Signal Processing 140 (2017) 12–31

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Diffusion wavelet basis algorithm for sparse representation of sensory data in *WSNs*

Cuicui Lv[a,*], Qiang Wang[a,*], Wenjie Yan[b], Yi Shen[a]

[a] Department of Control Science and Engineering, Harbin Institute of Technology, Harbin, China
[b] School of Computer Science and Engineering, Hebei University of Technology, Tianjin, China

ARTICLE INFO

Article history:
Received 4 October 2016
Revised 21 March 2017
Accepted 1 May 2017
Available online 1 May 2017

Keywords:
Compressive sensing
Wireless sensor networks
Diffusion wavelet
Numerical sparsity

ABSTRACT

Sparsity is a precondition for compressive data gathering in Wireless Sensor Networks. To solve the sparse representation problem of sensory data in temperature field, the paper puts forward a diffusion wavelet basis algorithm, which mainly includes the construction of diffusion operator and the filter of diffusion wavelet bases. The construction of diffusion operator takes spatial correlation between sensory data and the communication radius into account. Diffusion wavelet bases are generated from diffusion wavelet trees with different decomposition levels. The algorithm introduces numerical sparsity to evaluate the sparsity of approximately sparse data. Based on the measure of sparsity, diffusion wavelet basis with the better performance is extracted as the sparsifying basis. Synthetic data experiment and real data experiment are conducted to validate the performance of the sparsifying basis. The results of numerical experiments show that sensory data are approximately sparse, and the numerical sparsity in the proposed diffusion wavelet basis is less than that in the other types of the sparsifying bases. The recovery of sensory data considers the noiseless and noise cases. When BP and BPDN are used for data recovery, and the measurement matrix is sparse binary matrix, we observe that the proposed diffusion wavelet basis decreases relative error.

1. Introduction

Compressive sensing (CS) [1] provides a data gathering method for Wireless Sensor Networks (WSNs) [2], and can use the fewer linear measurements to recover these data. Consider a WSN with N sensor nodes. Each sensor node has a random ID that ranges from 1 to N, and sensor node with ID $i(1 \le i \le N)$ is denoted by s_i. The data readings of sensor nodes are a column vector $x = [x_1, x_2, \ldots, x_i, \ldots, x_N]^T$, where x_i is data reading of s_i. If x has a sparse representation in a basis $\Psi \in \mathbb{R}^{N\times N}$, it is written as

$$x = \Psi\Theta \tag{1}$$

where $\Theta \in \mathbb{R}^{N\times 1}$ is the coefficient vector of x, and $\|\Theta\|_0$ is the number of nonzero coefficients. When the value of $\|\Theta\|_0$ is equal to $k(k \ll N)$, x is said to be k-sparse in the basis Ψ. In practical, many signals are approximately sparse in an appropriate basis. If x only has $k(k \ll N)$ large coefficients in a basis, and the coefficients sorted in decreasing order of magnitude decrease quickly, x is approximately k-sparse.

Provided that x is sparse or approximately sparse, the linear measurement vector y, $y = [y_1, y_2, \ldots, y_i, \ldots, y_M]^T$, can be acquired by Eq. (2) in the noiseless case.

$$y = \Phi x = \Phi\Psi\Theta \tag{2}$$

where $\Phi \in \mathbb{R}^{M\times N}(k < M \ll N)$ is the measurement matrix. Given the column vector x and the measurement vector y, it is natural to attempt to recover x by solving the l_0-norm minimization problem: $\hat{\Theta}=argmin\|\Theta\|_0 s.t. y = \Phi\Psi\Theta$. This problem has been proved to be NP-hard. It is translated into the l_1-norm minimization problem: $\hat{\Theta}=argmin\|\Theta\|_1 s.t. y = \Phi\Psi\Theta$. x can be recovered from y by solving the l_1-norm minimization problem. The reconstruction algorithms contain Basis Pursuit(BP) [3], IT(Iterative thresholding) [4] and etc.

Due to the additive nature of radio waves in WSNs, the communication receiver circuitry of sink node usually generates the noise [5]. The noisy measurement can be expressed as

$$y = \Phi x + z \tag{3}$$

where z represents the measurement noise (additive uncorrelated i.i.d. Gaussian noise with the variance σ_n^2). Like the problem in Eq. (2), the problem in Eq. (3) can be solved by the noisy version BPDN of BP.

In compressive data gathering [6], two factors to be considered are the sparsifying basis and the measurement matrix. The mea-

* Corresponding authors.
E-mail addresses: wangqiang@hit.edu.cn, lvcuicuixuchuan@163.com (Q. Wang).

http://dx.doi.org/10.1016/j.sigpro.2017.05.001
0165-1684/

surement matrix includes dense random matrices and sparse random matrices [7]. Dense random matrices, such as independent and identically distributed Gaussian random matrix and Bernoulli matrix, satisfy RIP with overwhelming probability. However, the class of matrices have high computational complexity, which interferes with the hardware implementation [8]. Compared with them, sparse random matrices are sparser. For this kind of matrix, Wang et al. showed the performance of data recovery was comparable to the optimal k-term approximation [9]. Sparse binary matrix is in that category. Through extensive experiments on both synthetic and real data sets, Li and Qi showed the recovery accuracy of sensory data by sparse binary matrix outperformed existing sparse random matrices [10]. So sparse binary matrix is used for compressive data gathering and data recovery in the paper.

Sparse representation of sensory data is first to find the sparsifying basis of sensory data, and determine the coefficient vector based on Eq. (1). The goal of the paper is to construct the sparsifying basis of temperature data in the monitored filed. Because WSNs can be described by weighted graphs, and multiscale signal processing technique based on diffusion wavelet can be lifted to the graph for compression and approximation [11], diffusion wavelet is more attractive in this problem [12]. Diffusion wavelet applies a diffusion operator on the graph to a space of test functions at the finest scale, and uses its dyadic powers to induce a multiresolution analysis. Diffusion wavelet provides a paradigm of wavelet decomposition for multiscale analysis on graphs [13]. Xiang et al. [12] first adopted diffusion wavelet to design the sparsifying basis of sensory data in WSNs, and the experimental results showed that diffusion wavelet had the better performance in the sparse representation of sensory data. Inspired by the construction of diffusion operator [12], we study the effect factors of diffusion wavelet, and propose a diffusion wavelet basis algorithm. Based on the algorithm, we construct the sparsifying basis of sensory data, and use measure of sparsity to verify its performance. The main contributions of the paper are as follows.

(1) In the paper, the temperature filed monitored by a WSN is mapped into a graph. Based on graph theory and the deployment of sensor nodes, we put forward a method to measure the spatial correlation of sensory data. This method uses the spatial correlation between sensory data as the weight of an edge on the graph, and generates the diffusion operator restricted by the communication radius. It has been proved by theoretical analyses and the experiment that high powers of the constructed diffusion operator have low numerical rank.
(2) For approximately sparse temperature data in the sparsifying basis, we propose a diffusion wavelet basis algorithm to make temperature data sparser. In the algorithm, we introduce the numerical sparsity to evaluate the proportion of effectively large coefficients, and consider the decomposition level of diffusion wavelets for the optimization of the algorithm.
(3) The synthetic data experiment and real data experiment are carried out to verify the performance of the proposed algorithm in the noiseless and noise cases. In addition, we compare the proposed diffusion wavelet basis with other sparsifying bases in terms of numerical sparsity and data recovery. From the experimental results, we can see that numerical sparsity is minimal in the proposed diffusion wavelet basis, and relative error of data recovery is the less.

The rest of the paper is organized as follows. In Section 2, we construct the diffusion operator based on WSNs described by the weighted graphs. In Section 3, low numerical rank of high powers of diffusion operator is proved in theory. Given that premise, we propose the diffusion wavelet basis algorithm, and regard the basis as the sparsifying basis of sensory data. To verify the algorithm, synthetic data experiment and real data experiment are conducted in Section 4. Section 5 concludes the paper.

2. Diffusion wavelet based sparsity model

In the section, we introduce the construction of diffusion operator, which is the major part of sparsity model, and describe two estimation methods of the sparsifying basis: Gini Index(GI) and numerical sparsity s.

2.1. Construction of diffusion operator

Suppose sensor nodes are uniformly deployed to monitor a spatially varying temperature field. The field is regarded as an image with a degree of spatial correlation and sensory data correspond to the pixel values in the image [5]. In this case, the temperature field being monitored can be mapped to a graph on two-dimensional plane.

Network topology of a WSN is abstracted to a weighted undirected graph $G(V, E, \omega)$. It consists of a set of vertices V, a set of edges E and a weight function ω. V corresponds to nodes in WSNs, that is, $V = \{s_0, s_1, \ldots, s_i, \ldots, s_N\}$, where s_0 is sink node. For a general graph $G(V, E, \omega)$, if s_i and s_j can communicate with each other, an edge e_{ij} exists, and the set of all edges is expressed as $E = \{e_{ij} | 1 \le i < j \le N\}$. The weight function ω satisfies $\omega_{ij} = \omega_{ji}$ and $\omega_{ij} \ge 0$. The weighted adjacency matrix of $G(V, E, \omega)$ is Ω, $\Omega = [\omega_{ij}]$. For sensory data defined on $G(V, E, \omega)$, diffusion wavelet is introduced to construct an orthogonal basis for functions supported on the graph. The construction process of diffusion wavelet produces a basis tailored to $G(V, E, \omega)$ by analyzing the eigenvectors of diffusion operator O derived from weighted adjacency matrix. Based on the method in [12], the paper defines the weight ω_{ij} of e_{ij} as a function of the spatial correlation between sensory data of s_i and s_j.

$$\omega_{ij} = \begin{cases} d_{ij}^{-1} & i \ne j, d_{ij} \leqslant R \\ 0 & i \ne j, d_{ij} > R \\ 1 & i = j \end{cases} \tag{4}$$

Under the circumstances, the constructed $G(V, E, \omega)$ has the different meaning. $G(V, E, \omega)$ is used to denote the spatial correlation between sensory data of sensor nodes, and V not only includes sensor nodes but also includes their sensory data. Unlike the method in [12], the paper considers the communication radius R in weighted adjacency matrix Ω. The communication radius R called in the communications networks is no longer to measure whether two sensor nodes can communicate with each other, and it is used to bound the degree of spatial correlation in the paper. In Eq. (4), d_{ij} is Euclidean distance between s_i and s_j. The smaller the value of $d_{ij}(d_{ij} \le R)$ is, the larger the spatial correlation is. With the increase of d_{ij}, the spatial correlation between sensory data of s_i and s_j becomes smaller and smaller. When the value of d_{ij} is greater than R, spatial correlation between sensory data of s_i and s_j is set as zero. Compared with [12], the definition in Eq. (4) weakens the spatial correlation of sensory data from two sensor nodes, which are far away from each other. The value of R is determined based on measure of sparsity and the geometric structure of sensory data, which will be introduced in Section 3.

In graph theory, degree matrix is combined with adjacency matrix to construct Laplacian matrix of a graph. Each entry in the degree matrix is the number of edges attached to the vertex. For the weighted undirected graph $G(V, E, \omega)$, degree matrix D considers spatial correlation between sensory data. Each entry in degree matrix D is D_{ij}, defined in Eq. (5). Normalized Laplacian matrix of $G(V, E, \omega)$ is Λ, $\Lambda = [\lambda_{ij}]$. Different from the early work [14], the definition of degree matrix D in Eq. (5) can guarantee the spectral range

of Λ [15]. Diffusion operator O is $I-\Lambda$.

$$D_{ij}=\begin{cases}\sum_{p=1}^{N}\omega_{ip} & i=j\\ 0 & i\neq j\end{cases} \tag{5}$$

$$\lambda_{ij}=\begin{cases}1-\frac{1}{D_{ij}} & i=j\\ -\frac{\omega_{ij}}{\sqrt{D_{ii}\cdot D_{jj}}} & i\neq j\end{cases} \tag{6}$$

2.2. Measure of sparsity

The sparsity k of x in the sparsifying basis Ψ is usually measured by l_0 norm, $k=\|\Theta\|_0$ *s.t.* $x=\Psi\Theta$. For approximately sparse data, only fewer large coefficients concentrate a large proportion of energy. So we introduce Gini Index(*GI*) [16,17] and numerical sparsity [18].

Definition 1. Gini Index(*GI*): If the coefficient vector of x in the sparsifying basis Ψ is $\Theta=[\theta_1,\theta_2,\ldots,\theta_N]^\tau$, they are reordered from smallest to largest, $\left|\theta_{[1]}\right|\leqslant\left|\theta_{[2]}\right|,\ldots,\leqslant\left|\theta_{[N]}\right|$, where $[1],[2],\ldots,[N]$ are the new indices after reordering, then *GI* is

$$GI=1-2\sum_{f=1}^{N}\frac{\left|\theta_{[f]}\right|}{\|\Theta\|_1}\left(\frac{N-f+\frac{1}{2}}{N}\right) \tag{7}$$

Gini Index(*GI*) exhibits relative distribution of energy among the coefficients. Seen from the definition in Eq. (7), *GI* is normalized, and the value ranges from zero to one. The closer to one the value of *GI* is, the more concentrated the distribution of energy among these coefficients is. Further, the basis, where *GI* of the coefficients is the larger, can make x sparser. Considering the instability of $\|\Theta\|_0$ in practical application, numerical sparsity was proposed as an alternative measure of sparsity [18].

Definition 2. Numerical Sparsity: If the coefficient vector of x in the sparsifying basis Ψ is $\Theta\in\mathbb{R}^{N\times1}$, numerical sparsity s of x is

$$s=\frac{\|\Theta\|_1^2}{\|\Theta\|_2^2} \tag{8}$$

Numerical sparsity uses the ratio $\|\Theta\|_1^2/\|\Theta\|_2^2$ to replace the l_0 norm, and it is a soft version of l_0 norm. If x has k large coefficients and $N-k$ small coefficients in the sparsifying basis Ψ, then $s\approx k$. For any nonzero coefficient vector Θ, the l_1-norm and l_2-norm satisfy the inequality $\|\Theta\|_2\leqslant\|\Theta\|_1\leqslant\sqrt{N}\|\Theta\|_2$, and the value of s is [1, N]. Moreover, s has a sharp lower bound on $\|\Theta\|_0$, $s\leq\|\Theta\|_0$ [18]. Compared with *GI*, numerical sparsity provides an explicit estimation method of sparsity.

3. Diffusion wavelet basis algorithm

3.1. Description and analysis of algorithm

If high powers of diffusion operator have low numerical rank, they can be compressed by projecting on an appropriate subspace. Based on the assumption, Coifman and Maggioni [11] proposed the construction of diffusion wavelet tree. Enlightened by the approaches in [11,12], we propose a diffusion wavelet basis Algorithm 1.

Theorem 1. *If normalized Laplacian matrix Λ of $G(V, E, \omega)$ is constructed based on Eq. (6), and diffusion operator O is $I-\Lambda$, high powers of diffusion operator O have low numerical rank.*

Proof. For a weighted undirected graph $G(V, E, \omega)$, Λ is a real symmetric matrix. The eigenvalues of Λ are real number and satisfy $0=\lambda_1\leq\lambda_2\leq\cdots\leq\lambda_N=2$ [15]. The eigenvalues of diffusion operator O range from -1 to 1. Because diffusion operator O is also a real symmetric matrix, the scope of singular values is [0, 1], $1=\sigma_1\geq\sigma_2\geq\cdots\geq\sigma_N=0$.

Golub et al. [19] gave the definition of numerical rank, and pointed out that diffusion operator O had numerical rank r with parameters(ξ_1, ξ_2, l_2-norm) if and only if $\sigma_r\geqslant\xi_1>\xi_2\geqslant\sigma_{r+1}$. Diffusion operator O is used as a dilation operator by its dyadic powers. Relative to Λ, the eigenvalues of O^{2^j} are $(1-\lambda_1)^{2^j}$, $(1-\lambda_2)^{2^j},\ldots,(1-\lambda_N)^{2^j}$. For real symmetric matrix O^{2^j}, singular values are absolute values of the corresponding eigenvalues, and the range of singular values is [0, 1]. With the increase of powers, singular values become smaller and smaller. When the values of ξ_1 and ξ_2 are fixed, the value of numerical rank decreases. So the high powers of the constructed diffusion operator O have low numerical rank. □

Algorithm 1 Diffusion wavelet basis algorithm.

Input:
N, sensory data x, $G(V,E,\omega)$, the decomposition level γ of diffusion wavelet, and precision ϵ.

Output:
diffusion wavelet basis Ψ

The maximum value of Euclidean distances between any two nodes is d_{max}.
if the network is connected when r is $0.1\times d_{max}$ **then**
  R ranges from $0.1\times d_{max}$ to d_{max} by the step size 0.1.
else
  In the set $\{0.1\times d_{max}, 0.2\times d_{max},\ldots,1.0\times d_{max}\}$, the minimum value that makes the network connected is defined as R_{min}, and the coefficient (0.1,0.2,... ,1.0) in the set is marked as α.
  R ranges from R_{min} to d_{max} by the step size $0.1\times d_{max}$.
end if
for each value of R **do**
  Calculate each entry ω_{ij} in Ω, D_{ij} in D, and λ_{ij}.
  Normalized Laplacian matrix is $\Lambda=[\lambda_{ij}]$.
  $O=I-\Lambda$.
  Original range space of O is U_0, $U_0=\{\delta_{\mathbf{k}}\}_{\mathbf{k}\in G(V,E,\omega)}$.
  for $J=0$ to γ-1 **do**
    if the columns of O^{2^J} is below or equal to 1 **then**
      Break;
    end if
    $([U_{J+1}]_{U_J},[O^{2^J}]_{U_J}^{U_{J+1}})=SpQR([O^{2^J}]_{U_J}^{U_J},\epsilon)$
    if $J\geqslant1$ **then**
      $[U_{J+1}]_{U_0}=[U_J]_{U_0}[U_{J+1}]_{U_J}$
    end if
    $[V_J]_{U_J}=SpQR(I_{U_J}-[U_{J+1}]_{U_J}[U_{J+1}]_{U_J}^{\tau},\epsilon)$
    $[V_{J+1}]_{U_0}=[U_J]_{U_0}[V_{J+1}]_{U_J}$
    $\left[O^{2^{J+1}}\right]_{U_{J+1}}^{U_{J+1}}=([O^{2^J}]_{U_J}^{U_{J+1}}[U_{J+1}]_{U_J})^2$
  end for
  for $J=1$ to γ-1 **do**
    Extract the orthogonal bases of extended scaling functions $[V_0]_{U_0}$, $[V_1]_{U_0}$, ... , $[V_J]_{U_0}$, and extended wavelet functions $[U_J]_{U_0}$. Diffusion wavelet basis is the concatenation of their orthogonal bases.
    Calculate the sparsity (numerical sparsity) of sensory data x at the decomposition level J.
  end for
end for
Diffusion wavelet basis that makes the sparsity (numerical sparsity) of x minimal is considered as Ψ.

In the proposed algorithm, we use $[U_{J+1}]_{U_J}$ to represent U_{J+1} with respect to the basis U_J. $[O^{2^J}]_{U_J}^{U_{J+1}}$ is a representation of an ε-approximation of dyadic powers O^{2^J} with respect to U_J in the domain and U_{J+1} in the range. The representation of O^{2^J} onto the basis U_J is encoded in the matrix $[O^{2^J}]_{U_J}^{U_J}$. $[U_{J+1}]_{U_J}^{\tau}$ denotes the transpose of $[U_{J+1}]_{U_J}$.

According to the locations of sensor nodes, the algorithm first calculates the maximum value d_{max} of Euclidean distances among sensor nodes, and determines the minimum value R_{min} to guarantee the network connectivity. From the construction of diffusion operator, we can see that different communication radius R can generate different diffusion operators. For each value of R, the corresponding diffusion operator O is represented on the basis U_0, and its high powers have low numerical rank, which has been proved in Theorem 1. Through step13∼step24, a diffusion wavelet tree is generated. The tree includes the orthogonal bases of diffusion scaling function U_J and the orthogonal bases of wavelets V_J. *SpQR* is a function of computing a sparse *QR* decomposition. Take step17 as an example. *SpQR* decomposes $[O^{2^J}]_{U_J}^{U_J}$ into an orthogonal matrix $[U_{J+1}]_{U_J}$ and an upper triangular matrix $[O^{2^J}]_{U_J}^{U_{J+1}}$, such that $[O^{2^J}]_{U_J}^{U_J} =_{\varepsilon} [U_{J+1}]_{U_J}[O^{2^J}]_{U_J}^{U_{J+1}}$. More specific details about *SpQR* are available in [11]. Calculate the sparsity(numerical sparsity) of sensory data, and repeat the above steps. Based on the sparsity(numerical sparsity), we determine diffusion wavelet basis with the better performance. From the foregoing, the proposed algorithm is to look for the optimal values of α and γ from a finite set. Like greedy algorithm, it is difficult to prove the optimum mathematically, and the optimal values in the paper can be considered to be locally optimal.

The proposed algorithm requires the position information of sensor nodes to construct the diffusion operator. In a real sensor network, sink node gets these information by some localization techniques. Generally, the localization techniques of sensor nodes are categorized as the range-free and range-based. Range-free localization techniques utilize the radio connectivity to map the network topology, and the number of hops between two nodes is used as a distance metric [20]. Range-based localization techniques are to estimate the distance between two nodes, and compute the position of every node in the network. Compared with range-free localization techniques, range-based localization techniques are more accurate [21]. Range-based techniques include angle of arrival (AOA), time difference of arrival (TDOA), and time of arrival (TOA) and received signal strength indication (RSSI) [22]. Such as GreenOrbs, a large-scale sensor network system deployed in a forest, [23] inherited the advantages of two localization techniques, and proposed a Combined and Differentiated Localization approach. When the higher accuracy is required, Global Positioning System (GPS) could even be used in outdoor environments.

3.2. Time complexity of algorithm

The step 1 in the proposed algorithm is the calculation of Euclidean distance and determination of the maximum Euclidean distance. Because the network includes N sensor nodes and one sink node, the time complexity is $O((N+1)^2)$ and $O((N+1)log(N+1))$, respectively. In steps 2–7, Prim algorithm is used to validate the network connectivity, and the time complexity is $O((N+1)^2)$. The construction of normalized Laplacian matrix only relates to sensor nodes, and the time complexity is $O(N^2)$. Steps 13–24 are the generation of diffusion wavelet tree, and the time complexity is $O(N^3)$ in the worst case. The last step is to calculate the sparsity(numerical sparsity) of diffusion wavelet coefficients, and the time complexity is $O(N)$. As a result, the time complexity of the proposed algorithm is $O(N^3)$ in the worst case. Given the decomposition level γ, [12] has the same order of magnitude as the proposed algorithm in the construction of diffusion wavelet tree, and the time complexity is $O(N^3)$ in the worst case. However, the proposed algorithm adds to the steps 1–6, steps 25–28, and it needs to take more time to look for the optimal values of α and γ. Compared with [12], the runtime of the proposed algorithm is longer.

3.3. Sparse representation of sensory data

According to the proposed algorithm, we construct a diffusion wavelet basis Ψ with the better performance, and regard it as the sparsifying basis of x. If the data readings of sensor nodes are a column vector $x = [x_1, x_2, \ldots, x_i, \ldots, x_N]^T$, the coefficient vector Θ, $\Theta = [\theta_1, \theta_2, \cdots, \theta_N]^T$, is

$$\Theta = \Psi^{-1}x \tag{9}$$

The linear representation of x in the sparsifying basis Ψ is

$$x = x_1\theta_1 + x_2\theta_2 + \cdots + x_N\theta_N \tag{10}$$

As mentioned in the proposed algorithm, the numerical sparsity s of x is smaller than N. There are s effectively large coefficients in the coefficient vector, and the linear representation of x in the Ψ is sparse.

4. Experimental results and analysis

4.1. Experiment setting

In the section, we carry out the synthetic data experiment and real data experiment of temperature field to demonstrate the proposed algorithm. Among them, the temperature field in the synthetic data experiment is simulated by peaks function in MATLAB [24,25]. The temperature fields in real data experiment are from four scenarios: National Oceanic and Atmospheric Administration's National Weather Service [26], temperature distribution in USA [27], and Sea Surface Temperature(100 KM Global) [28]. These experiments are realized on the PC with inter(R) Core(TM) i3-3320 CPU@3.30GHz, and the experimental tool is MATLAB. Based on the approach in Section 2, we first construct the diffusion operator, and analyze the performance of numerical rank. Then represent sensory data in the proposed diffusion wavelet basis, and compare the sparsity(numerical sparsity) in different sparsifying bases. Finally we consider the recovery performance of sensory data by BP in the noiseless case and that by BPDN in the noise case. In addition, the proposed diffusion wavelet basis is compared with other sparsifying bases in terms of relative error.

4.2. Analysis of diffusion operator

Suppose N(1024) sensor nodes and one sink node are deployed in the area of 100 m × 100 m, as shown in Fig. 1. Sensor nodes obey the random uniform distribution, and their locations are known. Sink node is responsible for compressive data gathering and it is placed in the center (50,50) of the area. In the network, the maximum value of Euclidean distances between any two nodes is $d_{max}m$. The communication radius R among two nodes is first set as $0.1 \times d_{max}m$ and the network is connected. According to the approach in Section 2, we construct the diffusion operator O with the size $N \times N$. The singular values are sorted in descending order, as denoted in Fig. 2(a). We observe that the scope of singular values is [0, 1], i.e. $\sigma \in [0, 1]$. With the increase of powers of diffusion operator O, the singular values decay quickly, and the value of numerical rank r reduces. When the value of ξ_1 is 10^{-4} and ξ_2 is equal to 9.9999×10^{-4}, the values of r, σ_r and σ_{r+1}

Table 1
The powers of diffusion operator O and the numerical rank r.

The powers	r	$\sigma_r(\times 10^{-4})$	$\sigma_{r+1}(\times 10^{-5})$	r [12]	$\sigma_r(\times 10^{-4})$	$\sigma_{r+1}(\times 10^{-5})$
1	1023	1.5411	5.33321	1016	1.0347	9.1083
2	921	1.0316	9.92702	776	1.0024	9.7053
4	417	1.0033	9.93994	52	1.0322	9.9948
8	72	1.1336	9.72108	4	1.7505	5.6779
16	35	1.0557	5.546816	1	10000	3.4812×10^{-1}
32	20	1.4573	5.203732	1	10000	1.2119×10^{-6}
64	11	1.9242	4.697564	1	10000	2.0253×10^{-10}

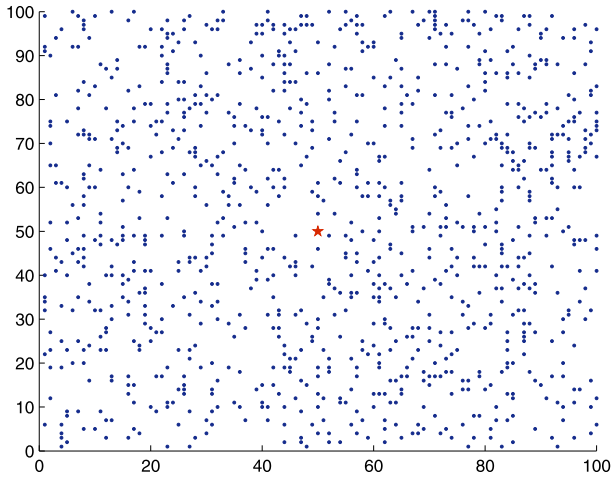


Fig. 1. The distribution of sensor nodes and one sink node.

are listed in Table 1. Compared with the lower powers of diffusion operator O, the higher powers have the lower numerical rank r, which corresponds to Theorem 1. For other values of R, the above steps are repeated and the same conclusion can also be drawn. Besides, Fig. 2(b) and Table 1 also show the diffusion operator O constructed by [12]. We observe that the diffusion operator O in [12] has the smaller numerical rank r.

4.3. Synthetic data experiment

In the experiment, we use peaks function in MATLAB to simulate the temperature field [24,25], as shown in Fig. 3(a). Fig. 3(b) depicts the temperature data of 1024 sensor nodes, and the temperature values are represented in different colors. In the proposed algorithm, the maximum value of the decomposition level γ and the value of ϵ are set as 8 and 10^{-6}, respectively.

4.3.1. Sparsity of sensory data

According to the proposed algorithm, we define $\alpha = R/d_{max}$. The values of α range from 0.1 to 1.0 by the step size 0.1. Because the construction of diffusion operator O is affected by the communication radius R, different values of α will generate different diffusion wavelet bases. For each value of α, the temperature data are represented in the corresponding diffusion wavelet basis. From the experiments, we discover that temperature data are approximately sparse in diffusion wavelet basis and record the numerical sparsity s. Based on these statistical data, Fig. 4 draws the relationship between the numerical sparsity s and α. In Fig. 4, the step size of α is 0.05 for the better observation. In addition, the effect of the decomposition level γ on the numerical sparsity s is also taken into account. For different values of γ, we observe from Fig. 4 that numerical sparsity s has the minimum value when the value of α is 0.1 and the value of γ is 2. If the value of γ is greater than 6, the condition in the step 14 of proposed algorithm is satisfied. In this case, diffusion wavelet basis, corresponding value of γ is greater than 6, is the same as that corresponding value of γ is 6, and the value of s doesn't change with the increase of γ. To verify the performance of numerical sparsity s, the coefficients of temperature data represented in diffusion wavelet basis are in descending order by the magnitude, as denoted in Fig. 5, where the value of α is 0.1. When the value of γ is 2, the magnitude of diffusion wavelet coefficients has a more rapid decay, which corresponds to the smaller value of s in Fig. 4. From the above analysis, one can draw a conclusion that the minimum value of s is obtained when the value of α is 0.1 and the value of γ is 2. In different diffusion wavelet bases, the diffusion wavelet basis corresponding to the minimum value of s is better. When GI is used as the evaluation index, the maximum value of GI is 0.8924. Here the value of α is 0.1 and the value of γ is 6. To compare the difference between numerical sparsity s and GI, Fig. 6 shows the diffusion wavelet coefficients sorted in decreasing order by the magnitude. Meanwhile, the values of GI and s are summarized in Table 2. When the value of α is 0.1, the value of GI corresponding decomposition level 2(i.e.$\gamma = 2$) is less than that corresponding decomposition level 6(i.e.$\gamma = 6$).

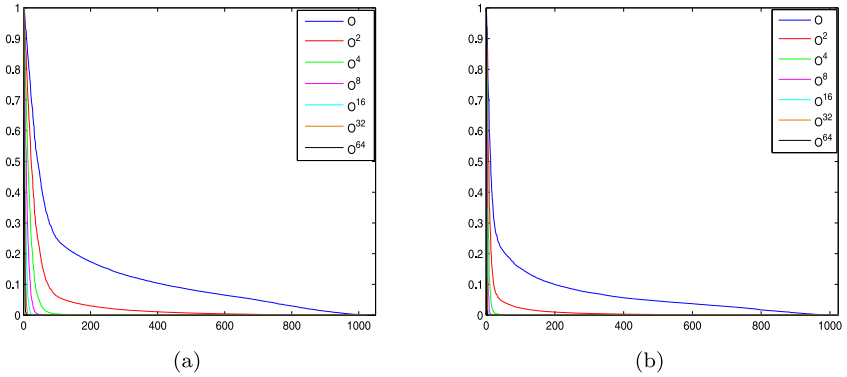
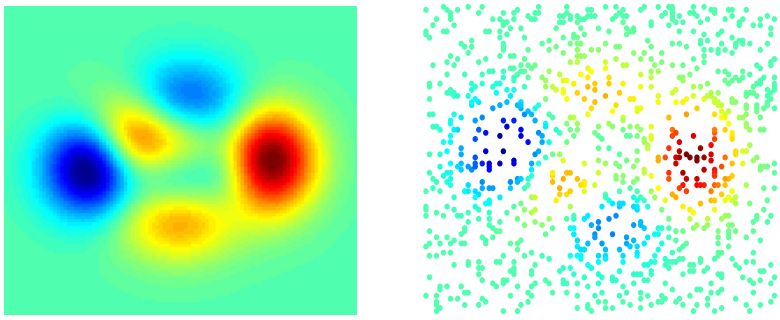


Fig. 2. The singular values sorted in descending order for different powers of diffusion operator O: (a) diffusion operator O when the value of R is $0.1 \times d_{max}m$ (b) diffusion operator O in [12].



(a) (b)

Fig. 3. The synthetic temperature field and synthetic temperature data: (a) the synthetic temperature field (b) the synthetic temperature data.

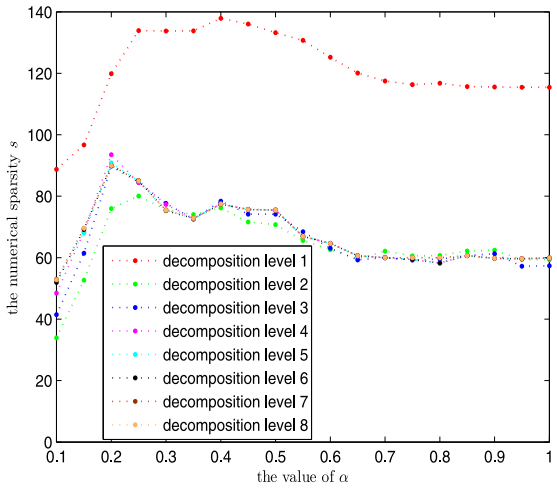


Fig. 4. The relationship between the numerical sparsity s and α.

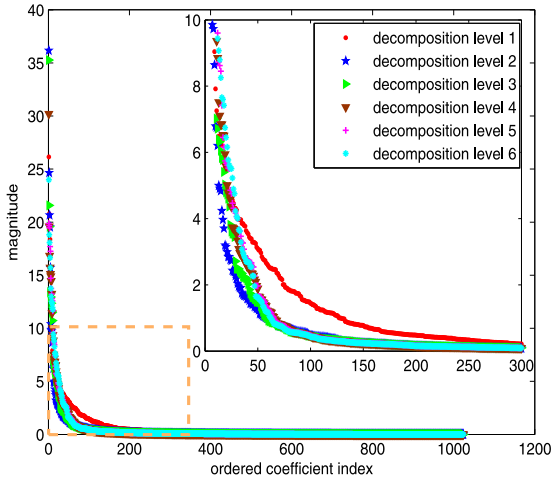


Fig. 5. Diffusion wavelet coefficients versus different values of decomposition level γ.

Table 2
The values of *GI* and s with respect to α and γ.

The values of α and γ	*GI*	s
$\alpha = 0.1, \gamma = 6$	0.8924	51.9474
$\alpha = 1.0, \gamma = 6$	0.8904	59.8835
$\gamma = 6$ [12]	0.8904	59.8835
$\alpha = 0.1, \gamma = 2$	0.8902	33.8886
$\alpha = 1.0, \gamma = 2$	0.8894	59.3483
$\gamma = 2$ [12]	0.8894	59.3483

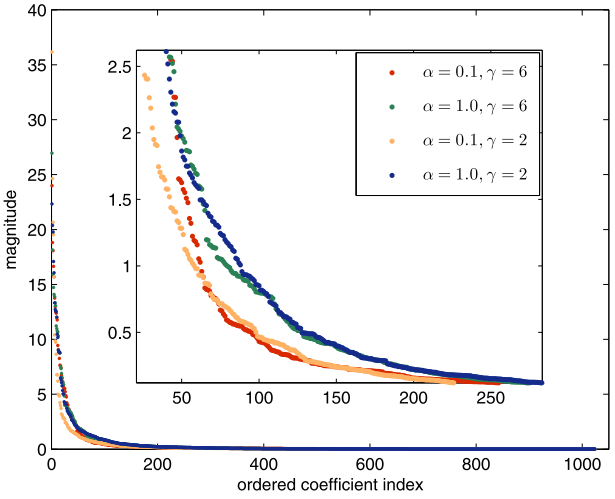


Fig. 6. Diffusion wavelet coefficients versus different values of α and γ.

This is because *GI* and numerical sparsity s have different focus. *GI* denotes relative distribution of energy among these coefficients, but numerical sparsity s represents the number of effectively large coefficients. The paper is to use large coefficients to solve the sparsity problem of sensory data in the sparsifying basis. By contrast, numerical sparsity is a more appropriate choice. If the value of α is 1.0 and the decomposition level γ is the same as that in [12], the constructed diffusion wavelet basis is the same as that in [12]. As can be seen from Table 2, the value of s in the proposed algorithm is 33.8886 and the value of s in [12] is 59.3483 when the value of γ is 2. If the value of α is 0.1 and the value of γ is the same as that in [12], the value of s in the proposed algorithm is less than that in [12]. For different values of α and γ, Fig. 6 shows the change of diffusion wavelet coefficients sorted in decreasing order by the magnitude. From Fig. 6, we can see that diffusion wavelet coefficients corresponding to the smaller value of s attenuate quickly.

To validate the performance of the proposed algorithm, the constructed diffusion wavelet basis is compared with seven sparsifying bases (i.e. spatial basis(identity matrix), Discrete Cosine Transformation(DCT) basis, Discrete Fourier Transform(DFT) basis, haar wavelet basis, bior3.1 wavelet basis, dictionary obtained by K-SVD [29] and diffusion wavelet basis [12]) [30]. Numerical experimental results are listed in Table 3, where the value of γ is 6. It can be seen that numerical sparsity s reaches the minimum value 51.9474 when the sparsifying basis is diffusion wavelet basis in the paper. The value of s in DFT basis is maximal, and *GI* reaches the minimum value 0.2906. The numerical sparsity s of the temperature data in spatial basis is 344.3421. DCT basis, DFT basis, haar wavelet

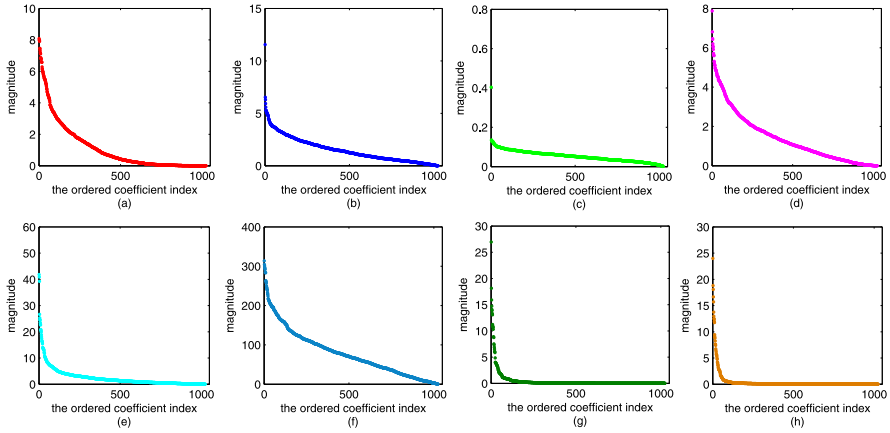


Fig. 7. The coefficients sorted in descending order of synthetic temperature data in eight sparsifying bases: (a) spatial basis (b) DCT basis (c) DFT basis (d) haar wavelet basis (e) bior3.1 wavelet basis (f) dictionary obtained by K-SVD (g) diffusion wavelet basis [12] (h) diffusion wavelet basis in the paper.

Table 3
The values of *s* and *GI* in eight sparsifying bases.

The sparsifying bases	*s*	*GI*
Spatial basis	344.3421	0.6657
DCT basis	624.3799	0.4265
DFT basis	790.6240	0.2906
Haar wavelet basis	544.8247	0.4954
Bior3.1 wavelet basis	289.1851	0.6069
Dictionary obtained by K-SVD	655.8218	0.4098
Diffusion wavelet basis [12]	59.8835	0.8904
Diffusion wavelet basis in the paper	51.9474	0.8924

basis, and dictionary obtained by K-SVD can't make the temperature data sparser than the spatial basis. Their performances are worse in terms of numerical sparsity *s*. Compared with them, the numerical sparsity *s* in bior3.1 wavelet basis reduces to 289.1851. In these sparsifying bases above, diffusion wavelet basis in the paper can make the value of *s* smaller. As mentioned in *Measure of Sparsity*, the smaller the value of *s* is, the sparser sensory data in the sparsifying basis are. Therefore, the temperature data are the sparser in diffusion wavelet basis. Fig. 7 shows the coefficients of temperature data represented in eight sparsifying bases. We can see that attenuation velocity of the coefficients in diffusion wavelet basis is much faster.

4.3.2. Recovery of sensory data

From Eqs. (2) and (3), we know that recovery of sensory data needs to consider the measurement matrix and the sparsifying basis. Sparse binary matrix with a fixed number of nonzero entries in each column satisfies RIP-1. Such matrices have a one-to-one correspondence with the adjacency matrices of fixed left degree expander graph [31]. In the experiment, the measurement matrix Φ is sparse binary matrix with 8 ones in each column, and the sparsifying basis Ψ is the sparsifying basis mentioned above. The coherence between Φ and Ψ is

$$\mu(\Phi, \Psi) = \sqrt{N} \max_{1 \leqslant i,j \leqslant N} \left|\langle \varphi_i, \psi_j \rangle\right| \tag{11}$$

$\mu(\Phi, \Psi)$ measures the largest correlation between any two entries of Φ and Ψ. If the two matrices contain correlated elements, the value of $\mu(\Phi, \Psi)$ is large, Otherwise, it is small. However, the value of $\mu(\Phi, \Psi)$ has a limit, $\mu(\Phi, \Psi) \in [1, \sqrt{N}]$.

If the coefficients of x in the sparsifying basis Ψ is k-sparse, and the number of measurements M satisfies the condition in Eq. (12), x can be recovered with overwhelming probability by solving the l_1-norm minimization problem.

$$M \geqslant C \cdot \mu^2(\Phi, \Psi) \cdot k \cdot logN \tag{12}$$

where C is a positive constant. When the value of $\mu(\Phi, \Psi)$ is 1, M has the minimum value $C \cdot k \cdot logN$ [32]. Because synthetic temperature data are approximately sparse in the sparsifying basis Ψ, k is replaced by numerical sparsity s. Before compressive data gathering, we analyze numerical sparsity s of original sensory data, and bound the minimum value M_{min} by $C \cdot s \cdot logN$. In order to guarantee recovery performance of sensory data, the value of M should be greater than or equal to M_{min}. Once the value of M is determined, sink node takes charge of collecting the measurements.

For WSNs, the measurements are gathered by compressive data gathering algorithms [12,33–35], which are beyond the scope of the paper. Generally speaking, in each measurement, the weighed sums of random projections(the product of sensory data and the corresponding entries in sparse binary matrix) are transmitted to sink node based on compressive data gathering algorithms. After M measurements, sink node collects the measurement vector y of the size $M \times 1$, and recovers these data by BP or BPDN. Relative error ε of recovering sensory data is defined in Eq. (13).

$$\varepsilon = \frac{\|x - \hat{x}\|_2}{\|x\|_2} \tag{13}$$

where x are sensory data and $\hat{x}$ are the recovered sensory data.

As listed in Table 2, diffusion wavelet basis with the better performance is obtained when the value of α is 0.1 and the value of γ is 2. In the diffusion wavelet basis, the numerical sparsity s of temperature data is 33.8886. According to Eq. (12), we know that the value of M_{min} is $C \cdot s \cdot logN$. If the value of C is 1, M_{min} becomes $s \cdot logN \approx 339$. When the value of α is 1.0, and the decomposition level γ is the same as that in [12], the proposed diffusion wavelet basis is the same as that in [12]. In the noiseless case, Fig. 8 draws the recovery results of sensory data by BP, and each subgraph corresponds to diffusion wavelet basis in Table 2. In the second subfigure of Fig. 8, we make a comparison between the proposed diffusion wavelet basis and diffusion wavelet basis in [12]. By contrast, the proposed diffusion wavelet basis can reduce relative error by 0.0038–0.1033. In order to observe the changing trends of relative error, the number of measurements M ranges from 100 to 700. For the noiseless and noise cases, Figs. 9 and 10 depict the comparison results of relative error in eight sparsifying bases. In Fig. 9, the value of α is 0.1, and the value of γ is 6. Seen from Fig. 9, the relative error decreases with the increase of M, and it reaches the minimum in diffusion wavelet basis of the paper. If the value of C is 1, the value of M_{min} is 519. When the value of M isn't less than 500, relative error changes among 0.0281–0.0731. Otherwise, relative error ranges from 0.0572 to 0.7573. When the measurement noise is added to, relative error will change, but the comparative results remain the same. When the value of γ is 2, we

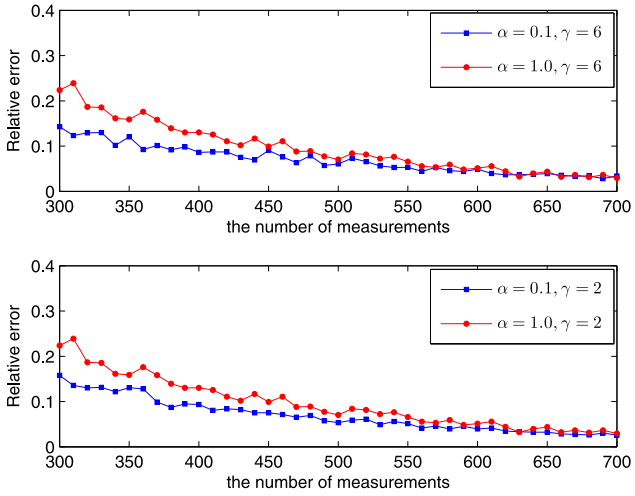


Fig. 8. The comparison results of relative error between diffusion wavelet basis in the paper and that in [12].

obtain diffusion wavelet basis with the better performance. However, both haar wavelet basis and bior3.1 wavelet basis also change. The above experiment is repeated and the results of relative error are shown in Fig. 10. From Fig. 10(a), we can see that relative error in the proposed diffusion wavelet basis is less than other seven sparsifying bases. When the value of M isn't less than 300, the range of relative error is 0.0260–0.1577 in the noiseless case. Otherwise, relative error changes from 0.1444 to 0.3896. The results in Fig. 10(b) are almost the same as those in the noiseless case. Moreover, we discover that relative error in the proposed diffusion wavelet basis is smaller than that in [12], and the difference becomes smaller with the increase of M. When the values of M are 200 and 520, Figs. 11–14 show the visual effects of the recovered temperature data in the noiseless and noise cases, respectively. In Fig. 11, relative error in spatial basis, DCT basis, DFT basis, haar wavelet basis, bior3.1 wavelet basis, dictionary obtained by K-SVD, diffusion wavelet basis [12], and the proposed diffusion wavelet basis is 0.9963, 1.0534, 0.8933, 0.9967, 1.0749, 0.8848, 0.4470 and 0.2539, respectively. From Fig. 11, it can be seen that the recovery performance of the proposed diffusion wavelet basis is better than that in [12]. Fig. 12 takes the measurement noise into account. Relative error corresponding to the eight sparsifying bases is 0.9943, 1.0545, 0.8872, 0.9962, 1.0712, 0.8838, 0.4464 and 0.2503. With the increase of M, relative error is decreasing. When the value of M is 520, the visual effects of the recovered temperature data in diffusion wavelet basis [12] and the proposed diffusion wavelet basis are almost the same, as shown in Figs. 13 and 14. For other values of M, Tables 4 and 5 give the numerical results of relative error corresponding to eight sparsifying bases in the noiseless and noise cases. When the measurement matrices are Gaussian random matrix and Bernoulli matrix, we carry out the recovery experiments of sensory data at the same condition, and show their recovery results in Fig. 15. Seen from four sub-figures, the curves of relative error have the same trends as those in Fig. 10. The proposed diffusion wavelet basis can make the temperature data sparser, and the recovery performances are better than other sparsifying bases.

4.4. Real data experiment

The real data of temperature filed come from National Oceanic and Atmospheric Administration's National Weather Service [26], temperature distribution in USA [27], and Sea Surface Temperature(100 KM Global) [28]. Fig. 16(a) is the temperature filed image of CONUS area provided by National Oceanic and Atmospheric Administration's National Weather Service [26]. Fig. 16(b) is temperature distribution in USA provided by NOAA [37]. Fig. 16(c) and (d) are Sea Surface Temperature(100 KM Global) of two areas. There are 1024 sensor nodes deployed in the area and their locations are the same as those in Fig. 1. Like synthetic data experiment, real data experiments are divided into the sparsity of sensory data and recovery of sensory data.

4.4.1. Sparsity of sensory data

We use real temperature data to verify the proposed algorithm. The maximum value of the decomposition level γ and the value of ϵ are set as 8 and 10^{-6}, respectively. Because the layout of all the nodes is the same as that in the synthetic data experiment, the constructed diffusion operator O in the step 11 and corresponding diffusion wavelet tree in step13∼step24 are the same. However, the value of numerical sparsity s at the same decomposition level γ is different.

Based on the experiments, we observe that real temperature data are approximately sparse in eight sparsifying bases. In the experiments, the value of γ in [12] is set to be the same as that in the proposed algorithm for the best results. Table 6 separately lists the values of s and GI in each sub-figure of Fig. 16. It can be seen that diffusion wavelet basis has an advantage over other six sparsifying bases. For sensory data in Fig. 16(a) and (b), the proposed algorithm obtains the better diffusion wavelet basis when the value of α is 1.0. Because the value of γ is set as the same, diffusion wavelet basis [12] has the same values of s and GI as the proposed diffusion wavelet basis in Fig. 16(a) and (b). For diffusion wavelet

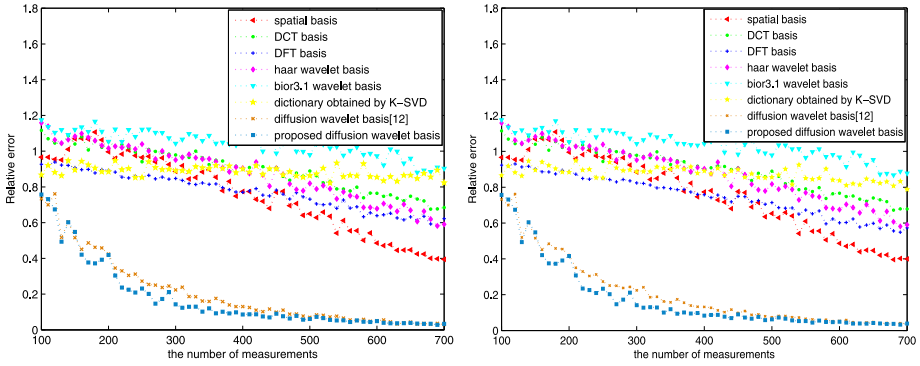


Fig. 9. The comparison results of relative error in eight sparsifying bases when the values of α and γ are 0.1 and 6 separately (a) without the measurement noise (b) with the measurement noise($\sigma_n = 0.05$) [36].

Table 4
The relative error of temperature data without the measurement noise in eight sparsifying bases when the measurement matrix is sparse binary matrix with a fixed number of nonzero entries in each column (A-spatial basis, B-DCT basis, C-DFT basis, D-haar wavelet basis, E-bior3.1 wavelet basis, F-dictionary obtained by K-SVD, G-diffusion wavelet basis [12], H-the proposed diffusion wavelet basis).

M	A	B	C	D	E	F	G	H
100	0.9662	1.1167	0.9262	1.1353	1.1347	0.8669	0.6914	0.3811
120	0.9542	1.0455	0.9388	1.1305	1.1276	0.8815	0.6470	0.3896
140	1.0104	1.0859	0.9203	1.0970	1.1044	0.9473	0.5836	0.3529
160	1.0718	1.0811	0.8962	1.1094	1.1290	0.9434	0.4351	0.2603
180	1.1070	1.0568	0.8777	1.0285	1.1178	0.9053	0.4085	0.2775
200	0.9963	1.0534	0.8933	0.9967	1.0749	0.8848	0.4470	0.2539
220	0.9839	1.0367	0.8552	1.0235	1.0531	0.8962	0.3013	0.1863
240	0.9749	0.9955	0.8608	1.0353	1.0956	0.8527	0.2821	0.2047
260	0.9706	1.0095	0.8648	1.0565	1.0961	0.9408	0.2900	0.1444
280	0.9624	0.9933	0.8547	0.9776	1.0721	0.8970	0.2599	0.1701
300	0.8872	0.9791	0.8475	0.9444	1.0522	0.8905	0.2548	0.1577
320	0.8446	0.9651	0.8191	0.9813	1.0766	0.8863	0.1856	0.1303
340	0.8836	0.9629	0.8219	0.9237	1.0809	0.9076	0.1600	0.1219
360	0.8541	0.9494	0.8167	0.9148	1.0669	0.9197	0.1719	0.1279
380	0.8310	0.8842	0.7843	0.9344	1.0170	0.9175	0.1654	0.0870
400	0.7745	0.8983	0.7696	0.9086	1.0425	0.9007	0.1618	0.0934
420	0.7306	0.9220	0.7907	0.9054	1.0571	0.9185	0.1267	0.0839
440	0.6800	0.8988	0.7496	0.8365	1.0711	0.9014	0.1104	0.0751
460	0.7755	0.8594	0.7594	0.8931	1.0152	0.8674	0.1113	0.0713
480	0.7068	0.8774	0.7674	0.8315	0.9783	0.8613	0.1058	0.0691
500	0.6497	0.8885	0.7316	0.8303	0.9652	0.8633	0.0886	0.0534
520	0.6589	0.8228	0.7003	0.8160	0.9883	0.9046	0.0833	0.0609
540	0.5412	0.7891	0.6800	0.7365	0.9550	0.9338	0.0757	0.0559
560	0.5546	0.7771	0.6814	0.7889	0.9700	0.8649	0.0573	0.0414
580	0.5008	0.7394	0.6330	0.6477	0.9560	0.9297	0.0516	0.0395
600	0.4858	0.7645	0.6390	0.7121	0.9206	0.8544	0.0555	0.0395
620	0.4777	0.7554	0.6522	0.6986	0.9329	0.8578	0.0440	0.0349
640	0.4464	0.7221	0.6159	0.6451	0.9084	0.8647	0.0421	0.0324
660	0.4251	0.7322	0.6129	0.6349	0.8344	0.8470	0.0363	0.0285
680	0.4006	0.6757	0.5929	0.5695	0.8312	0.8554	0.0313	0.0263
700	0.3945	0.6844	0.6225	0.5736	0.8368	0.8235	0.0312	0.0260

Table 5
The relative error of temperature data with the measurement noise($\sigma_n = 0.05$) [36] in eight sparsifying bases when the measurement matrix is sparse binary matrix with a fixed number of nonzero entries in each column (A-spatial basis, B-DCT basis, C-DFT basis, D-haar wavelet basis, E-bior3.1 wavelet basis, F-dictionary obtained by K-SVD, G-diffusion wavelet basis [12], H-the proposed diffusion wavelet basis).

M	A	B	C	D	E	F	G	H
100	0.9664	1.1163	0.9233	1.1336	1.1332	0.8658	0.6891	0.3798
120	0.9542	1.0457	0.9302	1.1291	1.1277	0.8822	0.6460	0.3926
140	1.0119	1.0867	0.9168	1.0979	1.1032	0.9467	0.5805	0.3521
160	1.0711	1.0807	0.8865	1.1097	1.1294	0.9411	0.4358	0.2601
180	1.1085	1.0566	0.8742	1.0282	1.1171	0.9034	0.4056	0.2792
200	0.9943	1.0545	0.8872	0.9962	1.0712	0.8838	0.4464	0.2503
220	0.9801	1.0344	0.8492	1.0223	1.0514	0.8967	0.2995	0.1844
240	0.9770	0.9943	0.8542	1.0342	1.0947	0.8532	0.2810	0.2034
260	0.9665	1.0092	0.8425	1.0563	1.0960	0.9399	0.2895	0.1415
280	0.9627	0.9930	0.8430	0.9777	1.0694	0.8928	0.2567	0.1695
300	0.8861	0.9773	0.8238	0.9800	1.0747	0.8867	0.1826	0.1298
320	0.8423	0.9641	0.8119	0.9214	1.0856	0.9071	0.1620	0.1233
340	0.8790	0.9610	0.7975	0.9166	1.0656	0.9148	0.1769	0.1247
360	0.8525	0.9483	0.7865	0.9341	1.0156	0.9143	0.1674	0.0886
380	0.8253	0.8834	0.7728	0.9087	1.0359	0.8913	0.1652	0.0895
400	0.7738	0.8976	0.7581	0.9047	1.0547	0.9058	0.1273	0.0855
420	0.7312	0.9185	0.7561	0.8362	1.0677	0.8912	0.1088	0.0757
440	0.6757	0.8976	0.7185	0.8889	1.0156	0.8561	0.1107	0.0756
460	0.7691	0.8557	0.7475	0.8349	0.9746	0.8555	0.1080	0.0693
480	0.7129	0.8736	0.7257	0.8286	0.9681	0.8559	0.0872	0.0550
500	0.6452	0.8900	0.7141	0.8177	0.9924	0.9017	0.0796	0.0604
520	0.6531	0.8165	0.6829	0.8177	0.9924	0.9017	0.0796	0.0604
540	0.5468	0.7858	0.6497	0.7333	0.9547	0.9246	0.0717	0.0564
560	0.5521	0.7708	0.6519	0.7800	0.9679	0.8440	0.0618	0.0450
580	0.4998	0.7362	0.5956	0.6498	0.9512	0.9026	0.0518	0.0414
600	0.4861	0.7717	0.6017	0.7088	0.9180	0.8374	0.0540	0.0444
620	0.4833	0.7496	0.6229	0.6959	0.9254	0.8195	0.0476	0.0397
640	0.4471	0.7227	0.5886	0.6489	0.9012	0.8467	0.0481	0.0358
660	0.4211	0.7259	0.5697	0.6340	0.8126	0.8220	0.0439	0.0324
680	0.3928	0.6660	0.5560	0.5661	0.8271	0.8072	0.0386	0.0303
700	0.4008	0.6784	0.5732	0.5755	0.8264	0.7873	0.0353	0.0309

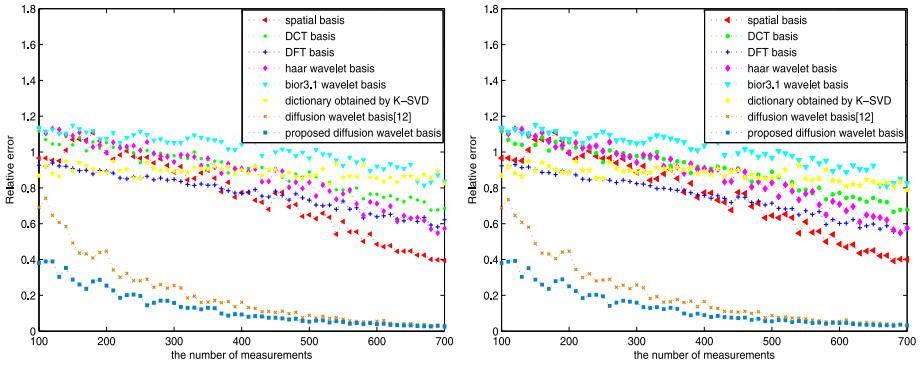


Fig. 10. The comparison results of relative error in eight sparsifying bases when the values of α and γ are 0.1 and 2 separately (a) without the measurement noise (b) with the measurement noise ($\sigma_n = 0.05$) [36].

Table 6
The values of s and GI of real temperature data in eight sparsifying bases.

The sparsifying bases	s	Fig. 16(a)	Fig. 16(b)	Fig. 16(c)	Fig. 16(d)
Spatial basis	s	1000.7978	901.1341	873.9807	967.2009
DCT basis	s	42.0655	125.5184	144.5129	73.3424
DFT basis	s	27.2497	117.0807	141.4851	59.3052
Haar wavelet basis	s	209.1440	253.0312	220.4146	96.1711
Bior3.1 wavelet basis	s	220.6655	258.1278	238.2169	144.8614
Dictionary obtained by K-SVD	s	660.6477	650.8165	633.0333	646.7721
Diffusion wavelet basis [12]	s	9.3906	14.7327	12.3609	10.1734
The proposed diffusion wavelet basis	s	9.3906	14.7327	11.2932	9.3255
The sparsifying bases	GI	Fig. 16(a)	Fig. 16(b)	Fig. 16(c)	Fig. 16(d)
Spatial basis	GI	0.0834	0.2111	0.2318	0.1357
DCT basis	GI	0.6086	0.4996	0.4928	0.5437
DFT basis	GI	0.4348	0.3434	0.3338	0.3699
Haar wavelet basis	GI	0.7365	0.6128	0.5776	0.6282
Bior3.1 wavelet basis	GI	0.7316	0.6268	0.6057	0.6443
Dictionary obtained by K-SVD	GI	0.4094	0.4155	0.4261	0.4176
Diffusion wavelet basis [12]	GI	0.7769	0.8196	0.8596	0.8412
Proposed diffusion wavelet basis	GI	0.7769	0.8196	0.8475	0.8346

basis [12], the values of s are 12.3609 and 10.1734 in Fig. 16(c) and (d), respectively. While the values of s in the proposed diffusion wavelet basis are 11.2932 and 9.3255, respectively. Compared with diffusion wavelet basis [12], the proposed diffusion wavelet basis reduces the value of s in the two sub-figures, and makes real temperature data sparser.

In the following analysis, we take real temperature data in Fig. 16(a) for example. With the increase of α, the numerical sparsity s will change, and the changes are depicted in Fig. 17. When the decomposition level γ is 1, the numerical sparsity s of real temperature data is maximal, which signifies that real temperature data are the least sparse. This phenomenon also appears in synthetic data experiment. The reason is that diffusion wavelet tree only includes orthogonal bases of scaling functions if the value of γ is 1. While an integral diffusion wavelet tree should consist the orthogonal bases of scaling functions and the orthogonal bases of wavelets. Scaling functions are smoothly bumped functions with some oscillations(approximation), and wavelets are smoothly localized oscillatory(details) functions at the same scale. If only the orthogonal bases of scaling function exist, details of real temperature data represented in diffusion wavelet basis are lost, which results in the increase of s. This is also the reason why the scope of J is $1 \sim \gamma - 1$ in step 25. When the value of α is 1.0 and the value of γ is 3, the proposed algorithm can obtain diffusion wavelet basis with the better performance. If GI is the evaluation index, the values of α and γ corresponding to the maximum value of GI are 0.1 and 2, respectively. Their diffusion wavelet coefficients sorted in decreasing order by the magnitude are denoted in Fig. 18, and

Table 7
The values of GI and s with respect to α and γ.

The values of α and γ	GI	s
$\alpha = 0.1, \gamma = 2$	0.8448	60.2793
$\alpha = 1.0, \gamma = 2$	0.7890	10.8081
$\gamma = 2$ [12]	0.7890	10.8081
$\alpha = 0.1, \gamma = 3$	0.8215	37.0535
$\alpha = 1.0, \gamma = 3$	0.7769	9.3906
$\gamma = 3$ [12]	0.7769	9.3906

the numerical results are given in Table 7. When the value of γ in [12] is the same as that in the proposed diffusion wavelet basis, sensory data in Fig. 16(a) have the same diffusion wavelet coefficients in two diffusion wavelet bases. When the value of α is 0.1, the proposed diffusion wavelet basis is better in terms of GI. However, the numerical sparsity is more efficient in the proportion of effectively large coefficients than GI. As shown in Fig. 18, the coefficients with respect to the proposed diffusion wavelet basis contain the fewer large coefficients, and real temperature data are sparser in the proposed diffusion wavelet basis.

For sensory data in Fig. 16(a), we compare the proposed diffusion wavelet basis with other seven sparsifying bases, and denote the comparative results of the coefficients in Fig. 19. The numerical comparisons are summarized in Table 6, where the value of α is 1.0 and the value of γ is 3 in Fig. 16(a). Seen from Table 6, the numerical sparsity s of real temperature data in the spatial basis is 1000.7978, which is in close proximity to 1024, that is, real

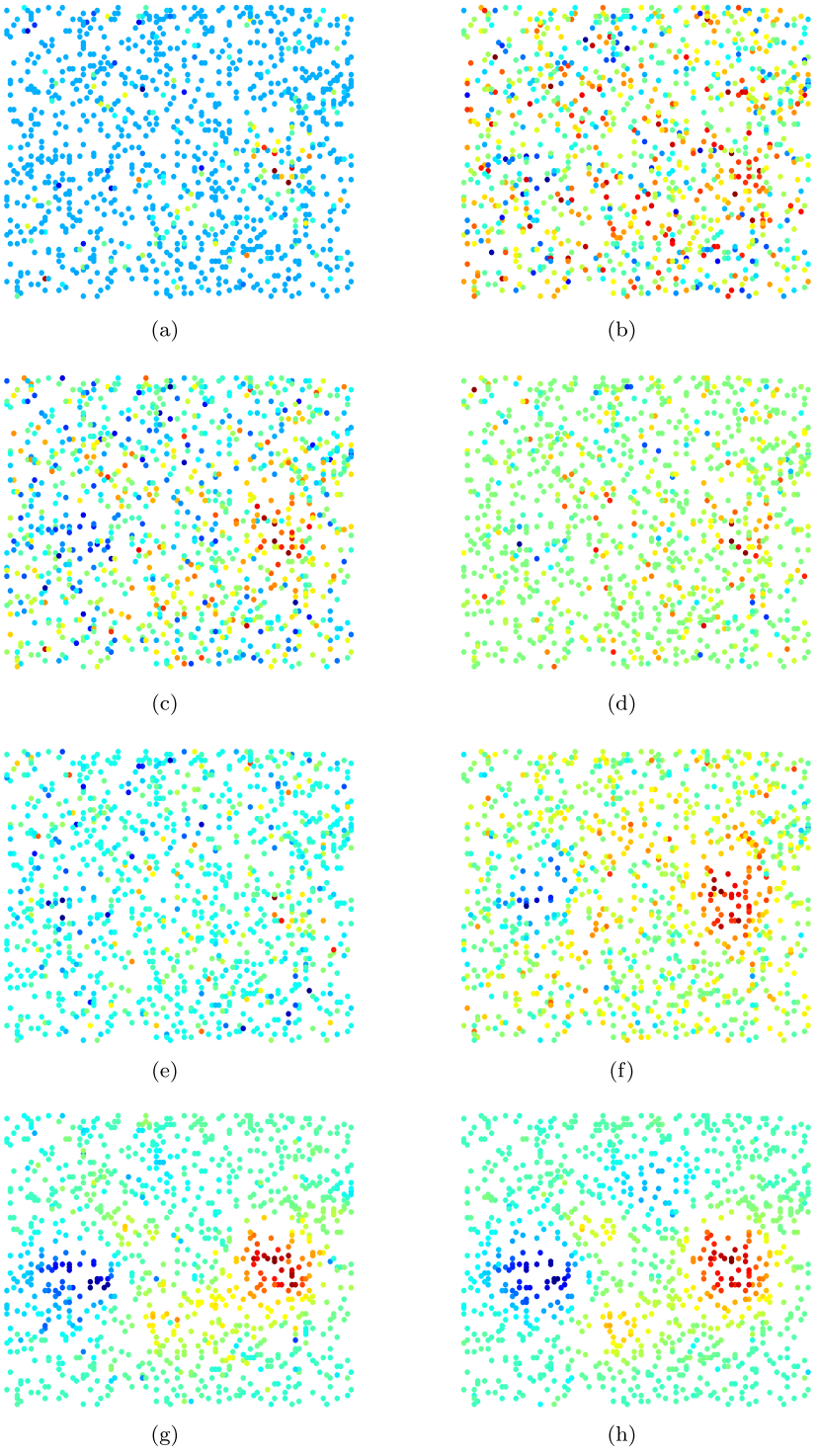


Fig. 11. The recovered temperature data by BP in eight sparsifying bases when the value of M is 200 without the measurement noise: (a) spatial basis (b) DCT basis (c) DFT basis (d) haar wavelet basis (e) bior3.1 wavelet basis (f) dictionary obtained by K-SVD (g) diffusion wavelet basis [12] (h)the proposed diffusion wavelet basis.

temperature data are the least sparse in the spatial basis. Like the spatial basis, the numerical sparsity s of real temperature data in dictionary obtained by K-SVD is larger than other six sparsifying bases. In haar wavelet basis, the value of s is 209.1440. Bior3.1 wavelet basis is almost the same as haar wavelet basis. Compared with DFT basis, real temperature data have more large coefficients in DCT basis, which leads to the increase of s. The proposed diffusion wavelet basis can make the numerical sparsity s reach the minimum value 9.3906. So the performance of the proposed diffusion wavelet basis is better than the other seven sparsifying bases. For the best results, diffusion wavelet basis [12] has the same performance as the proposed diffusion wavelet basis.

4.4.2. Recovery of sensory data

In the recovery of sensory data, the measurement matrix and the sparsifying basis need to be considered. In real data experiments, the measurement matrix Φ is the same as that in the synthetic data experiment, and the sparsifying basis Ψ is the proposed

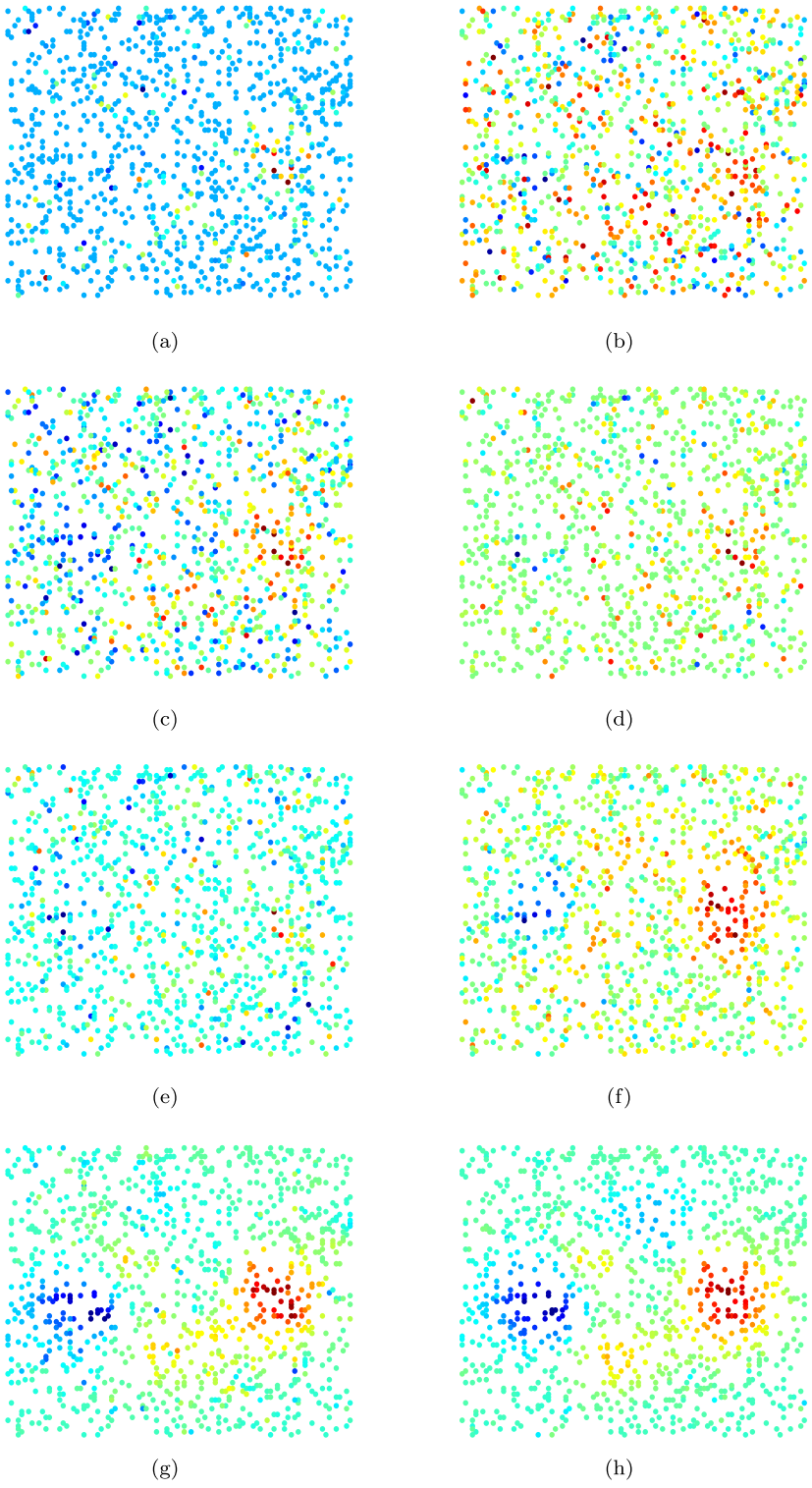


Fig. 12. The recovered temperature data by BPDN in eight sparsifying bases when the value of M is 200 with the measurement noise($\sigma_n = 0.05$) [36]: (a) spatial basis (b) DCT basis (c) DFT basis (d) haar wavelet basis (e) bior3.1 wavelet basis (f) dictionary obtained by K-SVD (g) diffusion wavelet basis [12] (h) the proposed diffusion wavelet basis.

diffusion wavelet basis. Besides, the other seven sparsifying bases mentioned above are used for comparison.

Similarly, sensory data in Fig. 16(a) are taken for example. From Table 6, it can be seen that numerical sparsity s of real temperature data in the proposed diffusion wavelet basis is 9.3906. Based on Eq. (12) and the lower bound of $\mu(\Phi, \Psi)$, the minimum value M_{min} is determined by $C \cdot s \cdot logN$. When the value of C is 1, M_{min} is equal to $s \cdot logN \approx 94$. If the value of α is 0.1, and the value of γ is 2, sensory data in Fig. 16(a) have the maximum value of GI. If the value of α is 1.0, and the value of γ is 3, the data in Fig. 16(a) have the minimum value of s. When the numerical sparsity s reaches the minimum value, we obtain the diffusion wavelet basis with better performance. Corresponding to GI and s, Fig. 20 depicts the recovery performance of sensory data

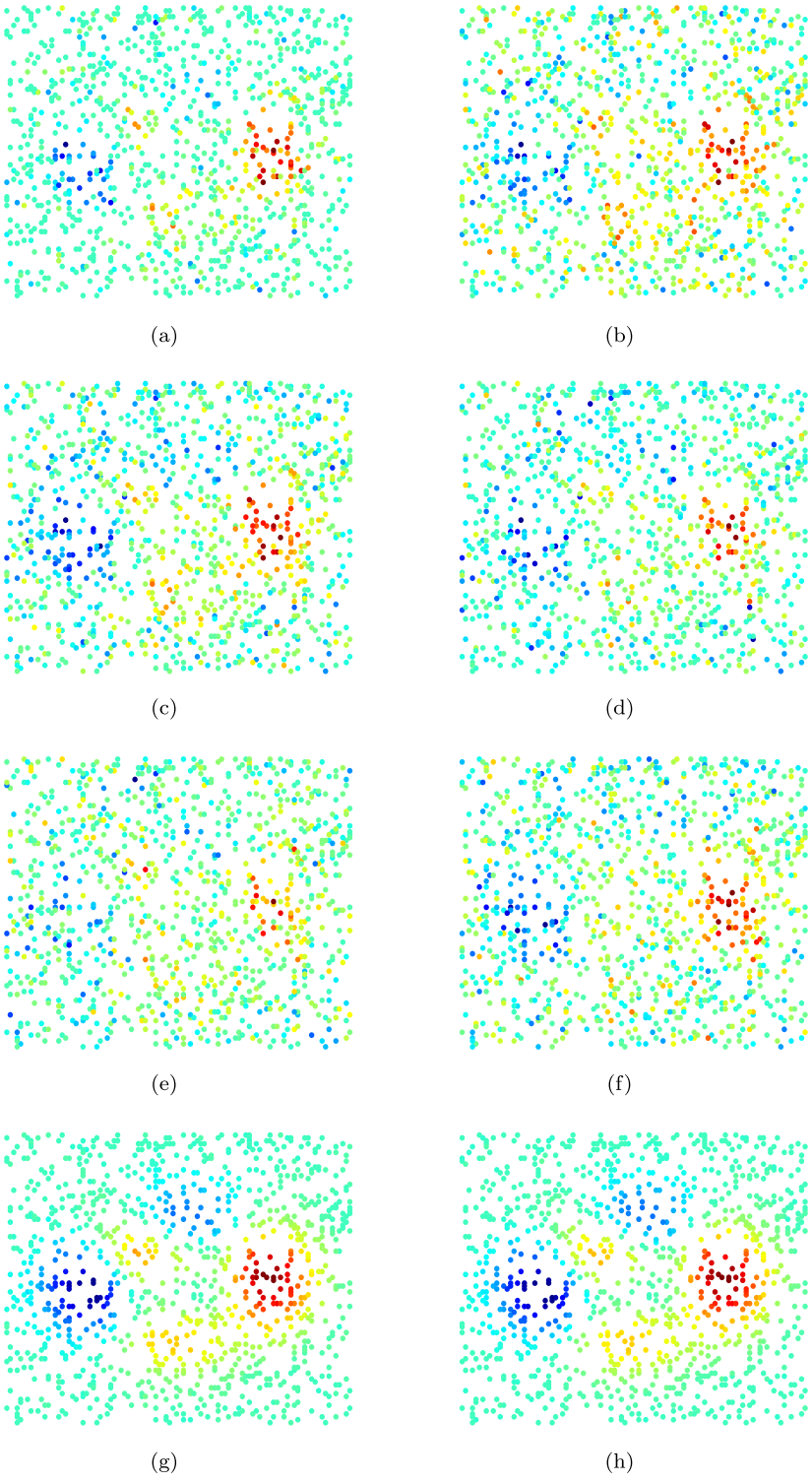


Fig. 13. The recovered temperature data by BP in eight sparsifying bases when the value of M is 520 without the measurement noise: (a) spatial basis (b) DCT basis (c) DFT basis (d) haar wavelet basis (e) bior3.1 wavelet basis (f) dictionary obtained by K-SVD (g) diffusion wavelet basis [12] (h) the proposed diffusion wavelet basis.

in Fig. 16(a) by BP. In the first sub-figure, the proposed diffusion wavelet basis decreases the relative error by 0.0098–0.1301 than diffusion wavelet basis [12] with the maximum value of *GI*. In the second sub-figure, compared with the diffusion wavelet basis [12] with the maximum value of *GI*, the proposed diffusion wavelet basis reduces the relative error by 0.0086–0.1010. For sensory data in Fig. 16, we conduct the recovery experiments in the noiseless and noise cases. Fig. 21 draws the relative error between the sensory data in Fig. 16 and the recovered sensory data in the noiseless case, and gives the comparison results between the eight sparsifying bases. Fig. 21 includes four sub-figures, and each sub-figure separately denotes the relative error of the recovery of real temperature data in each sub-figure. We have analyzed the recovery of sensory data in Fig. 16(a) above. When the value of α is 1.0 and the value of γ is 4, diffusion wavelet basis with the better performance is obtained for sensory data in Fig. 16(b). When the value

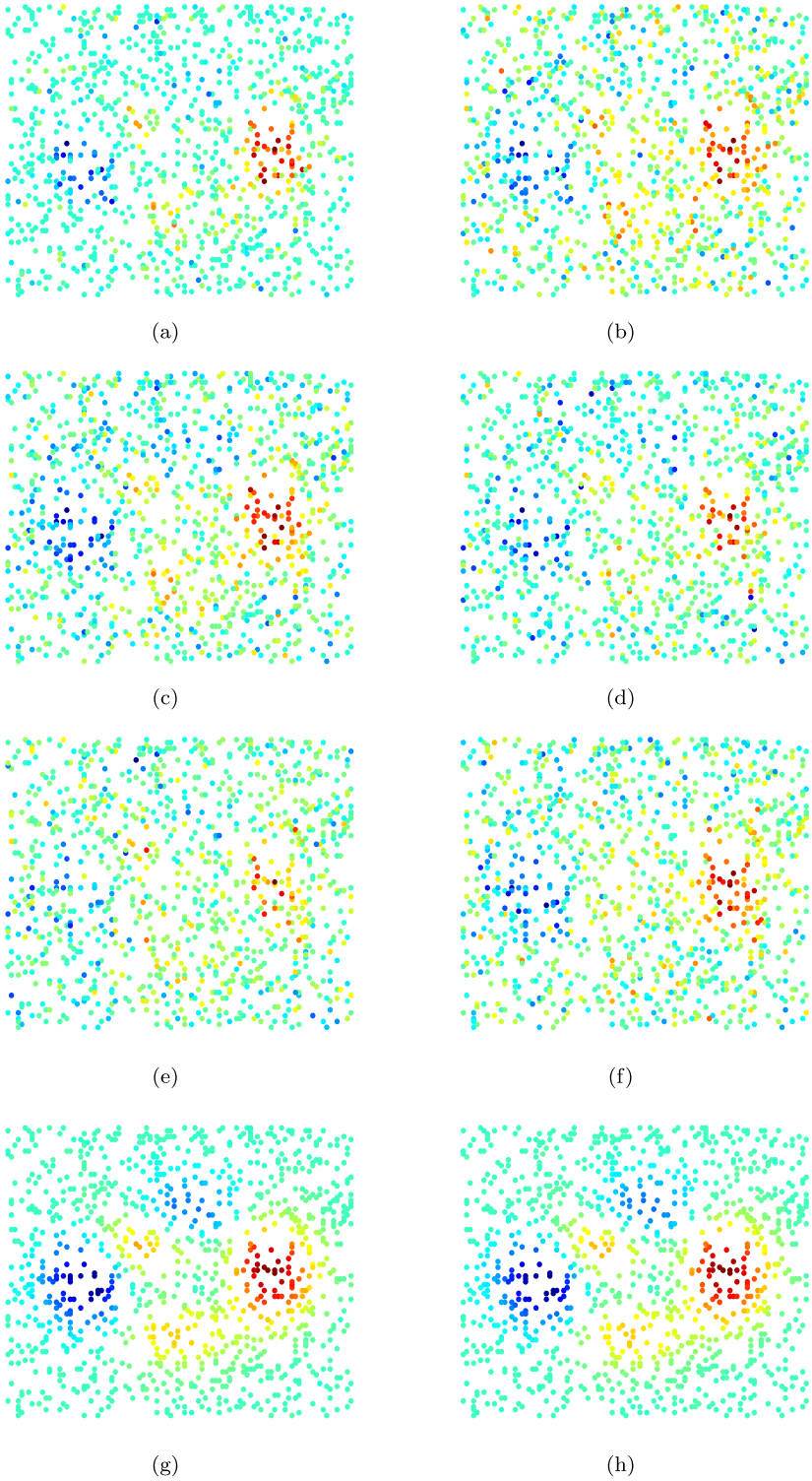


Fig. 14. The recovered temperature data by BPDN in eight sparsifying bases when the value of M is 520 with the measurement noise($\sigma_n = 0.05$) [36]: (a) spatial basis (b) DCT basis (c) DFT basis (d) haar wavelet basis (e) bior3.1 wavelet basis (f) dictionary obtained by K-SVD (g) diffusion wavelet basis [12] (h)the proposed diffusion wavelet basis.

of α is 0.2, sensory data in Fig. 16(c) and (d) are the sparsest. At this moment, the corresponding values of γ are 5 in Fig. 16(c), and 6 in Fig. 16(d). The numerical results of relative error are listed in Table 8.

With the increase of M, it can be seen from Fig. 21 that relative error decreases gradually. Because the number of sensor nodes N is 1024, the numerical sparsity s of real temperature data is a constant in the given sparsifying basis. If the value of M is larger, compressive data gathering can acquire the more information about original sensory data. In this case, relative error is smaller and the recovery of data is more accurate. However, if the value of M is smaller, it isn't enough for accurate recovery, and relative error becomes larger. In four sub-figures, the spatial basis and DFT basis have an advantage over four sparsifying bases(i.e. DCT basis, haar wavelet basis, bior3.1 wavelet basis and dictionary obtained by K-SVD). However, relative error in diffusion wavelet basis [12] and

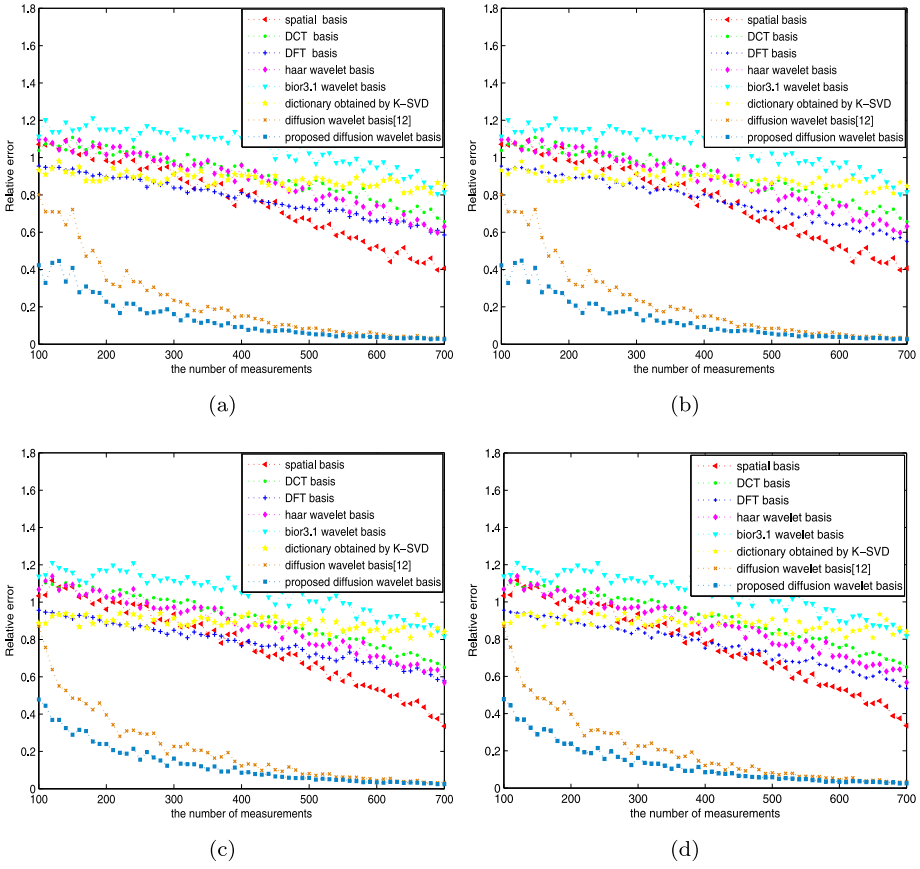


Fig. 15. The comparison results of relative error in eight sparsifying bases when the measurement matrices are Gaussian random matrix and Bernoulli matrix (a) Gaussian random matrix without the measurement noise (b) Gaussian random matrix with the measurement noise($\sigma_n = 0.05$) [36] (c) Bernoulli matrix without the measurement noise (d) Bernoulli matrix with the measurement noise($\sigma_n = 0.05$) [36].

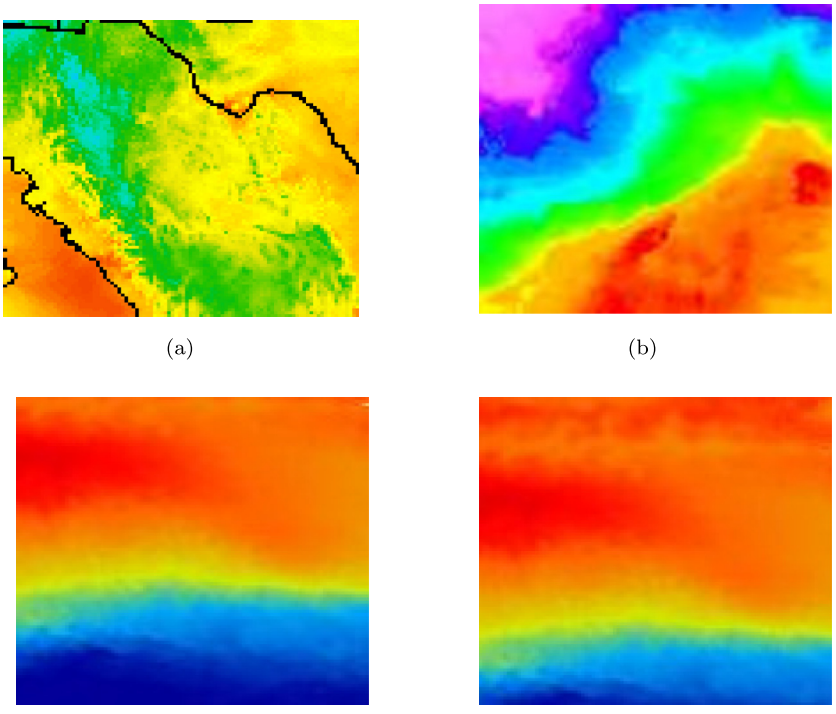


Fig. 16. (a) The temperature filed image of CONUS area, 8AM EST, Thu MAY 26, 2016 (b) temperature distribution in USA (c) Sea Surface Temperature(100 KM Global) (d) Sea Surface Temperature(100 KM Global).

Table 8
The relative error of the recovery of sensory data without the measurement noise in eight sparsifying bases.

The sparsifying bases	M	Fig. 21(a)	Fig. 21(b)	Fig. 21(c)	Fig. 21(d)
The spatial basis	100	0.1432	0.3246	0.3639	0.2236
DCT basis	100	0.2660	0.4573	0.4901	0.3804
DFT basis	100	0.1432	0.3246	0.3639	0.2236
Haar wavelet basis	100	0.3212	0.6674	0.5293	0.2675
Bior3.1 wavelet basis	100	0.3388	0.5648	0.5186	0.2838
Dictionary obtained by K-SVD	100	0.1775	0.4024	0.4314	0.2681
Diffusion wavelet basis [12]	100	0.1139	0.1884	0.1862	0.1776
Proposed diffusion wavelet basis	100	0.1139	0.1884	0.1594	0.1472
The spatial basis	150	0.1399	0.3186	0.3521	0.2168
DCT basis	150	0.2198	0.4296	0.4705	0.3465
DFT basis	150	0.1399	0.3186	0.3520	0.2168
Haar wavelet basis	150	0.3814	0.4753	0.4633	0.2623
Bior3.1 wavelet basis	150	0.3319	0.4977	0.4896	0.2738
Dictionary obtained by K-SVD	150	0.1678	0.3948	0.4270	0.2748
Diffusion wavelet basis [12]	150	0.1205	0.1320	0.1176	0.1265
Proposed diffusion wavelet basis	150	0.1205	0.1320	0.1017	0.1234
The spatial basis	200	0.1337	0.3092	0.3492	0.2150
DCT basis	200	0.2558	0.4405	0.4558	0.3255
DFT basis	200	0.1337	0.3091	0.3491	0.2150
Haar wavelet basis	200	0.2642	0.4446	0.4299	0.2528
Bior3.1 wavelet basis	200	0.2473	0.4572	0.4571	0.2718
Dictionary obtained by K-SVD	200	0.1808	0.3923	0.4411	0.2669
Diffusion wavelet basis [12]	200	0.0908	0.1143	0.1004	0.1236
Proposed diffusion wavelet basis	200	0.0908	0.1143	0.0922	0.1017
The spatial basis	250	0.1338	0.3048	0.3338	0.2027
DCT basis	250	0.2098	0.3988	0.4437	0.3011
DFT basis	250	0.1338	0.3049	0.3339	0.2027
Haar wavelet basis	250	0.2067	0.4109	0.4406	0.2574
Bior3.1 wavelet basis	250	0.2255	0.4230	0.4614	0.2718
Dictionary obtained by K-SVD	250	0.1710	0.3772	0.4449	0.2721
Diffusion wavelet basis [12]	250	0.0871	0.1135	0.0831	0.0922
Proposed diffusion wavelet basis	250	0.0871	0.1135	0.0738	0.0894
The spatial basis	300	0.1259	0.2844	0.3295	0.1954
DCT basis	300	0.1988	0.4159	0.4218	0.2835
DFT basis	300	0.1262	0.2844	0.3295	0.1954
Haar wavelet basis	300	0.2067	0.4031	0.4177	0.2446
Bior3.1 wavelet basis	300	0.2169	0.4520	0.4513	0.2676
Dictionary obtained by K-SVD	300	0.1784	0.3875	0.4534	0.2732
Diffusion wavelet basis [12]	300	0.0905	0.0946	0.0701	0.0865
Proposed diffusion wavelet basis	300	0.0905	0.0946	0.0644	0.0730

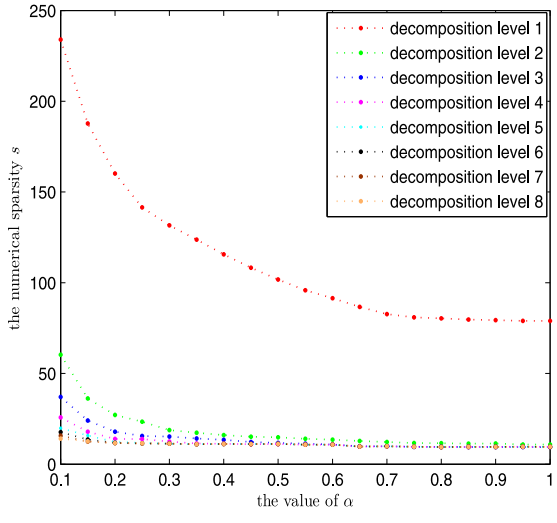


Fig. 17. The relationship between the numerical sparsity s and α in Fig. 16(a).

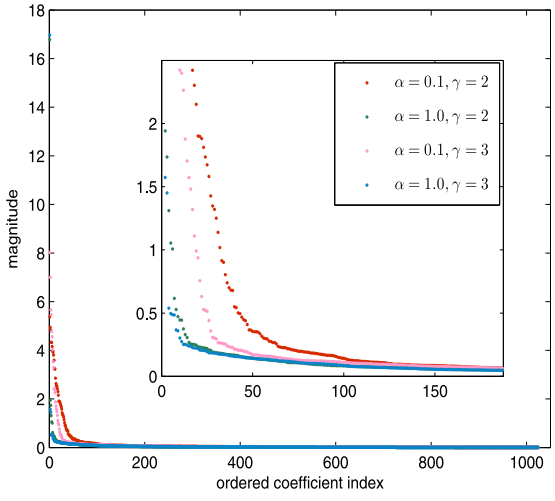


Fig. 18. Diffusion wavelet coefficients versus different values of α and γ in Fig. 17.

the proposed diffusion wavelet basis are less than that in the two sparsifying bases(spatial basis and DFT basis). For sensory data in Fig. 16(a) and (b), the proposed diffusion wavelet basis is gained when the value of α is 1.0. Under the best circumstance, the value of γ in diffusion wavelet basis [12] is the same as that in the proposed diffusion wavelet basis, and the recovery results are the same, as shown in Fig. 21(a) and (b). In diffusion wavelet basis [12], the numerical sparsity s of real temperature data in Fig. 21(c) and (d) are 12.3609 and 10.1734, respectively. If the proposed diffusion wavelet basis is used, the numerical sparsity s reduces to

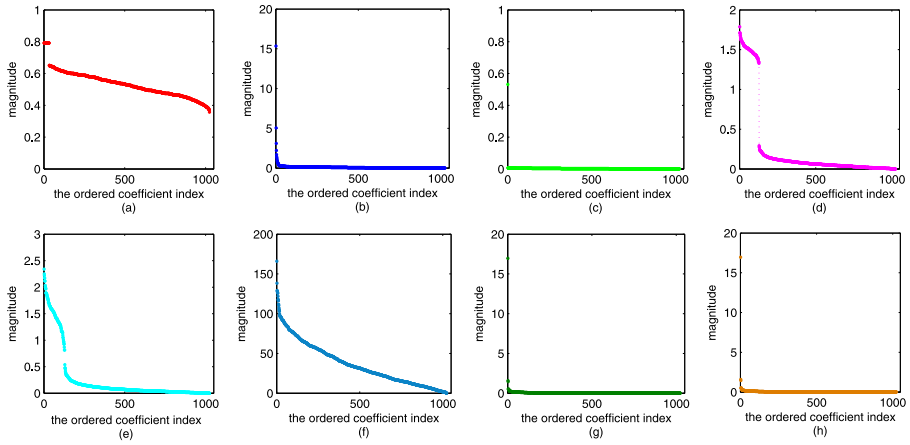


Fig. 19. The coefficients sorted in descending order of real temperature data in eight sparsifying bases: (a) spatial basis (b) DCT basis (c) DFT basis (d) haar wavelet basis (e) bior3.1 wavelet basis (f) dictionary obtained by K-SVD (g) diffusion wavelet basis [12] (h)the proposed diffusion wavelet basis.

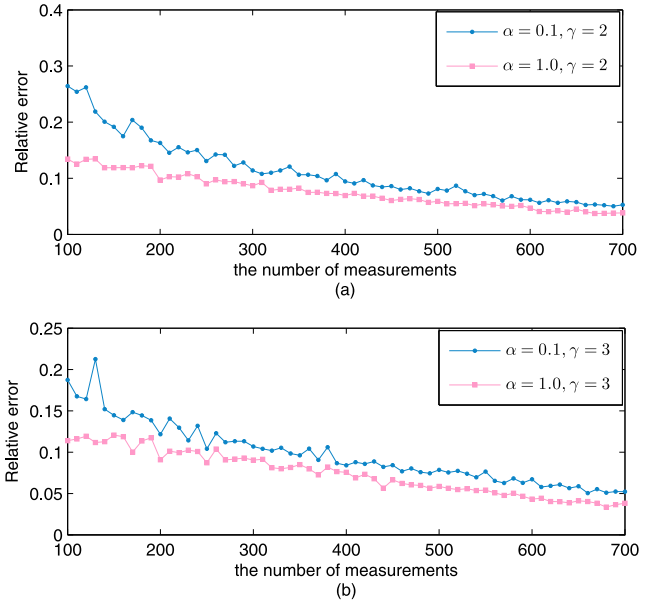


Fig. 20. The comparison results of the proposed diffusion wavelet basis and the diffusion wavelet basis [12].

11.2932 and 9.3255. When the value of M ranges from 100 to 300, we observe from Fig. 21(c) and (d) that relative error of the proposed diffusion wavelet is less than that of diffusion wavelet basis [12]. When the value of M becomes larger, their results are almost the same. As a result, diffusion wavelet basis constructed by the proposed algorithm can make sensory data sparser, and reduces the relative error when the value of M is smaller. Furthermore, when the measurement matrices are Gaussian random matrix and Bernoulli matrix, we also study the recovery performance of the proposed diffusion wavelet basis in the noiseless and noise cases, and show the relative error of data recovery in Fig. 22. In the same way, we can also draw the conclusion that the recovery performance of the proposed diffusion wavelet is better than other sparsifying bases.

When the measurement noise is considered and the measurement matrix is sparse binary matrix with a fixed number of nonzero entries in each column, Fig. 23 shows the comparison results of relative error in eight sparsifying bases, and Table 9 gives the numerical results of relative error. Compared with Fig. 21, the recovery performances in haar wavelet basis and bior3.1 wavelet basis are greatly influenced by the measurement noise, and the peak value of relative error appears. Like the curves in Fig. 21, spatial basis and DFT basis have the almost the same relative error. For diffusion wavelet basis [12], the value of γ is set as the same as the proposed diffusion wavelet basis for the best results. In Fig. 23(a) and (b), the proposed algorithm gets diffusion wavelet basis with the better performance when the value of α is 1.0. So the two diffusion wavelet bases have the same relative error in two sub-figures. Seen from Fig. 23(c) and (d), the proposed diffusion wavelet basis reduces relative error compared with other seven sparsifying bases. When the value of M ranges from 100 to 300 by the step size 50, relative error of the recovery of sensory data in Fig. 23(c) is 0.1594, 0.1017, 0.0922, 0.0738 and 0.0644 in the proposed diffusion wavelet basis. For sensory data in Fig. 23(d), relative error of data recovery is 0.1468, 0.1246, 0.1044, 0.0913 and 0.0765 when the proposed diffusion wavelet basis is chosen as the sparsifying basis. Through the experiments and analysis, we know that the proposed diffusion wavelet basis has the advantage over other seven sparsifying bases in the noiseless and noise cases.

5. Conclusions

The paper studies the effect of the communication radius and the wavelet decomposition level on the construction of diffusion operator, and proposes a diffusion wavelet basis algorithm to make sensory data on the graph represented in a sparser way. Based on the locations of sensor nodes and the communication radius, we analyze the performance of high powers of diffusion operator and conduct the experiments. Theoretical and experimental investigations prove that high powers of the constructed diffusion operator have low numerical rank. Based on measure of sparsity, the algorithm chooses the diffusion wavelet basis with better performance. From the results of synthetic data experiment and real data experiment, we see that sensory data are approximately sparse in different sparsifying bases. Compared with seven sparsifying bases such as the spatial basis, DCT basis, DFT basis, haar wavelet basis, bior3.1 wavelet basis, dictionary obtained by K-SVD and diffusion wavelet basis [12], sensory data are the sparsest in the proposed diffusion wavelet. Furthermore, we consider the effect of the noise on data recovery, and study the recovery of sensory data in the noise cases. Through the experiments, we discover that the proposed diffusion wavelet basis is superior to other seven sparsifying bases in terms of data recovery when the difference between diffusion wavelet basis [12] and the proposed diffusion wavelet basis is larger. If the difference is smaller, diffusion wavelet basis [12] have

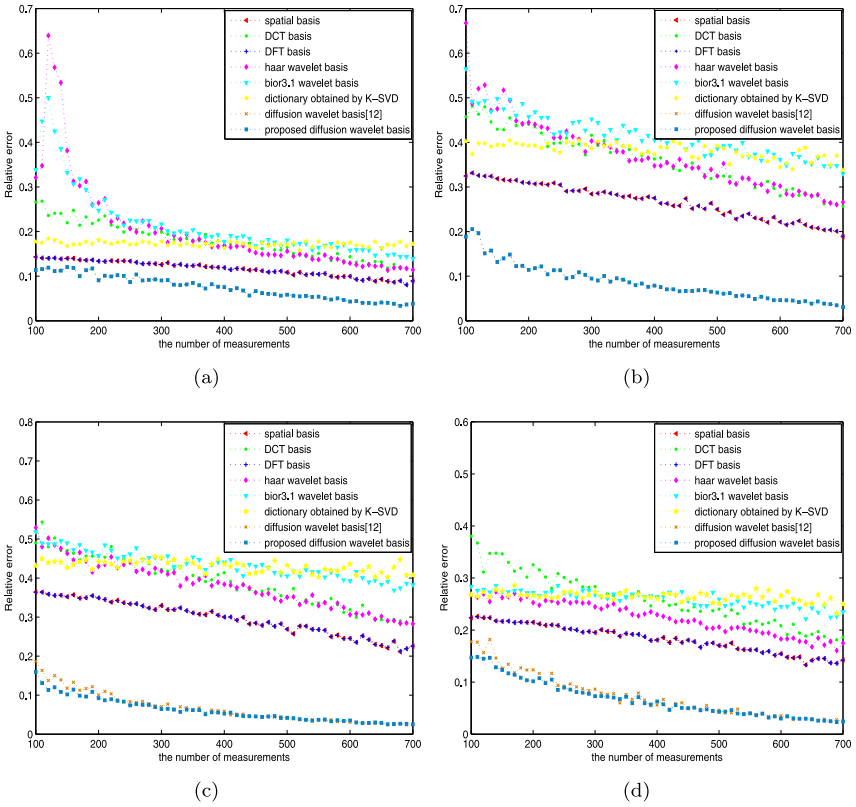


Fig. 21. The comparison results of relative error in eight sparsifying bases when the measurement matrix is sparse binary matrix with a fixed number of nonzero entries in each column (a) the recovery of sensory data in Fig. 16(a) without the measurement noise (b) the recovery of sensory data in Fig. 16(b) without the measurement noise (c) the recovery of sensory data in Fig. 16(c) without the measurement noise (d) the recovery of sensory data in Fig. 16(d) without the measurement noise.

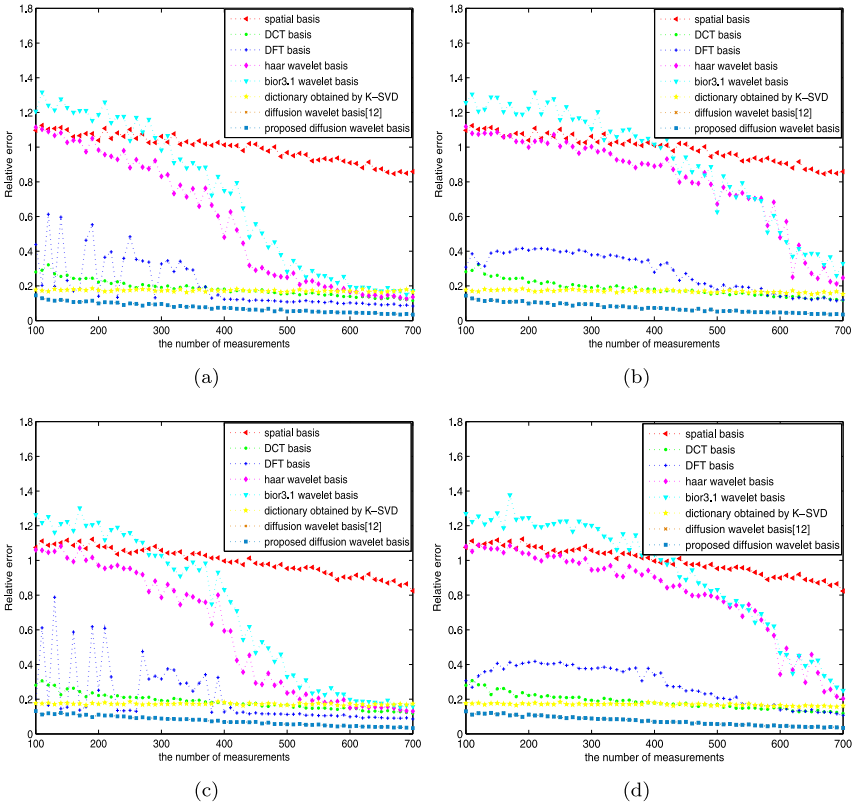


Fig. 22. The comparison results of relative error for sensory data in Fig. 16(a) when the measurement matrices are Gaussian random matrix and Bernoulli matrix (a)the recovery of sensory data without the measurement noise when the measurement matrix is Gaussian random matrix (b) the recovery of sensory data with the measurement noise ($\sigma_n = 0.05$) [36] when the measurement matrix is Gaussian random matrix (c) the recovery of sensory data without the measurement noise when the measurement matrix is Bernoulli matrix (d) the recovery of sensory data with the measurement noise ($\sigma_n = 0.05$) [36] when the measurement matrix is Bernoulli matrix.

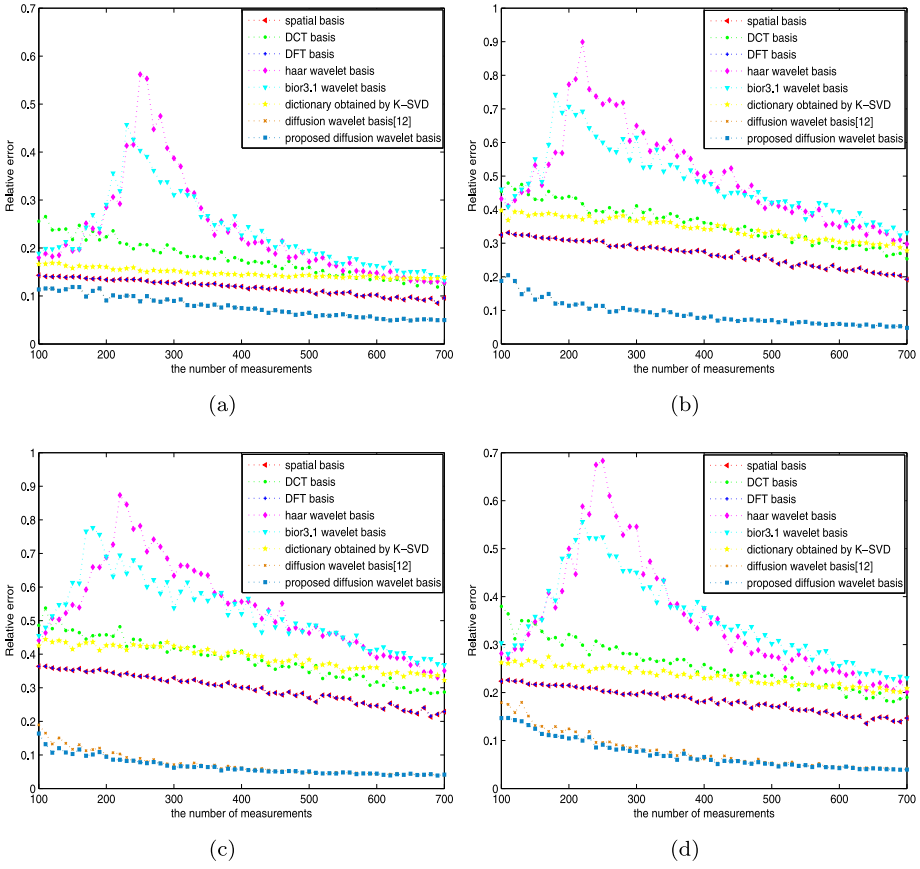


Fig. 23. The comparison results of relative error for sensory data in Fig. 16 when the measurement matrix is sparse binary matrix with a fixed number of nonzero entries in each column (a) the recovery of sensory data in Fig. 16(a) with the measurement noise($\sigma_n = 0.05$) [36] (b) the recovery of sensory data in Fig. 16(b) with the measurement noise($\sigma_n = 0.05$) [36] (c) the recovery of sensory data in Fig. 16(c) with the measurement noise($\sigma_n = 0.05$) [36] (d) the recovery of sensory data in Fig. 16(d) with the measurement noise($\sigma_n = 0.05$) [36].

Table 9
The relative error of the recovery of sensory data with the measurement noise($\sigma_n = 0.05$) [36] in eight sparsifying bases.

The sparsifying bases	M	Fig. 23(a)	Fig. 23(b)	Fig. 23(c)	Fig. 23(d)
The spatial basis	100	0.1432	0.3246	0.3640	0.2237
DCT basis	100	0.2557	0.4542	0.4859	0.3797
DFT basis	100	0.1432	0.3246	0.3639	0.2236
Haar wavelet basis	100	0.1799	0.4321	0.4410	0.2814
Bior3.1 wavelet basis	100	0.1887	0.4606	0.4549	0.3033
Dictionary obtained by K-SVD	100	0.1665	0.3973	0.4252	0.2623
Diffusion wavelet basis [12]	100	0.1135	0.1877	0.1896	0.1791
Proposed diffusion wavelet basis	100	0.1135	0.1877	0.1637	0.1468
The spatial basis	150	0.1400	0.3186	0.3521	0.2168
DCT basis	150	0.2177	0.4307	0.4721	0.3433
DFT basis	150	0.1400	0.3187	0.3521	0.2168
Haar wavelet basis	150	0.1943	0.5326	0.5467	0.3476
Bior3.1 wavelet basis	150	0.1984	0.5505	0.6130	0.3585
Dictionary obtained by K-SVD	150	0.1583	0.3865	0.4153	0.2648
Diffusion wavelet basis [12]	150	0.1188	0.1323	0.1169	0.1313
Proposed diffusion wavelet basis	150	0.1188	0.1323	0.1028	0.1246
The spatial basis	200	0.1338	0.3092	0.3492	0.2151
DCT basis	200	0.2234	0.4389	0.4577	0.3211
DFT basis	200	0.1339	0.3091	0.3491	0.2151
Haar wavelet basis	200	0.2852	0.7726	0.6883	0.4999
Bior3.1 wavelet basis	200	0.2908	0.7068	0.6907	0.4850
Dictionary obtained by K-SVD	200	0.1605	0.3795	0.4258	0.2584
Diffusion wavelet basis [12]	200	0.0908	0.1139	0.0978	0.1244
Proposed diffusion wavelet basis	200	0.0908	0.1139	0.0947	0.1044

(*continued on next page*)

Table 9 (*continued*)

The sparsifying bases	*M*	Fig. 23(a)	Fig. 23(b)	Fig. 23(c)	Fig. 23(d)
The spatial basis	250	0.1339	0.3049	0.3339	0.2028
DCT basis	250	0.2073	0.3934	0.4425	0.2975
DFT basis	250	0.1340	0.3050	0.3341	0.2028
Haar wavelet basis	250	0.5619	0.7141	0.7820	0.6830
Bior3.1 wavelet basis	250	0.4031	0.6001	0.6591	0.5240
Dictionary obtained by K-SVD	250	0.1527	0.3654	0.4247	0.2517
Diffusion wavelet basis [12]	250	0.0888	0.1133	0.0899	0.0960
Proposed diffusion wavelet basis	250	0.0888	0.1133	0.0782	0.0913
The spatial basis	300	0.1264	0.2845	0.3296	0.1956
DCT basis	300	0.1977	0.4115	0.4184	0.2801
DFT basis	300	0.1265	0.2846	0.3296	0.1956
Haar wavelet basis	300	0.3872	0.6497	0.6338	0.5456
Bior3.1 wavelet basis	300	0.3105	0.6147	0.5375	0.4505
Dictionary obtained by K-SVD	300	0.1531	0.3664	0.4246	0.2524
Diffusion wavelet basis [12]	300	0.0898	0.1001	0.0712	0.0880
Proposed diffusion wavelet basis	300	0.0898	0.1001	0.0620	0.0765

the same recovery performance as the proposed diffusion wavelet basis in the best case.

The paper only studies the construction of diffusion wavelet based sparsifying basis, and doesn't further research the design of the measurement matrix. For WSNs, the energy consumption and data gathering delay are also the important evaluation index. In the future study, the design of sparsifying basis and the measurement matrix are combined to improve the data gathering performance, which makes the work to be more meaningful.

Acknowledgments

We thank anonymous reviewers for careful and strict review, and the reviewers' comments improve our paper a lot. This work of the paper is supported by National Natural Science Foundation of China (No.61174016), Natural Science Foundation of Hebei Province of China (No.F2015202214), Hebei Province Science and Technology Support Program (No.15210506), and Natural Science Foundation of Tianjin (No.16JCQNJC00400).

References

[1] R.G. Baraniuk, Compressive sensing, IEEE Signal Process. Mag. 24 (4) (2008).
[2] J. Yick, B. Mukherjee, D. Ghosal, Wireless sensor network survey, Comput. Netw. 52 (12) (2008) 2292–2330.
[3] S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, Siam J. Sci. Comput. 20 (1) (1998) 33–61.
[4] I. Daubechies, M. Defrise, C.D. Mol, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint, Commun. Pure Appl. Math. 57 (11) (2004) 1413–1457.
[5] J. Haupt, W.U. Bajwa, M. Rabbat, R. Nowak, Compressed sensing for networked data, IEEE Signal Process. Mag. 25 (2) (2008) 92–101.
[6] J. He, G. Sun, Z. Li, Y. Zhang, Compressive data gathering with low-rank constraints for wireless sensor networks, Signal Process. 131 (2017) 73–76.
[7] W. Wang, M.J. Wainwright, K. Ramchandran, Information-theoretic limits on sparse signal recovery: dense versus sparse measurement matrices, IEEE Trans. Inf. Theory 56 (6) (2010) 2967–2979.
[8] H. Mamaghanian, N. Khaled, D. Atienza, P. Vandergheynst, Compressed sensing for real-time energy-efficient ecg compression on wireless body sensor nodes, IEEE Trans. Biomed. Eng. 58 (9) (2011) 2456–2466.
[9] W. Wang, M. Garofalakis, K. Ramchandran, Distributed sparse random projections for refinable approximation, in: International Symposium on Information Processing in Sensor Networks, 2007, pp. 331–339.
[10] S. Li, H. Qi, Distributed data aggregation for sparse recovery in wireless sensor networks, in: IEEE International Conference on Distributed Computing in Sensor Systems, 2013, pp. 62–69.
[11] R.R. Coifman, M. Maggioni, Diffusion wavelets, Appl. Comput. Harmon. Anal. 21 (1) (2006) 53–94.
[12] L. Xiang, J. Luo, C. Rosenberg, Compressed data aggregation: energy efficient and high fidelity data collection, IEEE/ACM Trans. Networking 21 (6) (2013) 1722–1735.
[13] Gudivada, S.K. Naidu, Applications of Diffusion Wavelets, University of York (2011).
[14] C. Lv, Q. Wang, W. Yan, R. Zhao, A sparse representation method of 2-d sensory data in wireless sensor networks, in: IEEE International Instrumentation and Measurement Technology Conference Proceedings, 2016.
[15] F. Chung, Spectral Graph Theory, 1997, p. 212.
[16] N. Hurley, S. Rickard, Comparing measures of sparsity, IEEE Trans. Inf. Theory 55 (10) (2008) 55–60.
[17] D. Zonoobi, A.A. Kassim, Y.V. Venkatesh, Gini index as sparsity measure for signal reconstruction from compressive samples, IEEE J. Sel. Topics Signal Process. 5 (5) (2011) 927–932.
[18] M.E. Lopes, Estimating unknown sparsity in compressed sensing, in: International Conference on Machine Learning, 2012, pp. 217–225.
[19] G.H. Golub, V.C. Klema, G.W. Stewart, Rank Degeneracy and Least Squares Problems, Stanford University, 1976.
[20] N. Salman, M. Ghogho, A.H. Kemp, Optimized low complexity sensor node positioning in wireless sensor networks, IEEE Sensors J. 14 (1) (2014) 39–46.
[21] Á. Lédeczi, M. Maróti, Wireless sensor node localization, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 370 (1958) (2012) 85–99.
[22] N.A. Alrajeh, M. Bashir, B. Shams, Localization techniques in wireless sensor networks, Int. J. Distrib. Sensor Netw. 17 (2013) (2014) 216–221.
[23] J. Zhao, W. Xi, Y. He, Y. Liu, X.Y. Li, L. Mo, Z. Yang, Localization of wireless sensor networks in the wild:pursuit of ranging quality, IEEE/ACM Trans. Netw. 21 (1) (2013) 311–323.
[24] X. Liu, J. Luo, C. Deng, A.V. Vasilakos, W. Lin, Dual-level compressed aggregation: recovering fields of physical quantities from incomplete sensory data, Comput. Sci. (2011) 182–190.
[25] L. Xiang, J. Luo, C. Deng, A.V. Vasilakos, Deca: recovering fields of physical quantities from incomplete sensory data, in: Sensor, Mesh and Ad Hoc Communications and Networks, Annual IEEE Communications Society Conference on, 2012, pp. 182–190.
[26] National oceanic and atmospheric administration's national weather service, 2016. (http://graphical.weather.gov/sectors/conus.php?element=T). [Online; accessed 5-December-2016].
[27] Temperature distribution in usa, 2016. (http://www.weather.gov).
[28] Sea surface temperature(100 km global), 2016. (http://www.class.ngdc.noaa.gov/saa/products/search?sub_id=0&datatype_family=SST100&submit.x=18&submit.y=7).
[29] M. Aharon, M. Elad, A. Bruckstein, K-svd: an algorithm for designing overcomplete dictionaries for sparse representation, IEEE Trans. Signal Process. 54 (11) (2006) 4311–4322.
[30] A. Tabibiazar, O. Basir, Energy-efficient compressive state recovery from sparsely noisy measurements, IEEE Trans. Instrum. Meas. 61 (61) (2012) 2392–2400.
[31] B. Bah, J. Tanner, Vanishingly sparse matrices and expander graphs, with application to compressed sensing, IEEE Trans. Inf. Theory 59 (11) (2013) 7491–7508.
[32] E. Candés, J. Romberg, Sparsity and incoherence in compressive sampling, Inverse Prob. 23 (3) (2006) 969–985. (17).
[33] C. Luo, F. Wu, J. Sun, C.W. Chen, Compressive data gathering for large-scale wireless sensor networks, in: Proceedings of the 15th Annual International Conference on Mobile Computing and Networking, ACM, 2009, pp. 145–156.
[34] D. Ebrahimi, C. Assi, Compressive data gathering using random projection for energy efficient wireless sensor networks, Ad Hoc Netw. 16 (2013) 105–119.
[35] D. Ebrahimi, C. Assi, Optimal and efficient algorithms for projection-based compressive data gathering, IEEE Commun. Lett. 17 (8) (2013) 1572–1575.
[36] D.M. Malioutov, S.R. Sanghavi, A.S. Willsky, Sequential compressed sensing, IEEE J. Sel. Topics Signal Process. 4 (2) (2010) 435–444.
[37] Noaa datasets, 2016. (http://www.noaa.gov/).