

Article

An Online Dictionary Learning-Based Compressive Data Gathering Algorithm in Wireless Sensor Networks

Donghao Wang, Jiangwen Wan *, Junying Chen and Qiang Zhang

School of Instrumentation Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China; dhwang@buaa.edu.cn (D.W.); woshicjy@126.com (J.C.); youleyuanzq@buaa.edu.cn (Q.Z.)

* Correspondence: jwwan@buaa.edu.cn; Tel.: +86-10-8233-9889

Academic Editor: Leonhard M. Reindl

Received: 30 May 2016; Accepted: 14 September 2016; Published: 22 September 2016

Abstract: To adapt to sense signals of enormous diversities and dynamics, and to decrease the reconstruction errors caused by ambient noise, a novel online dictionary learning method-based compressive data gathering (ODL-CDG) algorithm is proposed. The proposed dictionary is learned from a two-stage iterative procedure, alternately changing between a sparse coding step and a dictionary update step. The self-coherence of the learned dictionary is introduced as a penalty term during the dictionary update procedure. The dictionary is also constrained with sparse structure. It's theoretically demonstrated that the sensing matrix satisfies the restricted isometry property (RIP) with high probability. In addition, the lower bound of necessary number of measurements for compressive sensing (CS) reconstruction is given. Simulation results show that the proposed ODL-CDG algorithm can enhance the recovery accuracy in the presence of noise, and reduce the energy consumption in comparison with other dictionary based data gathering methods.

Keywords: wireless sensor networks; sparse representation; compressive sensing; data gathering; online dictionary learning

1. Introduction

Wireless sensor networks (WSNs) which are composed of lots of tiny, resource-constrained and cheap sensor nodes are self-organized networks. These nodes are always deployed in distributed mode to perform various applications, such as healthcare monitoring, transportation systems, industry service, and environmental measurement of humidity or temperature data [1]. In each case, efficient data gathering for target information is one of the primary missions.

In a typical scenario, WSNs have many ordinary sensor nodes and a base station named sink node. The ordinary nodes can only perform simple measurement and communication tasks since they are always equipped with limited power supply and most of the time, it is difficult to replace or recharge the battery. In contrast, the sink node is capable of performing complex operations since it is usually supplied with unlimited resources. Thus how to balance the energy consumption and develop energy efficient data collection protocols is still a research hotspot.

To reduce energy consumption for data gathering in WSNs, distributed source coding (DSC) [2] was proposed to compress the raw data between the ordinary nodes. DSC-based data collection protocols are composed of two important procedures. The first step is the collection of spatial-temporal correlation properties of the raw data. The second is a coding step which is based on the Slepian-Wolf coding theory. The coding process imposes no communication burden among sensor nodes, but the data correlation of the whole network must be calculated at the sink node before data collection, which results in a relatively high computational cost.

In recent years, compressive sensing has emerged as a new approach for signal acquisition, which can guarantee exact signal reconstruction with a small number of measurements [3]. Data compression and data collection are integrated into a single procedure in compressive sensing-based data gathering methods. Moreover, the high computational burdens are transferred to the base station. Finally, the incomplete data can be recovered by various complicated reconstruction algorithms at the sink node. Nevertheless, to ensure exact reconstruction, the key point is that the signals are required to be sparse in some base dictionary. Sparse representation expresses signals as sparse linear combinations of the basis. Therefore, dictionary learning for sparse signal representation is one of the core problems of compressive sensing.

The paper presents an ODL-CDG algorithm. The proposed algorithm aims to reduce the energy consumption for data gathering problem in WSNs. How to design ODL-CDG algorithm to be robust to environmental noise is also our objective. The main contributions of this paper can be summarized as follows:

- (1) Inspired by the periodicity of nature signals, the learned dictionary is constrained with sparse structure where each atom is a sparse linear combination of the base dictionary. We first apply the sparse structured dictionary in the compressive data gathering process.
- (2) The self-coherence of the learned dictionary is introduced as a penalty during the optimization procedure, which reduces the reconstruction error caused by ambient noise.
- (3) In respect of the sparse structured dictionary D and the Gaussian observation matrix Φ , it's theoretically demonstrated that the sensing matrix $P = \Phi D$ meets the property of RIP with very high probability. What's more, the lower bound of necessary number of measurements for exact reconstruction is given.
- (4) With these consideration, the online dictionary learning algorithm is designed to improve the adaptability of data gathering algorithm for a variety of practical applications. The training data is gathered in parallel with the compressive sensing procedure, which reduces the enormous energy consumption.

The remainder of this paper is organized as follows: in Section 2, we review the previous works involving dictionary learning and energy-efficient data gathering problem in wireless sensor networks. Section 3 presents the mathematical formulation of the problem in detail. Section 4 demonstrates the RIP property of the sensing matrix. In Section 5, the optimized solution of our proposed ODL-CDG problem is given and the convergence property is also analyzed. In Section 6, the performances of the ODL-CDG algorithm are verified on synthetic datasets and the real datasets by experimental results. Finally, conclusions are drawn and future work proposed in Section 7.

2. Related Work

In the past few years, much effort has gone into designing data gathering techniques with the aim of reducing energy consumption in WSNs. Luo et al. [4] first proposed a complete design for compressive sensing-based data gathering (CDG) in large scale wireless sensor networks. In CDG, the sensor readings are assumed to be spatially correlated. Moreover, the communication cost is reduced and the load balance is achieved simultaneously. Liu et al. [5] introduced a novel compressed sensing method called expanding window compressed sensing (EW-CS) to improve recovery quality for non-uniform compressible signals. Shen et al. [6] proposed non-uniform compressive sensing (NCS) for signal reconstruction in the WSNs. The spatio-temporal correlation of the sensed data and the network heterogeneity were both taken into consideration in NCS, which leads to significantly less samples. In [7], the authors presented a quantitative analysis of the primary energy consumption of WSNs. They pointed out that the compressed sensing and distributed compressed sensing may act as energy efficient sensing approaches in comparison with other compression techniques.

The abovementioned compressive sensing-based data gathering methods can relieve the energy shortage problem and prolong network lifespan. However, these methods have limitations in that the

signals are assumed only to be sparsely represented in a specified basis, etc., in a wavelet, Discrete Cosine Transform (DCT) or Fourier basis. Actually, a single predetermined basis may not be able to sparsely represent all types of signals, since there is a wide variety of applications for WSNs.

To adapt to signals of enormous diversity and dynamics, dictionary learning from a set of training signals has received lots of attention. The goal is to train a dictionary that can decompose the signals using a few atoms. The K-SVD [8] method is one of the well-known dictionary learning algorithms, which can lead to much more compact representation of signals. Duarte et al. [9] proposed to train the dictionary and optimize the sampling matrix simultaneously. The motivation is to make the mutual coherence between the dictionary and the projection matrix as minimal as possible. Christian et al. [10] presented a dictionary learning algorithm called IDL which made a trade-off between the coherence of the dictionary to the observations of signal class and the self-coherence of the dictionary atoms. To accelerate the convergence of K-SVD, an overcomplete dictionary was proposed in [11]. The authors suggested updating the atoms sequentially, thus leading to much better learning accuracy when compared with K-SVD. In [12], a new dictionary learning framework for distributed compressive sensing application was presented utilizing the data correlation between intra-nodes and inter-nodes, which resulted in improved compressive sensing (CS) performance.

However, the case where there is no access to the original data is not taken into account in the above work. What's more, obtaining the full original data may be costly in wireless sensor networks. That is our motivation to learn the dictionary from a compressive sensing approach. Studer et al. [13] investigated dictionary learning from sparsely corrupted signals or compressed signals. In [14], the authors further extended the problem of compressive dictionary learning based on sparse random projections. The idea was coming from their previous paper [15], where the compressive K-SVD (CK-SVD) algorithm was proposed to learn a dictionary using compressive sensing measurements. Aghagolzadeh et al. [16] associated the spatial diversity of compressive sensing measurements without additional structural constraints on the learned dictionary, which guaranteed the convergence to a unique solution with high probability.

Nevertheless, the environmental noise is not considered in the methods mentioned above. As analyzed in Section 3, the reconstruction error caused by environmental noise is positively correlated with the self-coherence of the learned dictionary. Thus, the self-coherence of the learned dictionary is added as a penalty term during the dictionary updating step. The novel dictionary is also imposed by structural constraints.

3. Problem Formulation

In this section, we introduce the related issues in respect to dictionary learning and compressive sensing theory. The final form of ODL-CDG problem is formulated in detail. The main notations of the paper are summarized in Table 1.

Table 1. Summary of notations.

M	Number of necessary measurements
N	Number of sensor nodes
K	Number of atoms of dictionary D
L	Length of training data vectors
λ_A	Sparse coefficient regularization parameter
λ_Θ	Structured dictionary regularization parameter
L_f	Lipschitz constant
Φ	Measurement matrix
Ψ	Orthonormal basis dictionary
P	Sensing matrix
D	Structured dictionary
X	Data matrix
\hat{X}	Estimated data matrix
Θ	Sparse atom representation dictionary

3.1. Compressive Sensing

Compressive sensing (CS) theory builds on the surprising revelation that a sparse signal can be recovered from a much smaller number of sampling values. Let $x \in \mathbb{R}^N$ be the original signal vector, which denotes sensor readings gathered in wireless sensor networks. Suppose $\Phi \in \mathbb{R}^{M \times N}$ ($M < N$) is the measurement matrix with independent and identically distributed Gaussian entries and unit norm columns. Thus the lower-dimensional linear measurement vector $y \in \mathbb{R}^M$ can be obtained from the following standard measurement model:

$$y = \Phi x + e \quad (1)$$

where $e \sim N(0, \sigma^2) \in \mathbb{R}^M$ is a white Gaussian noise vector.

Since sensor readings have spatial correlation, the signal vector x is assumed to be K -sparse in a given orthonormal basis $\Psi = [\Psi_1 \ \Psi_2 \ \dots \ \Psi_N]$, $\Psi_i \in \mathbb{R}^N$. That is:

$$x = \Psi \theta \quad (2)$$

where vector $\theta = [\theta_1, \theta_2, \dots, \theta_N]^T$ is the corresponding sparse coefficients, with the constraint of $\|\theta\|_0 = K \ll N$. The orthonormal basis Ψ can be constructed from various bases: DCT, wavelets, curvelets, etc.

As the number of equations M is much smaller than the number of variables N , the reconstruction of original signal x is an under-determined problem. An initial approach to solve the problem of recovering x depends on solving the following l_0 minimization problem:

$$\min_{\theta \in \mathbb{R}^N} \|\theta\|_0, \text{ s.t. } \|y - \Phi \Psi \theta\|_2 \leq \eta \quad (3)$$

where η is the expected noise on the measurements, $\|\bullet\|_0$ denotes the number of nonzero entries of vector θ , and $\|\bullet\|_2$ counts the standard Euclidean norm. The above problem is NP-hard, so it's numerically unstable to seek a global solution. To get an approximate solution, various greedy algorithms could be employed, like compressive sampling matching pursuit (CoSaMP) [17], Orthogonal Matching Pursuit with Replacement (OMPR) [18], and stagewise orthogonal matching pursuit (StOMP) [19].

Fortunately, the above problem is equivalent to the following l_1 minimization problem under certain conditions. Thus the recovery can be obtained using linear programming (LP) techniques, searching for resolution of:

$$\min_{\theta \in \mathbb{R}^N} \|\theta\|_1, \text{ s.t. } \|y - \Phi \Psi \theta\|_2 \leq \eta \quad (4)$$

If matrix $P = \Phi \Psi$ satisfies the RIP [20], the solutions of optimization Equations (3) and (4) coincide with each other. The definition of RIP is as follows:

Definition 1. (Restricted Isometry Property): Let $P = \Phi \Psi$ be an $M \times N$ matrix and let Θ be the sparse vector. The number of nonzero entries of vector Θ is no larger than K . Define the K -restricted isometry constant δ_K as the smallest constant that satisfies:

$$(1 - \delta_K) \|\theta\|_2^2 \leq \|\theta\|_2^2 \leq (1 + \delta_K) \|\theta\|_2^2 \quad (5)$$

Then the matrix $P = \Phi \Psi$ is said to satisfy the K -restricted isometry constant with the constant δ_K .

3.2. The Conventional Dictionary Learning Methods

Although exact recovery on account of a fixed sparse representation dictionary is guaranteed for inherently sparse signals or compressible signals, natural signals tend to have various types. Consequently, a single fixed dictionary would not be enough to sparsely represent all types of signals.

Hence, much work has been done to achieve sparse redundant dictionary using dictionary learning methods, since it can enhance their ability to adapt to different types of signals.

Let $\{\mathbf{x}^i\}_{i=1}^L$ denote training data for dictionary learning. $\mathbf{x}^i \in \mathbf{R}^N$ represents a data vector and L represents the amount of training data. Thus the data matrix $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^L] \in \mathbf{R}^{N \times L}$ is obtained. Then the general form of conventional dictionary learning methods can be rewritten as:

$$\min_{\mathbf{D}, \mathbf{C}} \|\mathbf{X} - \mathbf{DC}\|_F^2, \quad \text{s.t. } \forall i, \|\mathbf{c}_i\|_0 \leq S \quad (6)$$

where $\|\bullet\|_F$ represents matrix Frobenius norm, $\mathbf{D} \in \mathbf{R}^{N \times K}$ denotes the sparse redundant dictionary, and $\mathbf{C} \in \mathbf{R}^{K \times L}$ denotes the sparse matrix.

However, it is a fact that the original training data may not be available, or the cost for obtaining enough original data is high. In this paper, we are interested in training a sparse representation dictionary with only a few of CS measurements, which are linear projections of the original signals \mathbf{X} onto a random sensing matrix Φ . The problem of learning a dictionary $\mathbf{D} \in \mathbf{R}^{N \times K}$ from a series of linear and non-adaptive measurements is defined as:

$$\mathbf{y}^i = \Phi \mathbf{D} \mathbf{a}^i, \quad i = 1, \dots, L \quad (7)$$

with $\mathbf{y}^i \in \mathbf{R}^M$ representing the compressed version of \mathbf{x}^i and \mathbf{a}^i representing the sparse column vector of the sparse matrix.

3.3. Sparse Structured Dictionary

In the sparse dictionary model \mathbf{D} , it is assumed that each atom of the dictionary can be expressed as a sparse linear combination of few atoms in a fixed base dictionary Ψ . The dictionary is therefore expressed as:

$$\mathbf{D} = \Psi \Theta \quad (8)$$

where $\Theta \in \mathbf{R}^{N \times p}$ ($p \geq N$) is the atom representation dictionary which is assumed to be sparse. Obviously, the base dictionary Ψ should contain some prior knowledge about the training signals. The matrix Ψ itself can also act as the sparse representation dictionary, such as the (DCT) dictionary or the overcomplete dictionary learned through other methods. The sparse dictionary model can adapt to various signals by modifying the matrix Θ . As a general, with another Θ columns added to the base dictionary, the matrix Θ can be expressed as the following structure:

$$\Theta = [I_{N \times N} | \Sigma_{N \times N}] \quad (9)$$

where Σ is assumed to be sparse and normalized.

3.4. The Recovery Error Penalty

In practical applications, the compressive data gathering procedure is corrupted with ambient noise. As shown in Equation (1), \mathbf{e} represents the measurement noise. We employ the mean square error (MSE) to estimate the performance of reconstructing a sparse random vector θ in the presence of random Gaussian noise vector \mathbf{e} :

$$\text{MSE} = E_{\theta, \mathbf{e}} [\|\hat{\theta} - \theta\|_2^2] \quad (10)$$

where $\hat{\theta}$ denotes an estimation value of θ and $E_{\theta, \mathbf{e}}(\bullet)$ denotes the mathematical expectation concerning the joint distribution of the random vectors θ and \mathbf{e} . The well-known oracle estimator assumes that the position of non-zero entries of sparse vector θ is known as a priori knowledge. The prior support is defined as a set $\Gamma \subset \{1, 2, \dots, N\}$. Thus Equation (1) can be expressed in the following form:

$$\mathbf{y} = \Phi \mathbf{D} \mathbf{I}_{\Gamma} \theta_{\Gamma} + \mathbf{e} \quad (11)$$

where \mathbf{I}_Γ denotes the matrix obtained by just preserving the corresponding columns of the identity matrix on the support of Γ , and $\boldsymbol{\theta}_\Gamma \in \mathbf{R}^S$ denotes the vector obtained by deleting the set of entries out of the support Γ . Then the formulation of the oracle estimator is given as follows [21]:

$$\text{MSE}^{\text{oracle}} = \sigma^2 \text{Tr} \left[\left(\mathbf{I}_\Gamma^T \mathbf{D}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \mathbf{D} \mathbf{I}_\Gamma \right)^{-1} \right] \quad (12)$$

where $\text{Tr}(\bullet)$ denotes the trace of a matrix.

To mitigate the MSE caused by ambient noise, we consider making $\text{Tr} \left[\left(\mathbf{I}_\Gamma^T \mathbf{D}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \mathbf{D} \mathbf{I}_\Gamma \right)^{-1} \right]$ as small as possible. As theoretically discussed in [22], the term $\text{Tr} \left[\left(\mathbf{I}_\Gamma^T \mathbf{D}^T \boldsymbol{\Phi}^T \boldsymbol{\Phi} \mathbf{D} \mathbf{I}_\Gamma \right)^{-1} \right]$ is positively correlated with the self-coherence of the sparse dictionary. Therefore, the penalty term $\| \mathbf{D}^T \mathbf{D} - \mathbf{I} \|_F^2$ is introduced here to constrain the self-coherence of the learned dictionary.

3.5. The Final Form of ODL_CDG Problem

In this section, the method for training a sparse representation dictionary with only a few CS measurements is presented. The CS measurements can be arbitrary linear combinations of the original signals. The self-coherence of the learned dictionary is introduced to reduce the recovery error. Thus, in consideration of the sparse structured constraint and the self-coherence of the learned dictionary, the sparse dictionary model can be obtained from low dimensional CS measurements. Finally, the following optimization problem is obtained:

$$\min_{\mathbf{A}, \mathbf{D}} \left\{ \frac{1}{2} \| \mathbf{Y} - \boldsymbol{\Phi} \mathbf{D} \mathbf{A} \|_F^2 + \| \mathbf{D}^T \mathbf{D} - \mathbf{I} \|_F^2 + \lambda_A \| \mathbf{A} \|_1 \right\}, \quad \text{s.t. } \| \boldsymbol{\theta} \|_1 \leq \varepsilon_\theta \quad (13)$$

4. Necessary Guarantees for Signal Reconstruction

Cai et al. [23] proved that Basis Pursuit algorithm could guarantee the reconstruction of Equation (4) in the following theorem:

Theorem 1. Assume that the measurement matrix $\boldsymbol{\Phi}$ satisfies $\delta_{2S}(\boldsymbol{\Phi}) < 1/\sqrt{2}$ for some $S \in \mathbb{N}$. Let $\boldsymbol{\theta}$ be a S -sparse vector. Then, the recovery error of Equation (4) is bounded by the noise level.

The reconstruction of signals that are sparse in an orthonormal basis is guaranteed by the above theorem. Nevertheless, we mainly stress the problem that signals are not sparse in an orthonormal basis but rather in a redundant dictionary $\mathbf{D} \in \mathbf{R}^{N \times 2N}$, as described in Sections 3.2 and 3.3. For the sake of elaboration convenience, the following two lemmas are given:

Lemma 1. Let entries of $\boldsymbol{\Phi} \in \mathbf{R}^{M \times N}$ be independent normal variables with mean zero and variance n^{-1} . Let \mathbf{D}_Λ , i.e., $|\Lambda| = S$, be the submatrix extracted from columns of the redundant matrix \mathbf{D} . Define the isometry constant $\delta_\Lambda = \delta_\Lambda(\mathbf{D})$ and $v := \delta_\Lambda + \delta + \delta_\Lambda \delta$ for $0 < \delta < 1$. Then:

$$(1 - v) \| \boldsymbol{\theta} \|_2^2 \leq \| \boldsymbol{\Phi} \mathbf{D}_\Lambda \boldsymbol{\theta} \|_2^2 \leq (1 + v) \| \boldsymbol{\theta} \|_2^2 \quad (14)$$

with probability exceeding:

$$1 - 2 \left(1 + \frac{12}{\delta} \right)^S e^{-\frac{c}{9} \delta^2 M} \quad (15)$$

where c is a positive constant, in particular c equals $7/18$ for the Gaussian matrix $\boldsymbol{\Phi}$.

Proof. The proof of lemma 1 can be found in [24]. \square

Lemma 2. The restricted isometry constant of a redundant dictionary \mathbf{D} with coherence μ is bounded with:

$$\delta_S(\mathbf{D}) \leq (S-1)\mu \quad (16)$$

Proof. This can be obtained from the proof [25]. \square

Theorem 2. Assume that the structure of our redundant dictionary is $\mathbf{D} = \mathbf{\Psi}\mathbf{\Theta} = \mathbf{\Psi}[\mathbf{I}_{N \times N}\mathbf{\Sigma}_{N \times N}]$, where $\mathbf{\Psi}$ is the orthogonal base dictionary, i.e., discrete cosine transform (DCT) based on \mathbf{R}^N for $N = 2^{2p+1}$. The number of atoms is $K = 2^{2p+2}$. Suppose that the sparsity of the signal is smaller than 2^{p-4} , the necessary sampling number that could guarantee signal reconstruction is obtained by:

$$M \geq C_1 (4S (2p \log 2 - \log S) + C_2 + t) \quad (17)$$

with the constants $C_1 \approx 524.33$ and $C_2 \approx 5.75$.

Proof. For $t > 0$, assume that the local isometry constant of matrix $\mathbf{P} = \mathbf{\Phi}\mathbf{\Psi}$ is $\delta_\Lambda(\mathbf{P})$, which is no larger than $\delta_\Lambda(\mathbf{D}) + \delta + \delta_\Lambda(\mathbf{D})\delta$ with probability at least $1 - e^{-t}$. \square

By Lemma 1, we obtain that:

$$\begin{aligned} P(\delta_\Lambda(\mathbf{P}) > \delta_\Lambda(\mathbf{D}) + \delta + \delta_\Lambda(\mathbf{D})\delta) \\ \leq 2 \left(1 + \frac{12}{\delta}\right)^S e^{-\frac{\epsilon}{9}\delta^2 M} \end{aligned}$$

Thus the global isometry constant $\delta_S(\mathbf{A}) := \sup_{|\Lambda|=S} \delta_\Lambda(\mathbf{A})$, $S \in N$ is bounded over all $\binom{K}{S}$ possibilities. So:

$$\begin{aligned} P(\delta_S(\mathbf{P}) > \delta_S(\mathbf{D}) + \delta + \delta_S(\mathbf{D})\delta) \\ \leq 2 \binom{K}{S} \left(1 + \frac{12}{\delta}\right)^S e^{-\frac{\epsilon}{9}\delta^2 M} \end{aligned}$$

Using the Stirling formula, and confining the above term to less than e^{-t} , the following inequation is obtained:

$$\begin{aligned} 2 \left(\frac{eK}{S}\right)^S \left(1 + \frac{12}{\delta}\right)^S e^{-\frac{\epsilon}{9}\delta^2 M} &< e^{-t} \\ \Rightarrow 2 \left(\frac{K}{S}\right)^S \left(e \left(1 + \frac{12}{\delta}\right)\right)^S &< e^{-t + \frac{\epsilon}{9}\delta^2 M} \\ \Rightarrow \log 2 + S \log \left(\frac{K}{S}\right) + S \log \left(e \left(1 + \frac{12}{\delta}\right)\right) &< -t + \frac{\epsilon}{9}\delta^2 M \\ \Rightarrow \frac{\epsilon}{9}\delta^2 M \geq S \log \left(\frac{K}{S}\right) + S \log \left(e \left(1 + \frac{12}{\delta}\right)\right) \log 2 + t \\ \Rightarrow M \geq 9\delta^{-2}c^{-1} \left(S \log \left(\frac{K}{S}\right) + S \log \left(e \left(1 + \frac{12}{\delta}\right)\right) \log 2 + t\right) \\ \Rightarrow M \geq 9\delta^{-2}c^{-1} \left(S \log \left(\frac{K}{S}\right) + \log \left(2e \left(1 + \frac{12}{\delta}\right)\right) + t\right) \\ \Rightarrow M \geq C_1 (S \log \left(\frac{K}{S}\right) + C_2 + t) \end{aligned}$$

The above theoretical derivation states that $\delta_S(\mathbf{P})$ is less than $\delta_S(\mathbf{D}) + \delta + \delta_S(\mathbf{D})\delta$ with probability at least $1 - e^{-t}$ when $M \geq C_1 \left(S \log \left(\frac{K}{S}\right) + C_2 + t\right)$.

Let μ be the coherence of dictionary \mathbf{D} , we assume:

$$S-1 \leq \frac{1}{10}\mu^{-1} \quad (18)$$

Then combining with Lemma 2, we can obtain:

$$\delta_S(\mathbf{D}) \leq (S-1)\mu \leq \frac{1}{10} \quad (19)$$

Thus, defining $\delta = 7/33$ yields:

$$\begin{aligned} \delta_S(\mathbf{P}) &\leq \delta_S(\mathbf{D}) + \delta + \delta_S(\mathbf{D})\delta \\ &\leq \frac{1}{10} + \frac{7}{33} \left(1 + \frac{1}{10}\right) = \frac{1}{3} \end{aligned} \quad (20)$$

As demonstrated by Theorem 1, the necessary number of samples to have $\delta_{2S}(\Phi) < 1/\sqrt{2}$ is:

$$M \geq C_1 (\text{Slog}(K/s) + C_2 + t) \quad (21)$$

Replacing S in Equation (21) with $2S$, and plugging $K = 2^{2p+2}$ and $\delta = 7/33$ into Equation (21), the necessary number of samples is finally available. That is:

$$M \geq C_1 (2S(2p\log 2 + \log 2 - \log S) + C_2 + t) \quad (22)$$

5. The Solution of the ODL-CDG Algorithm

Considering the demonstration in Section 4, we conclude that the ODL-CDG algorithm is feasible with a sufficient number of measurements. To get the solution by existing methods, the optimization Problem (13) is reformulated as an unconstrained optimization problem. Therefore, the cost objective function of final form of ODL-CDG problem is:

$$\min_{\mathbf{A}, \Theta} \left\{ \frac{1}{2} \|\mathbf{Y} - \Phi \Psi \Theta \mathbf{A}\|_F^2 + \|(\Psi \Theta)^T (\Psi \Theta) - \mathbf{I}\|_F^2 + \lambda_A \|\mathbf{A}\|_1 + \lambda_\Theta \|\Theta\|_1 \right\} \quad (23)$$

where λ_A and λ_Θ represent the sparse coefficient regularization parameter and the structured dictionary regularization parameter, respectively. The ODL-CDG algorithm is solved in a two-step iterative approach, which alternates between sparse coding and dictionary update procedures.

5.1. Sparse Coding

In the sparse coding step, sparse coefficients are obtained using the dictionary \mathbf{D} which is computed from the previous step:

$$\min_{\mathbf{A}} \left\{ \frac{1}{2} \|\mathbf{Y} - \Phi \Psi \Theta \mathbf{A}\|_F^2 + \lambda_A \|\mathbf{A}\|_1 \right\} \quad (24)$$

The above optimization Problem (24) can be successfully computed using Matching Pursuit LASSO (MPL) [26]. MPL can greatly speed up the convergence of the Problem (24) when employed in batch-mode.

5.2. Dictionary Update

In the dictionary update step, with the consideration of the sparsity constraint on the dictionary, the following optimization equation is obtained:

$$\min_{\Theta} \left\{ \frac{1}{2} \|\mathbf{Y} - \Phi \Psi \Theta \mathbf{A}\|_F^2 + \|(\Psi \Theta)^T (\Psi \Theta) - \mathbf{I}\|_F^2 + \lambda_\Theta \|\Theta\|_1 \right\} \quad (25)$$

The object function in Equation (25) can be rewritten into the following form:

$$\min_{\Theta} F(\Theta) = f(\Theta) + g(\Theta) \quad (26)$$

with:

$$f(\Theta) = \frac{1}{2} \|Y - \Phi \Psi \Theta A\|_F^2 + \|(\Psi \Theta)^T (\Psi \Theta) - I\|_F^2 \text{ and } g(\Theta) = \lambda_\Theta \|\Theta\|_1 \quad (27)$$

Obviously, the accelerated proximal gradient (APG) approach [27,28] can be used to solve the above unconstrained non-smooth convex problem, where both f , and g are convex. Furthermore, f is Lipschitz continuous:

$$\|\nabla f(\Theta_1) - \nabla f(\Theta_2)\|_2 \leq L_f \|\Theta_1 - \Theta_2\|_2, \forall \Theta_1, \Theta_2 \in \mathbb{R}^{N \times p} \quad (28)$$

where $\|\bullet\|_2$ denotes the standard Euclidean norm and $L_f > 0$ is the Lipschitz constant of ∇f .

Since $g(\Theta)$ is nonsmooth, it is difficult to directly minimize the objective function $F(\Theta)$. Instead, $F(\Theta)$ is approximated locally as a quadratic function at point Y_k and we try to repeatedly solve:

$$\Theta_{k+1} = \arg \min_{\Theta \in \mathbb{R}^{N \times p}} Q(\Theta, Y_k) \doteq f(Y_k) + \langle \nabla f(Y_k), \Theta - Y_k \rangle + \frac{L_f}{2} \|\Theta - Y_k\|_F^2 + g(\Theta) \quad (29)$$

For convenience, let $S_\lambda(y)$ denote the soft-thresholding operator [29], which is defined as follows:

$$S_\lambda(y) = \begin{cases} y - \varepsilon, & \text{if } y > \varepsilon \\ y + \varepsilon, & \text{if } y < -\varepsilon \\ 0, & \text{otherwise} \end{cases} \quad (30)$$

where $y \in \mathbb{R}$ and $\varepsilon > 0$. This operator can be also useful when applied elementwise to vectors and matrices.

Let $G \in \mathbb{R}^{N \times p}$. Then:

$$S_{\lambda_\Theta}(G) = \arg \min_{\Theta \in \mathbb{R}^{N \times p}} \left\{ \frac{1}{2} \|\Theta - G\|_F^2 + \lambda_\Theta \|\Theta\|_1 \right\} \quad (31)$$

where $S_{\lambda_\Theta}(G)$ is the soft-thresholding operator, as defined in Equation (30).

Proposition 1. Let $Y_k \in \mathbb{R}^{N \times p}$, $g(\Theta) = \lambda_\Theta \|\Theta\|_1$. Assume $f(\Theta)$ is Lipschitz continuous with a positive constant L_f , then we have:

$$\arg \min_{\Theta \in \mathbb{R}^{N \times p}} Q(\Theta, Y_k) = S_{\lambda_{L_f}^{-1}}(G_k) \quad (32)$$

where $G_k = Y_k - \frac{1}{L_f} \nabla f(Y_k)$.

Proof.

$$\begin{aligned} Q(\Theta, Y_k) &\doteq f(Y_k) + \langle \nabla f(Y_k), \Theta - Y_k \rangle + \frac{L_f}{2} \|\Theta - Y_k\|_F^2 + g(\Theta) \\ &= f(Y_k) + \langle \nabla f(Y_k), \Theta - Y_k \rangle + \frac{L_f}{2} \|\Theta - Y_k\|_F^2 + \lambda_\Theta \|\Theta\|_1 \\ &= \frac{L_f}{2} \|\Theta - (Y_k - \frac{1}{L_f} \nabla f(Y_k))\|_F^2 - \frac{1}{2L_f} \|\nabla f(Y_k)\|_F^2 + f(Y_k) + \lambda_\Theta \|\Theta\|_1 \\ &= \frac{L_f}{2} \|\Theta - G\|_F^2 + \lambda_\Theta \|\Theta\|_1 + f(Y_k) - \frac{1}{2L_f} \|\nabla f(Y_k)\|_F^2 \end{aligned} \quad (33)$$

□

Then combined with Equation (32), the final solution is obtained:

$$\begin{aligned}
 \Theta_{k+1} &= \arg \min_{\Theta \in \mathbb{R}^{N \times p}} Q(\Theta, Y_k) \\
 &= \arg \min_{\Theta \in \mathbb{R}^{N \times p}} \left\{ \frac{L_f}{2} \|\Theta - G\|_F^2 + \lambda_\Theta \|\Theta\|_1 + f(Y_k) - \frac{1}{2L_f} \|\nabla f(Y_k)\|_F^2 \right\} \\
 &= \arg \min_{\Theta \in \mathbb{R}^{N \times p}} \left\{ \frac{L_f}{2} \|\Theta - G\|_F^2 + \lambda_\Theta \|\Theta\|_1 \right\} \\
 &= S_{\lambda L_f^{-1}}(G_k)
 \end{aligned} \tag{34}$$

However, it is not always easy to compute the Lipschitz constant L_f . The APG algorithm with a backtracking stepsize rule is employed in Algorithm 1. We appoint an initial estimate value of L_f and increase the estimate gradually until the violation rule is reached.

Finally the pseudo-code of the ODL-CDG algorithm is shown in Algorithm 2.

Algorithm 1. APG with backtracking.

Initialization: Let $L_0 > 0, \eta > 1$

While not converged **do**

1: Find the smallest nonnegative interger i_k with $\bar{L} = \eta^{i_k} L_{k-1}$ ($k \geq 1$) to satisfy

$$F(S_{\lambda \bar{L}}(G_k)) \leq Q(S_{\lambda \bar{L}}(G_k), Y_k)$$

2: Set $L_k = \eta^{i_k} L_{k-1}$ ($k \geq 1$) and compute

$$Y_k \leftarrow X_k + \frac{t_{k-1}-1}{t_k} (X_k - X_{k-1})$$

$$\Theta_{k+1} \leftarrow \arg \min_{\Theta} Q_{L_f}(\Theta, Y_k)$$

$$t_{k+1} \leftarrow \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$k \leftarrow k + 1$$

End While

Algorithm 2. ODL-CDG Algorithm.

Input: $Y, \Phi, \Psi, \lambda_A, \lambda_\Theta, T, \varepsilon_{stop}$

Output: D, \hat{X}

Main procedure:

While $t < T$ and $\|\Theta_{k+1} - \Theta_k\|_2 > \varepsilon_{stop}$ **do**

Sparse Coding using MPL:

$$\min_{A^{(t)}} \left\{ \frac{1}{2} \|Y - \Phi \Psi \Theta^{(t-1)} A^{(t)}\|_F^2 + \lambda_A \|A^{(t)}\|_1 \right\}$$

Dictionary Update using APG (see Algorithm 1)

$$\min_{\Theta^{(t)}} Q(\Theta, Y) = \arg \min_{\Theta \in \mathbb{R}^{N \times p}} \left\{ \frac{L_f}{2} \|\Theta^{(t)} - G^{(t-1)}\|_F^2 + \lambda_\Theta \|\Theta^{(t)}\|_1 \right\}$$

$t \leftarrow t + 1$

End while

$$\text{Then } D = \Psi [I_{N \times N} | \Sigma_{N \times N}] = [\Psi \ \Psi \Sigma]$$

$$\text{Compute } \hat{A}: \min_{\hat{A}} \left\{ \frac{1}{2} \|Y - \Phi D \hat{A}\|_F^2 + \lambda_A \|\hat{A}\|_1 \right\}$$

$$\hat{X} = D \hat{A}$$

5.3. Convergence Analysis

As described above, the ODL-CDG algorithm contains the sparse coding step and the dictionary learning step. In the sparse coding step, the convergence of the optimization Problem (24) is guaranteed by MPL [26]. Furthermore, in the dictionary update step, the sequence of function values $F(\Theta_k)$ produced by APG is non-increasing, since the Lipschitz constant L_f satisfies $L_0 \leq L_{f(k)} \leq \eta L_f$ for every $k \geq 1$. The convergence rate of APG with the backtracking rule is demonstrated as $O(k^{-2})$, that is to say $F(\Theta_k) - F(\Theta^*) \leq Ck^{-2}$ [27]. What's more, the convergence of the alternating minimization method is also studied in [11]. Therefore, the convergence of ODL-CDG algorithm can be guaranteed.

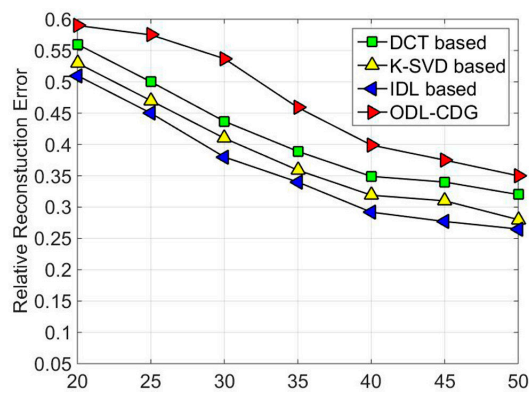
6. Simulation

This section presents our simulation results on synthetic data and the real datasets. The performance of the proposed dictionary is compared with a pre-specified dictionary, like the DCT dictionary, and other dictionary learning approaches, such as K-SVD, IDL and CK-SVD.

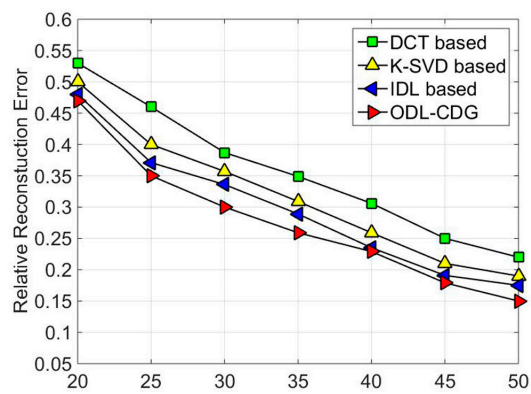
6.1. Recovery Accuracy

The initial basis Ψ is a 50×50 DCT matrix. A set of training signals $\{y_i\}_{i=1}^L$ is generated by a linear combination of the original synthetic data. The process is accomplished by applying a projection matrix Φ with independent and identically distributed Gaussian entries and column normalization. Input parameters to Algorithm 1 are $\lambda_A = 0.1$, $\lambda_\Theta = 0.05$, $\varepsilon_{stop} = 0.001$ and $T = 100$.

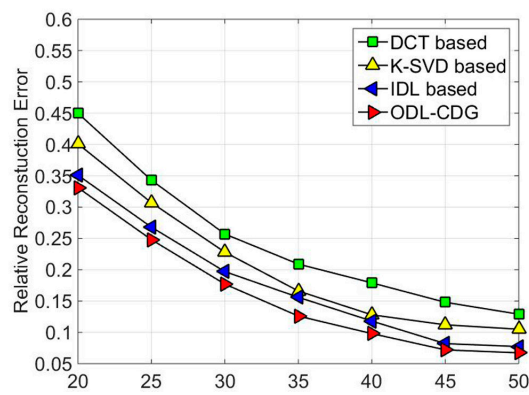
The performance is evaluated by using the relative reconstruction error, i.e., $\frac{\|\hat{X} - X\|_F}{\|X\|_F}$, where X and \hat{X} denote the original signal and the reconstructed signal, respectively. Each setting is averaged for 50 trials. The simulation results are presented in Figures 1 and 2. In Figure 1, each subgraph corresponds to a certain amount of sampling ratio. The signals are added with white Gaussian noise, which yields the signal-to-noise ratio (SNR) to range from 20 dB to 50 dB. As can be seen from Figure 1a, ODL has poor performance when the sampling ratio is low, but the ODL dictionary outperforms both the DCT dictionary and the K-SVD method in the relative reconstruction error, when the sampling ratio is high (high than 20%). The fixed dictionary, DCT, is the worst case. That's because the DCT dictionary using a fixed structure cannot sparsely represent synthetic data of various diversities. In comparison, K-SVD is better than the DCT dictionary. This is because K-SVD can adapt to sparsely represent the synthetic data by training. Since the IDL algorithm trains the dictionary using the self-coherence constraint term, the relative reconstruction error is smaller than DCT and K-SVD. Similar results can be obtained from Figure 2, where the relative reconstruction errors of DCT, K-SVD, IDL and ODL are obtained with sampling ratios of 10%, 15%, 20%, 25%, 30% and 40%, respectively. As we can see in our simulations, the results of ODL are much worse than DCT, K-SVD, and IDL when the sampling ratios is quite low (less than 10%). But ODL outperforms these algorithms compared by gradually increasing the sampling ratio.



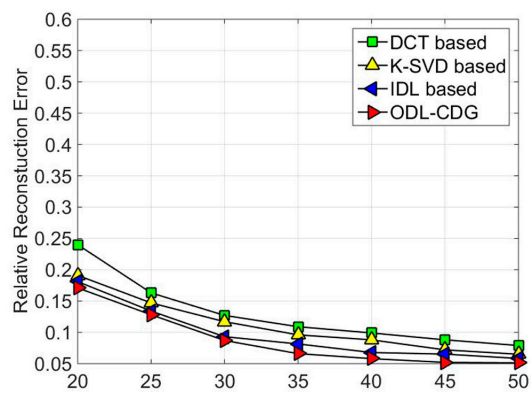
(a) Signal to noise ratio (dB) (Sampling ratio = 10%)



(b) Signal to noise ratio (dB) (Sampling ratio = 20%)

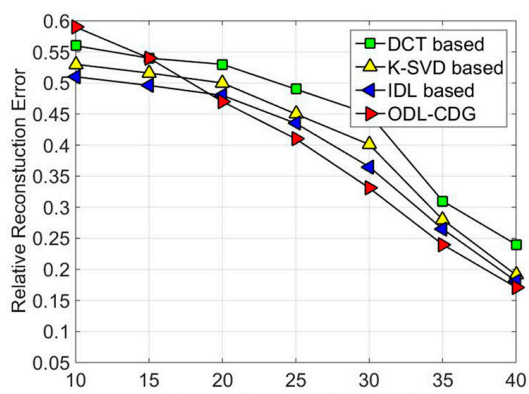


(c) Signal to noise ratio (dB) (Sampling ratio = 30%)

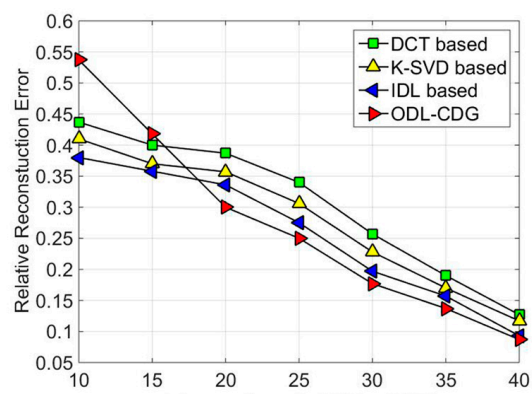


(d) Signal to noise ratio (dB) (Sampling ratio = 40%)

Figure 1. The relative reconstruction error of DCT, K-SVD, IDL and ODL-CDG under different signal-to-noise ratio: (a) Sampling Ratio = 10%; (b) Sampling Ratio = 20%; (c) Sampling Ratio = 30%; (d) Sampling Ratio = 40%.



(a) Sampling ratio (SNR = 20 dB)



(b) Sampling ratio (SNR = 30 dB)

Figure 2. Cont.

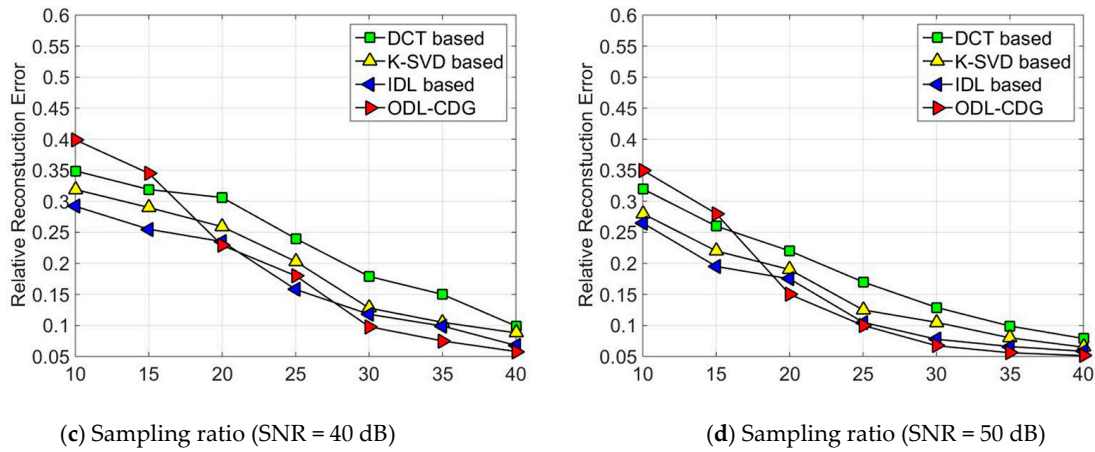


Figure 2. The relative reconstruction error of DCT, K-SVD, IDL and ODL-CDG under different sampling ratio: (a) SNR = 20 dB; (b) SNR = 30 dB; (c) SNR = 40 dB; (d) SNR = 50 dB.

6.2. Impact of Regularization Parameters on Sparse Representation Error

The performance of ODL-CDG algorithm may also be highly influenced by the setting value of regularization parameters λ_A and λ_θ . In this experiment, we analyze how the selected regularization parameters affect the sparse representation error. The optimal parameters for ODL-CDG algorithm is determined. The datasets used in this section are collected from the IntelLab [30]. We select the temperature values and the humidity values of size 54×100 between 28 February and 27 March 2004. The time interval is 31 s. We first solve the following sparse representation problem on training data $\{x^i\}_{i=1}^L$:

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^L}{\operatorname{argmin}} \|D\hat{\theta} - x^i\|_2, \text{ s.t. } \|\theta\|_0 \leq S \quad (35)$$

where S denotes the sparsity of the coefficient θ .

Then, the sparse representation error of the learned dictionary is evaluated using the root mean square error (RMSE), which is defined as follows:

$$RMSE = \frac{1}{\sqrt{L}} \sum_{i=1}^L \|D\hat{\theta} - x^i\|_2 \quad (36)$$

where L is the amount of training data. We average the experiment 50 times for every training data vector. Figure 3 shows the simulation results. In general, the sparse representation error is becoming larger as the parameter λ_A gradually increases. That's because regularization parameter λ_A determines the sparsity of the sparse coefficient. To constrain the coefficients to be sparser, we need to increase λ_A to a specific threshold. However, as can be seen from Figure 3, the sparse representation error increases tremendously as λ_A exceeds 0.1. In Figure 4, parameter λ_θ shows that it has the same trend as λ_A in impacting the sparse representation error. Based on the above discussion, we set the regularization parameter as a relatively small value, such as $\lambda_A = 0.1$ and $\lambda_\theta = 0.05$. These are also the optimal parameter values we set in Section 6.1.

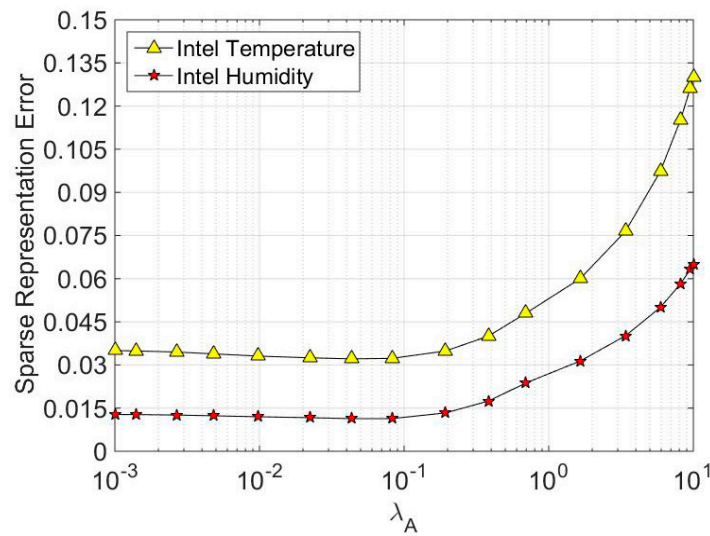


Figure 3. The impact of sparse coefficient regularization parameter λ_A on sparse representation error.

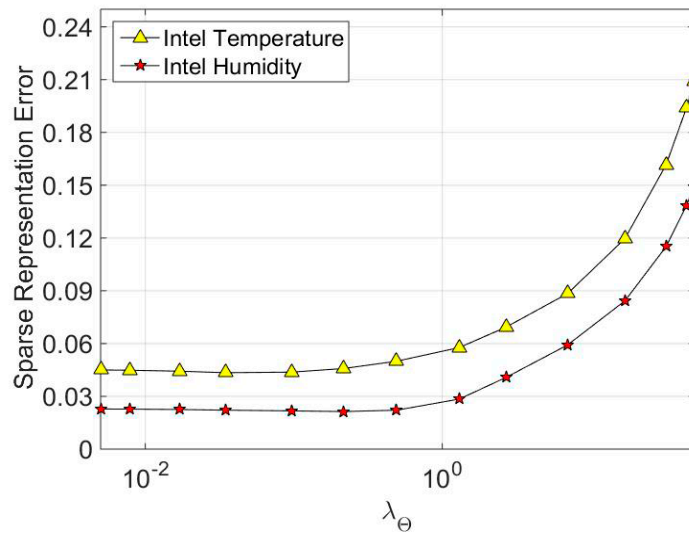


Figure 4. The impact of structured dictionary regularization parameter λ_Θ on sparse representation error.

6.3. Energy Consumption Comparison

In this subsection, we simulate the energy consumption of the ODL-CDG algorithm. The simulation platform is MATLAB. Suppose 500 nodes are randomly deployed in a 1000 m × 1000 m area and the sink node is deployed in the center. The random topology of these sensor nodes is shown in Figure 5. The communication range is 50 m and the initial energy is 2 J. The original data used in this section is synthesized of multiple data sets. Thus they cannot be sparsely represented in a predefined dictionary. To evaluate the energy consumption of ODL-CDG, we employ the same energy model in [31]:

$$E_{trans} = \begin{cases} l \times (E_{elec} + E_{mp} \times d^2), & \text{if } d \leq d_{Thres} \\ l \times (E_{elec} + E_{mp} \times d^4), & \text{if } d > d_{Thres} \end{cases} \quad (37)$$

$$E_{rec} = l \times E_{elec} \quad (38)$$

where E_{trans} denotes the energy consumption of transmitting l bits of data to another node within distance d , E_{rec} denotes the energy consumption of receiving l bits of data, E_{elec} denotes the energy consumption of the modular circuit, and E_{mp} denotes the energy consumed by the power amplifying

circuit. The parameters input to the ODL-CDG algorithm are the same as in Section 6.1 and the related parameters are listed in Table 2.

Table 2. Experimental parameters.

Parameter Name	Value
Node number	500
Transmission range	50 m
Initial energy	2 J
Data Size	1024 bit
E_{elec}	50 nJ/bit
E_{amp}	0.1 nJ/(bit·m ²)

It is regarded as a successful reconstruction when the relative reconstruction error is smaller than 0.1. Figure 6 shows the energy consumption of ODL-CDG algorithm compared with other dictionary learning-based data gathering methods. Note that the K-SVD-based data gathering method requires one to access the whole data, so the original training data should be transmitted to the sink node by multi-hop paths. Moreover, the K-SVD dictionary should be updated in time since the synthesized data contains large diversities. Thus, the initial step of dictionary learning before data gathering may consume large amounts of energy. Therefore, the energy consumption of the K-SVD-based data gathering method is significantly larger than that of CK-SVD and the ODL-CDG. Figure 6 shows that ODL-CDG algorithm achieves the best energy savings. That's because the dictionary in the ODL-CDG algorithm is learned in the process of compressive data gathering, which greatly reduces the energy consumption for raw data transmission through the entire network. Similarly, the total energy consumption of CK-SVD is enlarged with the increase of successful reconstruction number. But its energy consumption is still high than ODL-CDG as can be seen from Figure 6. The reason is that the introduced self-coherence penalty term can restrain the reconstruction error in the ODL-CDG algorithm, so the ODL-CDG algorithm should collect much fewer CS measurements than the CK-SVD-based data gathering method for the same reconstruction accuracy.

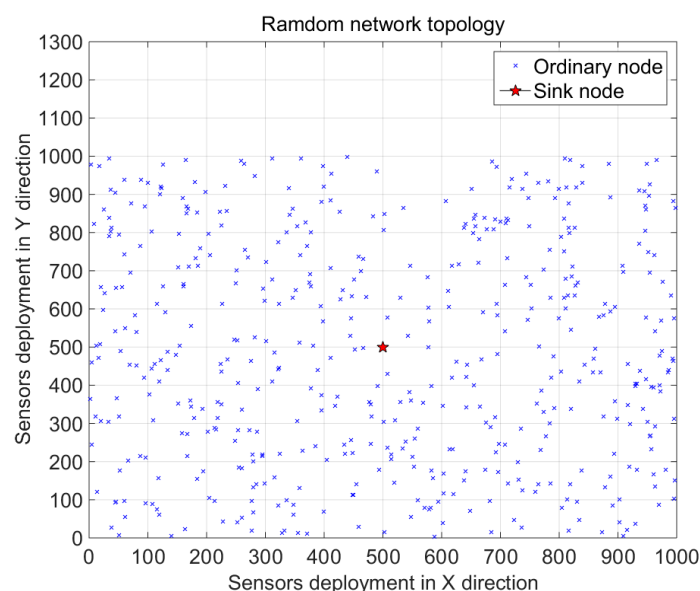


Figure 5. Random deployment of 500 sensor nodes.

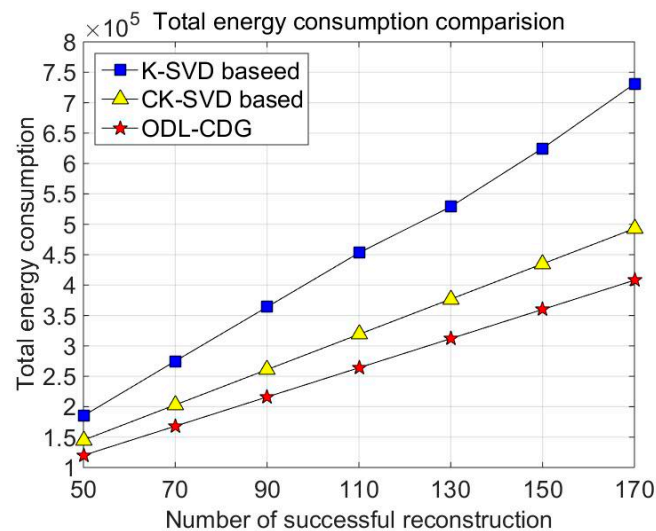


Figure 6. The total energy consumption of different dictionary learning based data gathering methods.

In Figure 7, the impact of different dictionary learning-based data gathering methods on the lifespan of nodes is studied. The node is supposed to survive when its energy is higher than zero. As can be seen from Figure 7, the ODL-CDG algorithm outperforms the other methods since the proposed dictionary has better adaptability to various signals. Thus the ODL-CDG algorithm reduces the energy consumption and prolongs the network lifespan.

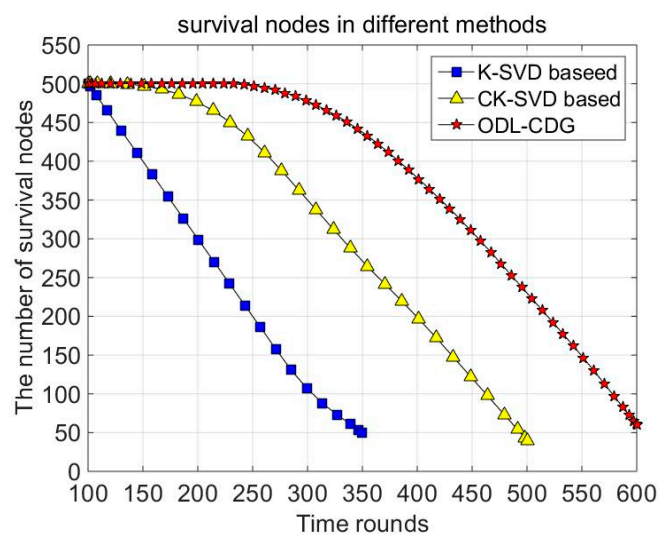


Figure 7. The number of survival nodes in different methods.

7. Conclusions and Future Work

In this paper, we propose the ODL-CDG algorithm for energy efficient data collection in WSNs. The training signals for dictionary learning are obtained by a compressive data gathering approach, which greatly reduces the energy consumption. Inspired by the periodicity of natural signals, the learned dictionary is constrained with a sparse structure. To reduce the recovery error caused by environmental noise, the self-coherence of the learned dictionary is also introduced as a penalty term during the dictionary optimization procedure. Simulation results show the online dictionary learning algorithm outperforms both pre-specified dictionaries, like the DCT dictionary, and other dictionary learning approaches, like K-SVD, IDL and CK-SVD. The energy consumption of the ODL-CDG

algorithm is significantly less than that of K-SVD-based and CK-SVD-based data gathering methods, which helps to enhance the network lifetime. In the future, we intend to employ other measurement matrices, such as the sparse measurement matrices, to further reduce the energy consumption. What's more, how to apply the proposed algorithm to real large-scale WSNs is also a potential research direction.

Acknowledgments: This work is supported by the National Natural Science Foundation of China under Grant No. 61371135 and Beihang University Innovation & Practice Fund for Graduate under Grant YCSJ-02-2016-04. The authors are grateful to the anonymous reviewers for their intensive reviews and insightful suggestions, which have improved the quality of this paper significantly.

Author Contributions: Donghao Wang made main contribution to the original ideas on the original optimization methods, detailed algorithm implementation, and the manuscript writing. Jiangwen Wan conceived the scope of the paper, and reviewed and revised the manuscript in detail. Junying Chen performed the experiments. Qiang Zhang analyzed the data.

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

References

1. Rault, T.; Bouabdallah, A.; Challal, Y. Energy efficiency in wireless sensor networks: A top-down survey. *Comput. Netw.* **2014**, *67*, 104–122. [[CrossRef](#)]
2. Cheng, S.T.; Shih, J.S.; Chou, C.L.; Horng, G.J.; Wang, C.H. Hierarchical distributed source coding scheme and optimal transmission scheduling for wireless sensor networks. *Wirel. Pers. Commun.* **2012**, *70*, 847–868. [[CrossRef](#)]
3. Baraniuk, R. Compressive sensing. *IEEE Signal Process. Mag.* **2007**, *24*, 4. [[CrossRef](#)]
4. Luo, C.; Wu, F.; Sun, J.; Chen, C.W. Compressive data gathering for large-scale wireless sensor networks. In Proceedings of the 15th ACM International Conference on Mobile Computing and Networking, Beijing, China, 20–25 September 2009; pp. 145–156.
5. Liu, Y.; Zhu, X.; Zhang, L.; Cho, S.H. Expanding window compressed sensing for non-uniform compressible signals. *Sensors* **2012**, *12*, 13034–13057. [[CrossRef](#)] [[PubMed](#)]
6. Shen, Y.; Hu, W.; Rana, R.; Chou, C.T. Nonuniform compressive sensing for heterogeneous wireless sensor networks. *IEEE Sens. J.* **2013**, *13*, 2120–2128. [[CrossRef](#)]
7. Razzaque, M.A.; Dobson, S. Energy-efficient sensing in wireless sensor networks using compressed sensing. *Sensors* **2014**, *14*, 2822–2859. [[CrossRef](#)] [[PubMed](#)]
8. Aharon, M.; Elad, M.; Bruckstein, A. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **2006**, *54*, 4311–4322. [[CrossRef](#)]
9. Duarte-Carvajalino, J.M.; Sapiro, G. Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. *IEEE Trans. Image Process.* **2009**, *18*, 1395–1408. [[CrossRef](#)] [[PubMed](#)]
10. Sigg, C.D.; Dikk, T.; Buhmann, J.M. Learning dictionaries with bounded self-coherence. *IEEE Signal Process. Lett.* **2012**, *19*, 861–864. [[CrossRef](#)]
11. Sadeghi, M.; Babaie-Zadeh, M.; Jutten, C. Learning overcomplete dictionaries based on atom-by-atom updating. *IEEE Trans. Signal Process.* **2014**, *62*, 883–891. [[CrossRef](#)]
12. Chen, W.; Wassell, I.; Rodrigues, M. Dictionary design for distributed compressive sensing. *IEEE Signal Process. Lett.* **2015**, *22*, 95–99. [[CrossRef](#)]
13. Studer, C.; Baraniuk, R.G. Dictionary learning from sparsely corrupted or compressed signals. In Proceedings of the 2012 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2012), Kyoto, Japan, 25–30 March 2012; pp. 3341–3344.
14. Pourkamali-Anaraki, F.; Becker, S.; Hughes, S.M. Efficient dictionary learning via very sparse random projections. In Proceedings of the 11th International Conference on Sampling Theory and Applications, SampTA 2015, Washington, DC, USA, 25–29 May 2015; pp. 478–482.
15. Pourkamali-Anaraki, F.; Hughes, S.M. Compressive k-svd. In Proceedings of the 2013 38th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013), Vancouver, BC, Canada, 26–31 May 2013; pp. 5469–5473.
16. Dictionary and Image Recovery from Incomplete and Random Measurements. Available online: <http://arxiv.org/pdf/1508.00278.pdf> (accessed on 15 September 2016).

17. Needell, D.; Tropp, J.A. Cosamp: Iterative signal recovery from incomplete and inaccurate samples. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 301–321. [[CrossRef](#)]
18. Jain, P.; Tewari, A.; Dhillon, I.S. Orthogonal matching pursuit with replacement. In Proceedings of the Advances in Neural Information Processing Systems, Granada, Spain, 12–14 December 2011; Curran Associates Inc.: Granada, Spain, 2011; pp. 1215–1223.
19. Donoho, D.L.; Tsaig, Y.; Drori, I.; Starck, J.L. Sparse solution of underdetermined systems of linear equations by stagewise orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2012**, *58*, 1094–1121. [[CrossRef](#)]
20. Candes, E.J.; Romberg, J.K.; Tao, T. Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **2006**, *59*, 1207–1223. [[CrossRef](#)]
21. Candes, E.; Tao, T. The dantzig selector: Statistical estimation when p is much larger than n . *Ann. Stat.* **2007**, *35*, 2313–2351. [[CrossRef](#)]
22. Li, G.; Zhu, Z.; Yang, D.; Chang, L.; Bai, H. On projection matrix optimization for compressive sensing systems. *IEEE Trans. Signal Process.* **2013**, *61*, 2887–2898. [[CrossRef](#)]
23. Cai, T.T.; Zhang, A. Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Trans. Inf. Theory* **2013**, *60*, 122–132. [[CrossRef](#)]
24. Rauhut, H.; Schnass, K.; Vandergheynst, P. Compressed sensing and redundant dictionaries. *IEEE Trans. Inf. Theory* **2008**, *54*, 2210–2219. [[CrossRef](#)]
25. Tropp, J.A. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inf. Theory* **2004**, *50*, 2231–2242. [[CrossRef](#)]
26. Tan, M.K.; Tsang, I.W.; Wang, L. Matching pursuit lasso part ii: Applications and sparse recovery over batch signals. *IEEE Trans. Signal Process.* **2015**, *63*, 742–753. [[CrossRef](#)]
27. Toh, K.C.; Yun, S. An accelerated proximal gradient algorithm for nuclear norm regularized linear least squares problems. *Pac. J. Optim.* **2010**, *6*, 615–640.
28. Jiang, K.; Sun, D.; Toh, K.C. An inexact accelerated proximal gradient method for large scale linearly constrained convex sdp. *SIAM J. Optim.* **2012**, *22*, 1042–1064. [[CrossRef](#)]
29. Balavoine, A.; Rozell, C.J.; Romberg, J. Discrete and continuous-time soft-thresholding for dynamic signal recovery. *IEEE Trans. Signal Process.* **2015**, *63*, 3165–3176. [[CrossRef](#)]
30. Madden, S. Intel Lab Data. Available online: <http://db.csail.mit.edu/labdata/labdata.html> (accessed on 15 September 2016).
31. Heinzelman, W.R.; Chandrakasan, A.; Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. In Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, Maui, HI, USA, 4–7 January 2000; p. 223.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).