\mathbf{P}_{\Lambda_K}^\perp (\mathbf{A}_{T \setminus \Lambda_K} \mathbf{s}_{T \setminus \Lambda_K} + \mathbf{v})\|_2 \\ &= \|\mathbf{A}_{T \cup \Lambda_K} \bar{\mathbf{s}}_{T \cup \Lambda_K}\|_2 - \|\mathbf{P}_{\Lambda_K}^\perp \mathbf{v}\|_2 \end{aligned} \quad (\text{E2})$$

where (E3) is from the triangle inequality and Definition 2. Using the result from Lemma 7, (E3) is lower bounded as:

$$\begin{aligned} \|\mathbf{r}_{\Lambda_K}\|_2 &= \|\mathbf{A}_{T \cup \Lambda_K} \bar{\mathbf{s}}_{T \cup \Lambda_K}\|_2 - \|\mathbf{P}_{\Lambda_K}^\perp \mathbf{v}\|_2 \\ &\geq \sqrt{1 - \delta_{|T \cup \Lambda_K|}} \|\bar{\mathbf{s}}_{T \cup \Lambda_K}\|_2 - \|\mathbf{P}_{\Lambda_K}^\perp \mathbf{v}\|_2 \end{aligned} \quad (\text{E3})$$

$$\geq \sqrt{1 - \delta_{|T \cup \Lambda_K|}} \|\bar{\mathbf{s}}_{T \cup \Lambda_K}\|_2 - \|\mathbf{v}\|_2 \quad (\text{E4})$$

$$\geq \sqrt{1 - \delta_{2K}} \|\bar{\mathbf{s}}_{T \cup \Lambda_K}\|_2 - \|\mathbf{v}\|_2 \quad (\text{E5})$$

where (E3) is from Definition 1, (E4) is because $\|\mathbf{P}_{\Lambda_K}^\perp \mathbf{v}\|_2 \leq \|\mathbf{v}\|_2$ and (E5) is from Lemma 1. From (E1) and (E5), we have:

$$\|\bar{\mathbf{s}}_{T \cup \Lambda_K}\|_2 \geq \frac{2\|\mathbf{v}\|_2}{\sqrt{1 - \delta_{2K}}} \quad (\text{E6})$$

Using the result from Lemma 7, we additionally have:

$$\|\mathbf{s}_{T \setminus \Lambda_K}\|_2 \geq \frac{\sqrt{2}(1 + \delta_K)\|\mathbf{v}\|_2}{\sqrt{(1 - \delta_{2K})(1 + \delta_K^2)}} \quad (\text{E7})$$

and since $\|\mathbf{s}_{T \setminus \Lambda_K}\|_2 \geq \min_{j \in T} |s_j|$, we have the desired result.

References

- Birnbaum, Y.; Wilson, J.M.; Fiol, M.; Luna, A.B.; Eskola, M.; Nikus, K. ECG diagnosis and classification of acute coronary syndromes. *Ann. Noninvasive Electrocardiol.* **2014**, *19*, 4–14.
- Singh, B.; Singh, D.; Jaryal, A.; Deepak, K. Ectopic beats in approximate entropy and sample entropy-based HRV assessment. *Int. J. Syst. Sci.* **2012**, *43*, 884–893.
- Steg, P.G.; James, S.K.; Atar, D.; Badano, L.P.; Blömstrom-Lundqvist, C.; Borger, M.A.; Di Mario, C.; Dickstein, K.; Ducrocq, G.; Fernandez-Aviles, F.; et al. ESC guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *Eur. Heart J.* **2012**, *33*, 2569–2619.
- Manina, G.; Agnelli, G.; Becattini, C.; Zingarini, G.; Paciaroni, M. 96 hours ECG monitoring for patients with ischemic cryptogenic stroke or transient ischaemic attack. *Intern. Emerg. Med.* **2014**, *9*, 65–67.
- Weder, M.; Hegemann, D.; Amberg, M.; Hess, M.; Boesel, L.; Abächerli, R.; Meyer, V.; Rossi, R. Embroidered electrode with silver/titanium coating for long-term ECG monitoring. *Sensors* **2015**, *15*, 1750–1759.
- Chen, Y.H.; de Beeck, M.; Vanderheyden, L.; Carrette, E.; Mihajlović, V.; Vanstreels, K.; Grundlehner, B.; Gadeyne, S.; Boon, P.; van Hoof, C. Soft, comfortable polymer dry electrodes for high quality ECG and EEG recording. *Sensors* **2014**, *14*, 23758–23780.
- Wang, C.; Lu, W.; Narayanan, M.R.; Redmond, S.J.; Lovell, N.H. Low-power technologies for wearable telecare and telehealth systems: A review. *Biomed. Eng. Lett.* **2015**, *5*, 1–9.
- Candes, E.J.; Tao, T. Decoding by linear programming. *IEEE Trans. Inf. Theory* **2005**, *51*, 4203–4215.
- Candes, E.J.; Romberg, J.; Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theory* **2006**, *52*, 489–509.
- Chen, S.S.; Donoho, D.L.; Saunders, M.A. Atomic Decomposition by Basis Pursuit. *SIAM J. Sci. Comput.* **1998**, *20*, 33–61.
- Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288.
- Candes, E.; Tao, T. The Dantzig Selector: Statistical Estimation When p Is Much Larger than n. *Ann. Stat.* **2007**, *35*, 2313–2351.
- Mamaghanian, H.; Khaled, N.; Atienza, D.; Vandergheynst, P. Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes. *IEEE Trans. Biomed. Eng.* **2011**, *58*, 2456–2466.

14. Craven, D.; McGinley, B.; Kilmartin, L.; Glavin, M.; Jones, E. Compressed sensing for bioelectric signals: A review. *IEEE J. Biomed. Health Inf.* **2015**, *19*, 529–540.
15. Zhang, Z.; Jung, T.P.; Makeig, S.; Rao, B.D. Compressed sensing for energy-efficient wireless telemonitoring of noninvasive fetal ECG via block sparse bayesian learning. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 300–309.
16. Cotter, S.F.; Rao, B.D. Application of tree-based searches to matching pursuit. In Proceedings of the IEEE International Conference Acoustics, Speech, and Signal Process, Salt Lake City, UT, USA, 7–11 May 2001; pp. 3933–3936.
17. Karabulut, G.Z.; Moura, L.; Panario, D.; Yongacoglu, A. Integrating flexible tree searches to orthogonal matching pursuit algorithm. *IEE Proc. Vis. Image Signal Process.* **2006**, *153*, 538–548.
18. Schniter, P.; Potter, L.C.; Ziniel, J. Fast bayesian matching pursuit. In Proceedings of the Information Theory and Applications Workshop, San Diego, CA, USA, 27 January–1 February 2008; pp. 326–333.
19. Kwon, S.; Wang, J.; Shim, B. Multipath matching pursuit. *IEEE Trans. Inf. Theory* **2013**, *56*, 4867–4878.
20. Karahanoglu, N.B.; Erdogan, H. A* orthogonal matching pursuit: Best-first search for compressed sensing signal recovery. *Digit. Signal Process.* **2012**, *22*, 555–568.
21. Cai, T.; Wang, L. Orthogonal matching pursuit for sparse signal recovery with noise. *IEEE Trans. Inf. Theory* **2011**, *57*, 4680–4688.
22. Needell, D.; Tropp, J.A. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Commun. ACM* **2010**, *53*, 93–100.
23. Dai, W.; Milenkovic, O. Subspace Pursuit for Compressive Sensing Signal Reconstruction. *IEEE Trans. Inf. Theory* **2009**, *55*, 2230–2249.
24. Karahanoglu, N.; Erdogan, H. Improving A*OMP: Theoretical and empirical analyses with a novel dynamic cost model. *J. Signal Process.* **2016**, *118*, 62–74.
25. Dechter, R.; Pearl, J. Generalized best-first search strategies and the optimality of A*. *J. ACM* **1985**, *32*, 505–536.
26. Jelinek, F. *Statistical Methods for Speech Recognition*; MIT Press: Cambridge, MA, USA, 1998.
27. Fincke, U.; Pohst, M. Improved methods for calculating vectors of short length in a lattice, including a complexity analysis. *Math. Comput.* **1985**, *44*, 463–471.
28. Hochwald, B.M.; Ten Brink, S. Achieving near-capacity on a multiple-antenna channel. *IEEE Trans. Commun.* **2003**, *51*, 389–399.
29. Shim, B.; Kang, I. Sphere Decoding With a Probabilistic Tree Pruning. *IEEE Trans. Signal Process.* **2008**, *56*, 4867–4878.
30. Lee, J.; Shim, B.; Kang, I. Soft-Input Soft-Output List Sphere Detection with a Probabilistic Radius Tightening. *IEEE Trans. Wirel. Commun.* **2012**, *11*, 2848–2857.
31. Lee, J.; Kwon, S.; Shim, B. A greedy search algorithm with tree pruning for sparse signal recovery. In Proceedings of the 2014 IEEE International Symposium on Information Theory (ISIT) 2014, Honolulu, HI, USA, 29 June–4 July 2014; pp. 1847–1851.
32. Chen, W.; Rodrigues, M.R.D.; Wassell, I.J. Projection Design for Statistical Compressive Sensing: A Tight Frame Based Approach. *IEEE Trans. Signal Process.* **2013**, *61*, 2016–2029.
33. Baraniuk, R.G.; Davenport, M.A.; DeVore, R.; Wakin, M.B. A Simple Proof of the Restricted Isometry Property for Random Matrices. *Constr. Approx.* **2008**, *28*, 253–263.
34. Candes, E.J. The restricted isometry property and its implications for compressed sensing. *C. R. Math.* **2008**, *346*, 589–592.
35. Satpathi, S.; Lochan Das, R.; Chakraborty, M. Improving the Bound on the RIP Constant in Generalized Orthogonal Matching Pursuit. *IEEE Signal Process. Lett.* **2013**, *20*, 1074–1077.
36. Taddei, A.; Distanto, G.; Emdin, M.; Pisani, P.; Moody, G.B.; Zeelenberg, C.; Marchesi, C. The European ST-T database: Standard for evaluating systems for the analysis of ST-T changes in ambulatory electrocardiography. *Eur. Heart J.* **1992**, *13*, 1164–1172.

