Preprints of the 10th IFAC International Symposium on Dynamics and Control of Process Systems
The International Federation of Automatic Control
December 18-20, 2013. Mumbai, India

An adaptive basis estimation method for compressed sensing with applications to missing data reconstruction

Satheesh K. Perepu * **Arun K. Tangirala** **

* *Dept. of Chemical Engineering, IIT Madras, Chennai, Tamilnadu 600036 India. (e-mail: satheesh841@ gmail.com).*
** *Dept. of Chemical Engineering, IIT Madras, Chennai, Tamilnadu 600036 India . (e-mail: arunkt@ iitm.ac.in).*

Abstract: The subject of compressed sensing, especially, the related concept of sparse representation has been growing into an exciting area with a diverse set of applications in the fields of image sensing and analysis, signal compression, network reconstruction, *etc*. The efficacy of the associated techniques depends on the ability to discover a suitable basis for a sparse representation of the underlying signal. This paper presents a method for discovering this basis adaptively from the data. Specifically, the method estimates the dictionary of basis functions that maps the sub-sampled signal to the sparse representation of the signal. We present an application of this technique to the reconstruction of missing data, which is an important problem in all data-driven methods. Two case studies, namely, the reconstruction of missing data in a liquid level system and missing pixels of a 2-D signal (image) are presented. Results show that the proposed algorithm outperforms the existing KSVD algorithm in terms of both accuracy and speed of the reconstruction.

Keywords: compressed sensing;random sampling;missing data;basis pursuit;level system;KSVD.

1. INTRODUCTION

Compressed sensing is an exciting branch of signal processing with a growing number of applications in signal compression, image analysis, *etc*. The problem of compressed sensing is primarily concerned with the recovery of the signal on a regular and finer grid from its measurement (or representation) available on an irregular grid of time or space or any other independent dimension (*e.g.*, frequency). Recovering signals from limited or irregularly spaced data is an old problem, but the formulation and the solution methodologies are novel (Donoho, 2006; Candes, 2008). Several methods have appeared since the introduction of this problem in what is considered as a seminal paper by Donoho (2006). An important requirement underlying the use of these techniques, however, is that there exists a domain in which the underlying signal has a sparse representation. This requirement actually stems from the original motivation for compressed sensing. If a regularly sampled signal can be represented by a fewer coefficients, typically at irregularly spaced locations and in a *transformed* domain (e.g., Fourier or wavelet domain), then it is efficient to store only those fewer numbers. While this is well-known, the contribution of compressed sensing theory has been in providing methods to (i) directly compute the significant coefficients without the need for obtaining the complete sampled signal and (ii) to recover the original signal from the significant coefficients or even from irregularly sampled data, both under the assumption of existence of a basis space for sparse representation.

Applications of compressed sensing can be broadly divided into (i) design of algorithms and hardware for efficient storage of signals and (ii) recovery of signals from irregularly spaced data using sparse representations. The present work belongs to the latter category. Irregularly spaced data, specifically missing data are a commonplace in process measurements due to a number of known reasons such as sensor malfunctions, power outages, transmission errors, *etc* (Guo et al., 2012). Developing methods for handling missing data has been a topic of research over the last four to five decades. A popular technique for handling the problem of simultaneous model development and missing data is the Expectation Maximisation (EM) algorithm (Gopaluni, 2008; Moon, 1996). The EM algorithm iteratively estimates the missing data and the model parameters so that the likelihood of the complete data is maximized. However, it is computationally expensive, making it prohibitive for online applications. Several applications such as control and monitoring require online recovery of lost data.

The problem of missing data reconstruction fits naturally into the compressed sensing framework. Recently developed compressed sensing techniques can be used effectively to estimate missing data [Romberg, J. and Wakin (2007)]. Donoho (2006) shows that $O(K \log N)$ non-adaptive measurements are enough to reconstruct a signal of length N samples, given the signal is K sparse in some basis, *i.e.*, its representation should have K zero-valued coefficients. Given the basis, various algorithms such as Basis Pursuit, Orthogonal Matching Pursuit (OMP), *etc* are available in literature to reconstruct sparse representation of the signal from available data. See Elad (2009) for a detailed discussion of these algorithms. On the other hand, finding a suitable sparsifying basis for a given application remains an open-ended problem.

A simple choice of basis is a fixed or pre-specified basis matrix (*e.g.*, of the sinusoidal family). However, as is known in several signal processing applications, a fixed basis is not necessarily and usually the most appropriate basis for a given process. The reason is that, a fixed basis, while being mathematically suitable, does not necessarily explain the physics of the process. An alternative is to use an adaptive basis, *i.e.*, one that is derived from the data, which is the line of approach used in this work. While an adaptive method aids in extracting the basis that is appropriate to the given application, there is also a provision to impose the amount of sparsity as desired by that application. In the compressed sensing literature, one comes across a few different adaptive techniques to construct the basis matrix for sparse representations. The Method of Optimal Dictionaries (or the Method of Optimal Directions), popularly known as MOD algorithm, developed by Engan et al. (1999) (see also Candes (2008)) is one of the simplest techniques available for this purpose. A widely used method is based on the K-SVD algorithm due to Aharon et al. (2006). This method involves iterative computation of Singular Value Decomposition (SVD) of residual matrix as well as computation of inverse of the obtained matrix. In a different work, Xie and Feng (2009) propose the Kernel Fuzzy Codebook Estimation (KFCE) algorithm, which integrates the distance kernel trick with fuzzy clustering algorithm to obtain the basis matrix . A major shortcoming that is common to all the aforementioned techniques is that they require complete data to construct the basis matrix, which restricts their usage in applications with missing data.

In the present work, we present a method to obtain an optimal estimate of sparsifying basis (matrix) for a given signal from its measurements where the data missing at random. The proposed methodology is independent of the locations of missing samples.

Two different scenarios, namely time domain and spatial domain sparsity, are taken up for study with demonstrations on two case studies, reconstruction of missing level readings from a single liquid-level system and reconstruction of missing pixel values in an image respectively.

This paper is organized as follows. In Section 2 the basics of compressed sensing are reviewed. The proposed technique for the dictionary estimation is discussed in Section 3. Section 4 presents the application of the method to the aforementioned case studies. Concluding remarks appear in Section 5.

2. FOUNDATIONS OF COMPRESSED SENSING

As described in §1, compressed sensing is concerned with representing a signal with as few irregularly spaced coefficients or samples as possible, in contrast to the standard representation of a signal using regularly spaced samples (of usually large size). It rests on an important requirement, which is that the signal of interest should have a *sparse* representation in some basis space. An N-length vector $\mathbf{x}$ is said to be sparse if the number of non-zero elements of $\mathbf{x}$ is much smaller than N. The discrete-time impulse sequence (Kronecker delta function) $\{x[k] : x[k] = 0, k \neq 0,\ x[0] = c, c \neq 0\}$, for example, has the maximum sparsity.

Several measures of sparsity exist. In general, a sparsity measure should possess certain attributes such as scaling invariance, sensitivity to re-distribution of amplitudes, *etc.* See Hurley and Rickard (2009) and the references therein for a comparison of sparsity measures and an in-depth treatment of this topic. In the compressed sensing literature, the l_0-norm is frequently used as a measure of sparsity, at least, for theoretical formulations. The zeroth norm does not follow the usual definition of a p-norm. The following definition is used instead. For any vector $\mathbf{x}$,

$$||\mathbf{x}||_0 = \#\{i | x_i \neq 0\} \tag{1}$$

where # denotes the cardinality of the set.

The mathematics of compressed sensing is as follows. Consider a regularly sampled signal, denoted by $\mathbf{s} \in \mathbb{R}^n$. Let $\mathbf{y} \in \mathbb{R}^m$ denote the signal obtained by irregularly sampling $\mathbf{s}$, with $(m \ll n)$. Note that the samples in $\mathbf{y}$ are irregularly spaced. The sub-sampling operation in matrix form can be written as

$$\mathbf{y} = \mathbf{L}\mathbf{s} \tag{2}$$

where $\mathbf{L} \in \mathbb{R}^{m \times n}$ is the sampling matrix consisting of 1's and 0's at the appropriate locations. Recovering the signal from (2) uniquely is not possible since we have lost information in moving from $\mathbf{s}$ to $\mathbf{y}$. Mathematically, it is an underdetermined problem, *i.e.*, it contains more number of unknowns than equations.

Now assume that the signal $\mathbf{s}$ is sparse in some basis set $\mathcal{B}$, and denote the associated representation by $\mathbf{x}$

$$\mathbf{s} = \mathbf{B}\mathbf{x} \tag{3}$$

where $\mathbf{B}$ is an invertible mapping from the space of $\mathbf{x}$ to the space of $\mathbf{s}$.

Combining (2) and (3) yields

$$\mathbf{y} = \mathbf{A}\mathbf{x} \tag{4}$$

where $\mathbf{A} = \mathbf{L}\mathbf{B}$. Since, it is insisted that $\mathbf{x}$ is sparse, the optimization problem to solve (4) is setup as

$$\arg\min_{\mathbf{x}} ||\mathbf{x}||_0 \qquad \text{such that} \qquad \mathbf{y} = \mathbf{A}\mathbf{x} \tag{5}$$

The matrix $\mathbf{A}$ is also known as the *overcomplete dictionary.*

Solving (5) is a non-convex NP-hard problem. Consequently, the zero-norm is replaced by a suitable norm. Donoho (2006) proposed the 1-norm minimization - a convex problem.

$$\arg\min_{\mathbf{x}} ||\mathbf{x}||_1 \qquad \text{such that} \qquad \mathbf{y} = \mathbf{A}\mathbf{x} \tag{6}$$

Candes (2008) proved that the zero-norm solution is identical to the 1-norm solution as long as $\mathbf{A}$ satisfies a condition known as the *restrictive isometric property* (RIP). A basis pursuit (BP) algorithm can be used to solve (6)

The optimization problem in (6) is re-written in the standard way as

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \lambda||\mathbf{x}||_1 \tag{7}$$

Two widely used techniques to solve (7) are the least absolute shrinkage and selection operator (LASSO) and basis pursuit de-noising (BPDN. We choose to employ the BPDN in this work due to its robustness (to noise) property. Another advantage is that even when $\mathbf{x}$ is not sufficiently sparse, $\mathbf{s}$ can be approximated with a small error.

Where the problem of interest is in recovering the original s, knowing $\mathbf{B}$ it can be estimated from

$$\hat{\mathbf{s}} = \mathbf{B}\hat{\mathbf{x}} \tag{8}$$

The reconstructed vector $\mathbf{s}$ depends on the choice of matrix $\mathbf{A}$. Hence matrix $\mathbf{A}$ should be chosen wisely.

Compressed sensing has diverse applications in different fields. To mention a few, data compression, image denoising, estimation of missing data *etc.* Data compression refers to encoding information using fewer bits than the original representation with minimal information loss. Donoho (2006) proved that $k \log n$ random Gaussian measurements, i.e., $k \log n$ randomly selected projections of the measurements on Gaussian basis functions, are enough to reconstruct n samples of a signal. In such cases, the sampling matrix $\mathbf{L}$ will be a random Gaussian matrix. Elad and Aharon (2006) successfully used compressed sensing techniques to remove text data from the image. They considered sampling matrix $\mathbf{L}$ as 2-D DCT matrix.

The present work is concerned with the application of the compressed sensing concepts to reconstruction of missing data, under an important assumption - the underlying signal has a sparse representation in a basis space. To the best of the authors' knowledge, there exists no significant work in the literature using these ideas although the problem (of missing data reconstruction) itself has been studied for decades now.

3. PROPOSED METHOD

The missing data problem is concerned with recovery of lost data. In terms of the previously introduced notations, the problem is that of recovering $\mathbf{s}$ from the incomplete signal $\mathbf{y}$. The proposed method first estimates $\mathbf{A}$ and then proceeds to estimating the basis $\mathbf{B}$ in an optimal manner.

Given a set of training samples $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^{N_j}$ in a group, the signal in the i^{th} block is given by,

$$\mathbf{y}_i = \mathbf{A}\mathbf{x}_i + \mathbf{v} \tag{9}$$

where $\mathbf{v} \sim \mathcal{N}(0, \sigma_v^2 \mathbf{I}_{N_j})$ (Gaussian white noise), $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N_j}$ is the sparse representation of the original signal in the dictionary $\mathbf{A}$.

As a first step. imagine the complete signal $\mathbf{s} \in \mathbb{R}^n$ to be made up of P blocks of equal length. Each block is to be reconstructed individually. Blocks that have samples at same locations (same time stamps) are concatenated to form a group. Assume G groups exist and $N_j, j = 1, \cdots, G$ such sub-blocks exist in each group. Note that $\sum_{j=1}^{G} N_j = m$.

The algorithm is applied to each group individually with the eventual idea of fusing estimates of $\mathbf{B}$ from each group.

The optimum value of $\mathbf{A}$ is obtained by solving the optimization problem,

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \left(\sum_{\mathrm{i}=1}^{\mathrm{N_j}} \min_{\mathbf{x}_\mathrm{i}} \|\mathbf{y}_\mathrm{i} - \mathbf{A}\mathbf{x}_\mathrm{i}\|_2^2 + \lambda ||\mathbf{x}_\mathrm{i}||_1 \right) \tag{10}$$

where the summation across blocks in each group is justified by the fact that the blocks contain uncorrelated errors.

The inner and outer minimization problems are solved iteratively. The BPDN algorithm is used to solve the inner minimization for a given $\mathbf{A}$ while the outer one provides the estimate of $\mathbf{A}$ given $\mathbf{x}$ as follows.

Let the individual residual vector be

$$\mathbf{e}_i = \mathbf{y}_i - \mathbf{A}\mathbf{x}_i$$

Substituting

$$\hat{\mathbf{A}} = arg \min_{\mathbf{A}} \sum_{\mathrm{i}=1}^{\mathrm{N_j}} \min_{\mathbf{x}_\mathrm{i}} \|\mathbf{e}_\mathrm{i}\|_2^2 \tag{11}$$

which can be written as

$$\hat{\mathbf{A}} = \arg \min_{\mathbf{A}} \|\mathbf{E}\|_\mathrm{F}^2 \tag{12}$$

Solving (12), we obtain the overcomplete dictionary

$$\hat{\mathbf{A}} = \mathbf{Y}\mathbf{X}(\mathbf{X}\mathbf{X}^T)^{-1} \tag{13}$$

Equation (13) and the solution to (7) are solved iteratively as prescribed by the algorithm in Figure 1, where $\mathbf{L}_j$ is the sampling matrix for the j^{th} group. The number of iterations is a user-defined parameter and is usually governed by the accuracy requirements.

Initialize $\mathbf{A}^0 \in \mathcal{R}^{m \times n}$, $k = 0$

(1) Normalise the columns in matrix $\mathbf{A}$

(2) Find column-wise sparse representation of $\mathbf{Y}$ in $\mathbf{A}$ as

$$\arg \min_{\mathbf{X}} ||\mathbf{y}_i - \mathbf{A}^j \mathbf{x}_i||_2^2 + \lambda \|x\|_1$$

(3) Update matrix $\mathbf{A}$ by using formula

$$\mathbf{A}^{k+1} = \mathbf{Y}\mathbf{X}^{k^T}(\mathbf{X}^k \mathbf{X}^{k^T})^{-1}$$

(4) End for loop

(5) Estimate basis matrix $\hat{\mathbf{B}}_j = pinv(\mathbf{L_j}) \times \mathbf{A^k}$

Fig. 1. Algorithm to find basis for particular group of data

After calculating overcomplete dictionary $\mathbf{A}$, the basis matrix $\mathbf{B}$ can be calculated as

$$\hat{\mathbf{B}}_j = pinv(\mathbf{L})\,\mathbf{A} \tag{14}$$

Each group gives rise to a different estimate $\mathbf{B}_j$. The overall estimate of $\mathbf{B}$ is obtained as a weighted average of all the estimates $\mathbf{B}_j, j = 1, \cdots, G$. The weight of a group is given by N_j/G. The proposed algorithm is illustrated with the help of following case studies.

4. RESULTS AND DISCUSSIONS

To test the performance of the proposed algorithm, it is applied to a signal for which the true dictionary is known. The signal is a mixture of three sinusoidal waves having frequency 0.1Hz, 0.2Hz and 0.3Hz respectively. 1000 samples of the signal is generated and 100 samples are taken for reconstruction. To make the simulation realistic, Gaussian

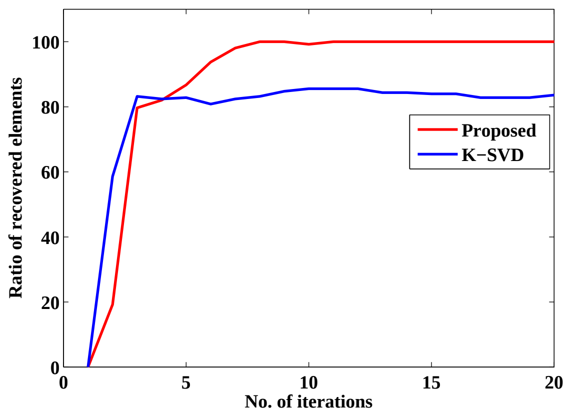

Fig. 2. Ratio of elements recovered from true dictionary

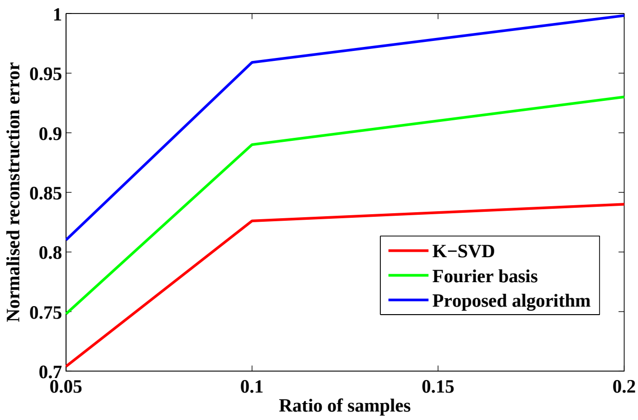

Fig. 3. Ratio of samples available (m/n) vs reconstruction error

white noise is added to the data. It is known that mixture of sinusoidal signals is sparse in Fourier basis. Every 10 samples of data is considered as a block. The graph below shows the percentage of recovered elements from the original dictionary for a particular group of samples. From Figure 2, it is clear that proposed algorithm recovers elements from true dictionary well when compared to K-SVD algorithm. To test the basis obtained from the above example, missing data reconstruction is done using the obtained basis. Different ratios of available to total number of samples are considered and normalised reconstruction errors for different ratios is plotted in Figure 3. For comparison purpose, signal is also reconstructed using Fourier basis. From the graph it is clear that reconstruction done using proposed algorithm estimated missing data with less error when compared to the reconstruction done using true Fourier basis as well as basis obtained using K-SVD. Reconstructed sparse vectors using proposed algorithm, KSVD as well as Fourier basis is shown in the Figure 4.
The proposed algorithm is tested on two scenarios. One is on reconstruction of missing level data of a single tank system and other is on reconstruction of missing pixels in an image. For comparison purposes, reconstruction is done using basis designed by K-SVD. The reconstruction error is calculated using the given formula

$$p = 1 - \frac{||\hat{s} - s||_2}{||s||_2} \tag{15}$$

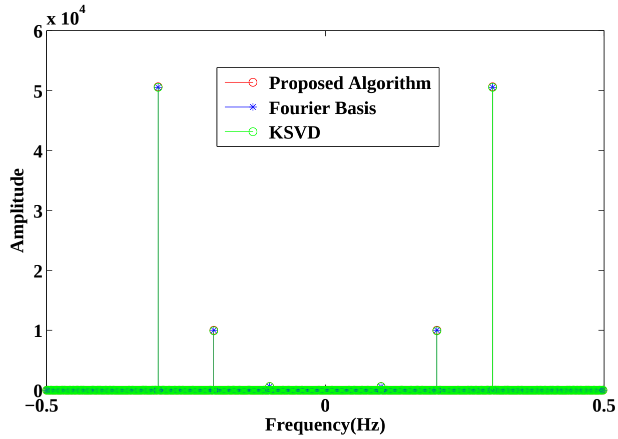

Fig. 4. Reconstruction of spectrum of sinusoidal signals

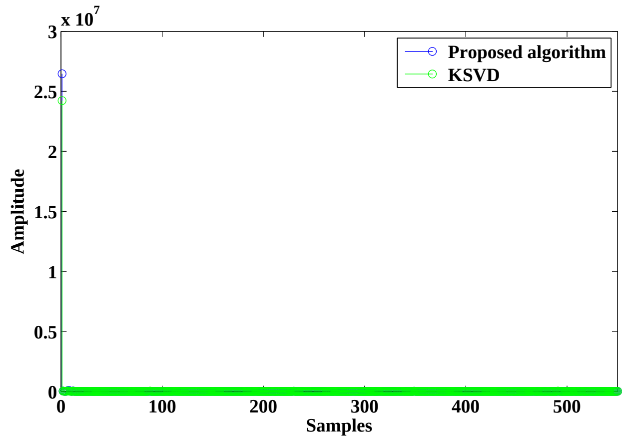

Fig. 5. Reconstruction of sparse vector using various algorithms

p can take values in between 0 to 1. The value of p close to 1 suggests good reconstruction. For illustrative purpose, reconstruction is done using basis designed by KSVD also. While designing basis using KSVD, it is assumed complete data is available.

4.1 Level Data

The objective of this example is to reconstruct steady state level data of a single tank system with irregularly spaced samples. The level data is collected from the liquid level system perturbed by a step change of 1V to the pump. After reaching steady state, 550 samples of level data are noted down.The data is sampled randomly to get 100 samples. Every 10 samples of data is considered as a block. The reconstruction is done using the basis obtained from the proposed algorithm as well as K-SVD. Figure 5 refers to sparse signal reconstruction of level data using proposed algorithm and KSVD. Figure 6 compares the reconstructed level data from proposed algorithm and K-SVD to the original level data. The normalised reconstruction errors and time taken for reconstruction are given in Table 1. Results show that reconstruction done using the basis obtained by the proposed algorithm is better than the reconstruction done by the basis obtained using K-SVD in terms of both normalised error and time taken for reconstruction.

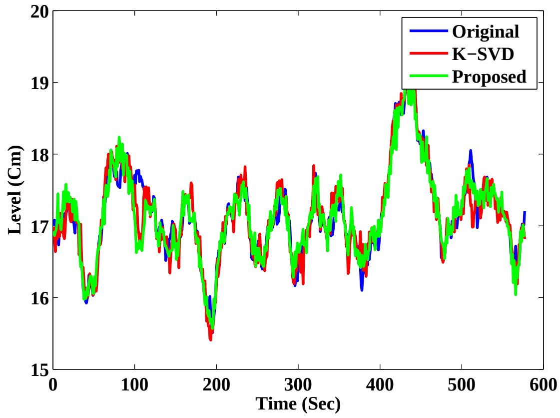

Fig. 6. Reconstruction of level data using basis obtained from various algorithms

Table 1. Reconstruction of level data

S. No.	Method used	Normalised error (p)	Time (Min)
1	Proposed method	0.958	15.46
2	K-SVD	0.916	24.15

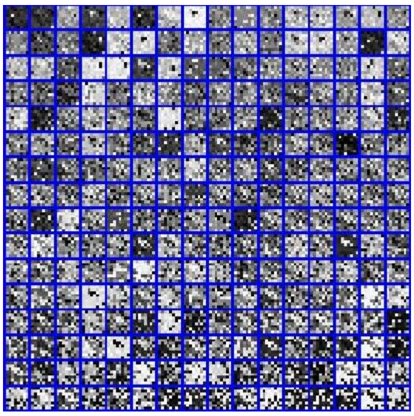

Fig. 7. Estimated basis Matrix using proposed algorithm with elements sorted by their variance

4.2 Image reconstruction

The objective of this example is to reconstruct 256×256 pixel image data by using 10000 pixel data. Every 16×16 pixels is considered as a block. The designed basis using the proposed algorithm and KSVD is shown in the Figures 7 and 8. Original image is shown in the Figure 9. Reconstructed sparse vector is shown in the Figure 10. The reconstructed images using proposed algorithm and KSVD are shown in the Figures 11 and 12 respectively. The normalised reconstruction errors and time taken for reconstruction are given in the Table 2. Results show that reconstruction done using the basis obtained by the proposed algorithm is better than the reconstruction done by the basis obtained using K-SVD in terms of both normalised error and time taken for reconstruction.

5. CONCLUSIONS

We presented a novel technique to construct basis from available data. The proposed algorithm is verified on the

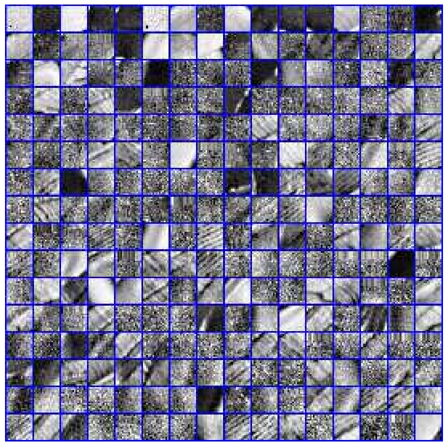

Fig. 8. Estimated basis Matrix using KSVD with elements sorted by their variance

Fig. 9. Original Image

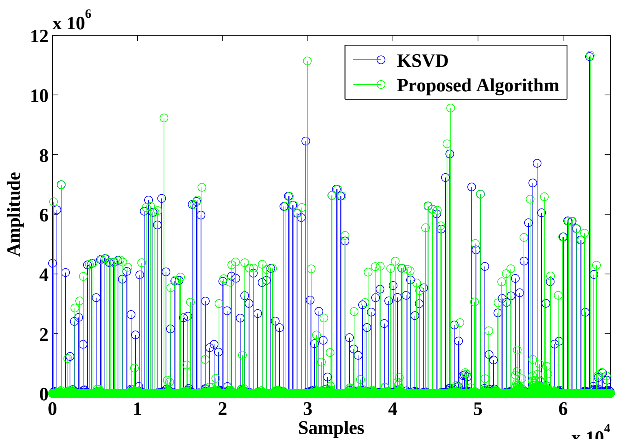

Fig. 10. Reconstructed sparse vector of image data using various algorithms

Fig. 11. Reconstructed Image using proposed algorithm

Fig. 12. Reconstructed Image using KSVD

Table 2. Reconstruction of image

S. No.	Method used	Normalised error (p)	Time (min)
1	Proposed method	0.9384	35.47
2	K-SVD	0.895	64.74

sinusoidal signal for which we know the true basis in which it is sparse. Basis designed using K-SVD algorithm is also used to compare the results obtained. Results show that when compared to K-SVD, the proposed technique recovered elements well from the true dictionary. The designed basis along with BPDN algorithm is used to reconstruct both real time data as well as image data. In both cases, the proposed algorithm outperforms the K-SVD algorithm both in terms of accuracy of reconstruction as well as time taken for reconstruction. In future, proposed algorithm can be modified such that it can be used to reconstruct the dynamic signals.

REFERENCES

Aharon, M., Elad, M., and Bruckstein, A. (2006). K -SVD : An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation. *IEEE Transactions on Signal Processing*, 54(11), 4311–4322.

Candes, E.J. (2008). The restricted isometry property and its implications for compressed sensing. *Theory of Signals/Mathematical Analysis*, 346, 589–592.

Donoho, D. (2006). Compressed sensing. *IEEE Transactions on Information Theory*, 52(4), 1289–1306.

Elad, M. (2009). *Sparse and Redundant Representations.* Springer.

Elad, M. and Aharon, M. (2006). Image denoising via sparse and redundant representations over learned dictionaries. *Image Processing, IEEE Transactions on*, 15(12), 3736–3745.

Engan, K., Aase, S., and Hakon Husoy, J. (1999). Method of optimal directions for frame design. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 5, 2443–2446 vol.5.

Gopaluni, R.B. (2008). A particle filter approach to identification of nonlinear processes under missing observations. *The Canadian Journal of Chemical Engineering*, 86(6), 1081–1092.

Guo, D., Zicheng, L., Xiaobo, Q., and Huang, L. (2012). Sparsity-Based Online Missing Data Recovery Using Overcomplete Dictionary. *Sensors Journal, IEEE*, 12(7), 2485–2495.

Hurley, N. and Rickard, S. (2009). Comparing measures of sparsity. *Information Theory, IEEE Transactions on*, 55(10), 4723–4741.

Moon, T. (1996). The Expectation Maximization Algorithm. *IEEE Signal Processing Magazine*, 47–60.

Romberg, J. and Wakin, M. (2007). Compressed Sensing: A Tutorial. In *IEEE Statistical Signal Processing Workshop,*.

Xie, Z. and Feng, J. (2009). KFCE: A dictionary generation algorithm for sparse representation. *Signal Processing*, 89(10), 2072–2077.