A Physics-Constrained Dictionary Learning Approach for Compression of Vibration Signals

Yanglong Lu
Woodruff School of Mechanical Engineering
Georgia Institute of Technology, Atlanta, GA
ylu381@gatech.edu

Yan Wang*
Woodruff School of Mechanical Engineering
Georgia Institute of Technology, Atlanta, GA
yan-wang@gatech.edu

*Contact author:
Prof. Yan Wang
Email: yan-wang@gatech.edu
Mail: 813 Ferst Drive NW, Atlanta, GA 30332-0405, USA.
Telephone: +1(404)894-4714
Fax: +1(404)894-9342

A Physics-Constrained Dictionary Learning Approach for Compression of Vibration Signals

Abstract

Monitoring the health condition of rotating machinery in manufacturing systems usually requires vibration signals to be continuously collected, transmitted, and stored. The available bandwidth in communication channels for transmission of a large amount of data is limited in an industry setting. Therefore, reducing the amount of data in communication and storage without sacrificing the amount of information collection is necessary. Here, a new technique called physics-constrained dictionary learning is proposed to reduce the volume of data in storage and communication using compressed sensing. In compressed sensing, the original signals can be reconstructed with a much smaller amount of data determined by a measurement matrix, if the representation of signals in the reciprocal space is sparse. The proposed physics-constrained dictionary learning approach optimizes the measurement and basis matrices simultaneously to improve the accuracy of reconstruction, where physical constraints of time stamps of sampling and sampling intervals are considered. New training algorithms are developed. The proposed scheme is applied to compress the vibration signals of roller bearings. It is shown that the reconstruction performance of the proposed scheme is significantly improved from traditional dictionary learning.

Keywords: Compressed sensing; Dictionary learning; Sparse coding; Data compression; Rotating machinery

1. Introduction

Monitoring the health condition of rotating machinery in manufacturing systems usually requires vibration signals to be continuously collected and transmitted. The available bandwidth in communication channels for transmission of a large amount of data is limited in an industry setting where different machine conditions need to be monitored. Therefore, reducing the amount of data in communication and storage without sacrificing the amount of information gathered and exchanged is necessary. In the most recent decade, compressive sampling or compressed sensing (CS) [1, 2] was developed as a new sampling approach which requires less data storage and the original signal can be reconstructed. If the original signal has a data size of N and its representation in the reciprocal space is sparse with only K non-zero coefficients ($K<<N$), standard CS for generic signals allows for robust recovery from M=O(Klog(N/K)) measurements.

The main idea of CS is as follows. Suppose that the original signal is represented in a discrete form as vector $\boldsymbol{s} \in \mathbb{R}^N$. It can be represented in the reciprocal space via transformations as $\boldsymbol{s} = \boldsymbol{\Psi\gamma}$, where $\boldsymbol{\Psi} \in \mathbb{R}^{N\times N}$ is the transformation or basis matrix and $\boldsymbol{\gamma} \in \mathbb{R}^N$ is the vector of coefficients. When the signal is projected into the M-dimensional measurement subspace ($M<N$) with measurement matrix $\boldsymbol{\Phi} \in \mathbb{R}^{M\times N}$ as $\boldsymbol{y} = \boldsymbol{\Phi s}$, the recovery of the original signal $\boldsymbol{s}$ from the measured data $\boldsymbol{y}$ is to solve the linear equations $\boldsymbol{y} = \boldsymbol{\Phi s} = \boldsymbol{\Phi\Psi\gamma}$. Theoretically the recovery can be precise when the vector of coefficients $\boldsymbol{\gamma}$ is sparse and the transformation and projection operations are incoherent. The basis matrix $\boldsymbol{\Psi}$ is often predefined as some known transformation matrices such as Fourier transformation, discrete cosine transformation, wavelet transformation, or some random matrices that satisfy the restricted isometry property.

One issue with the predefined basis matrix $\boldsymbol{\Psi}$ is that it is not directly related to the observed signals. If the sparsity level of the coefficient vector $\boldsymbol{\gamma}$ is low, CS will not perform well. Therefore, dictionary learning approaches have been developed to train a dictionary $\mathbf{D}$ specifically based on the collected data to replace $\boldsymbol{\Phi\Psi}$. Given P sets of collected data $\mathbf{Y} = [\boldsymbol{y}_1, \boldsymbol{y}_2 \dots \boldsymbol{y}_P] \in \mathbb{R}^{M\times P}$, the dictionary $\mathbf{D}$ needs to be optimized so that $\mathbf{Y}$ can be sparsely represented as $\mathbf{Y} = \mathbf{D}\boldsymbol{\Upsilon}$, where $\boldsymbol{\Upsilon} = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \dots \boldsymbol{\gamma}_P] \in \mathbb{R}^{W\times P}$ is a matrix that includes the sparse vector of coefficients for each collected dataset and $W \geq N$. Thus, dictionary $\mathbf{D} \in \mathbb{R}^{M \times W}$ is over-complete since $M < W$.

Although existing dictionary learning approaches have been applied to learn the dictionary $\mathbf{D}$ for natural images, vibration signals and others, they cannot be used to determine the locations of pixels to be measured and stored for two-dimensional images or the time stamps of measurements for one-dimensional signals, because the measurement matrix $\mathbf{\Phi}$ is not designed explicitly. For instance, to select data points for storage and communication from all collected vibration signals, the measurement matrix $\mathbf{\Phi}$ needs to be optimized to determine the total number of stored and transmitted measurements. For the measurement matrix $\mathbf{\Phi}$, there should be only one non-zero entry in each row of $\mathbf{\Phi}$ and other entries are zeros. The index of non-zero entry in each row indicates the time stamps of sampling or when to store and transmit data. Therefore, the optimization of $\mathbf{\Phi}$ and $\mathbf{\Psi}$ individually provides more physical meanings of the optimized dictionary. Furthermore, the columns of the trained dictionary $\mathbf{D}$ in traditional dictionary learning are not always orthogonal, which affects the CS performance. A well-designed measurement matrix $\mathbf{\Phi}$ can also improve the orthogonality of the columns in $\mathbf{D}$.

In this paper, a new physics-constrained dictionary learning scheme is proposed to reduce the amount of data in storage and communication. From all collected data points, only a few of them are stored and transmitted. The original signal can be reconstructed from the compressed data with CS. The actual storage space and communication cost are determined by the optimized measurement matrix $\mathbf{\Phi}$, and the signal can achieve a high sparsity level with respect to the optimized basis matrix. Some physical constraints such as the data storage space, the number of measurements, sensor accessibility, and the energy consumption of data collection can be considered in the learning process to optimize the basis and measurement matrices. Here, the minimum sampling interval between compressed data points is used as the physical constraint to demonstrate the new physics-constrained dictionary learning approach for vibration signals of roller bearings, which can reduce the redundancy for the storage and communication of temporally correlated data. The main contributions of this paper include the new formulation of dictionary learning subject to physical constraints, as well as new algorithms to simultaneously optimize the basis matrix $\mathbf{\Psi}$ for sparse representation and the measurement matrix $\mathbf{\Phi}$ for the physical time stamps of sampling.

In the remainder of the paper, the background of dictionary learning methods and their applications are introduced in Section 2. The proposed physics-constrained dictionary learning method is described in Section 3. The demonstration of its application to compress the roller bearing vibration signal for the storage and communication, and experimental results are given in Section 4.

2. Background

Various dictionary learning methods [3] have been developed to search for the sparsest representation of signals. The purpose is to find the optimal dictionary so that the sparsity is maximized for a specific type of signals. As a result, the original signals can be represented in a form of linear combinations of the learned dictionary and the sparse vector of coefficients. Some commonly used dictionary learning algorithms include the method of optimal directions [4], K-SVD [5], the online dictionary learning [6] and others. The training process is also based on the maximum likelihood [7], least-square error [8, 9], and hidden Markov model [10].

Dictionary learning methods have been applied in combination with CS. For conventional CS, the basis matrix is usually predefined, so it is not directly related to the observed signals. Therefore, dictionary learning approaches have been developed to improve the sparsity level of the coefficient vector with a trained dictionary specifically based on the collected data. For example, Chen et al. [7] applied the dictionary learning method to improve the CS performance in extracting impulse components from noisy vibration signals. Lorintiu et al. [11] reconstructed ultrasound data with CS and dictionary learning by K-SVD. It was shown that reconstruction errors are lower than conventional dictionaries based on Fourier or discrete cosine transformations. CS with learned dictionary was also applied for the reconstruction of magnetic resonance images [12-15], videos [16] and electrocardiogram signals [17], and image denoising [18-20]. The existing dictionary learning approaches can improve the performance of CS. However, they are limited in practical applications because the measurement matrix is not designed explicitly. The measurement matrix is necessary to determine the locations of pixels to be measured and stored for two-dimensional images or the time stamps of measurements for one-dimensional signals.

Instead of learning the dictionary, which is the combination of the measurement matrix and the basis matrix, approaches to design the measurement and basis matrices separately were also developed. Duarte-Carvajalino and Sapiro [21] simultaneously optimized the measurement matrix and basis matrix with a new scheme called coupled-KSVD. The incoherence between the measurement and basis matrices was improved and resulted in better reconstruction performance. Bai et al. [22] further improved the framework with analytical solutions to update the measurement and basis matrices. It was shown that the convergence and accuracy of the solutions were improved for reconstruction of natural images. Nevertheless, in the above approaches, the optimized measurement matrix is dense. The dense measurement matrix cannot be used to determine the locations or time stamps of measurements or sampling in physical experiments. To be physically meaningful, measurement matrices should have only one non-zero entry in each row. The index of non-zero entry in each row indicates the time stamps to sample and store signals. Furthermore, physical constraints such as the data storage space, the number of measurements, sensor accessibility, and the energy consumption of data collection are important but considered in the existing approaches. Physical constraints ensure that the optimal performance is realizable in practical applications.

The proposed physics-constrained dictionary learning framework optimizes the measurement and basis matrices simultaneously where the measurement matrix with only one non-zero entry in each row can directly indicate the time stamp of sampling. The physical constraint of the minimum sampling interval between stored and transmitted measurements is considered to reduce the redundancy for the storage and communication of temporally correlated data. The number of required measurements thus is optimized based on the physical constraints.

3. Methodology

The proposed physics-constrained dictionary learning scheme is to optimize the measurement matrix $\mathbf{\Phi}$ and the basis matrix $\mathbf{\Psi}$ simultaneously under the physical constraints related to the time stamps for sampling. It is formulated as

$$\min_{\Phi,\Psi,\Upsilon}\ \alpha\|\mathbf{S}-\boldsymbol{\Psi}\boldsymbol{\Upsilon}\|_F^2+\|\boldsymbol{\Phi}\mathbf{S}-\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\Upsilon}\|_F^2 \tag{1}$$

$$\text{subject to}\ \ \boldsymbol{\Phi}=f(\boldsymbol{\Psi}) \tag{2}$$

$$\|\boldsymbol{\gamma}_i\|_{\mathbf{0}}\leq\ T_0,\ \ \forall i \tag{3}$$

$$I_{ij}(\boldsymbol{\Phi})\geq\ r,\ \ \forall i,j \tag{4}$$

where F denotes the Frobenius norm, $\mathbf{S}=[\boldsymbol{s}_1,\boldsymbol{s}_2\ldots\boldsymbol{s}_P]\in\mathbb{R}^{N\times P}$ contains P sets of training signals and each set of signals has the length of N. $\boldsymbol{\Psi}\in\mathbb{R}^{N\times W}$ is the basis matrix with $W\ll P$ and $W>N$. $\boldsymbol{\Upsilon}=[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\ldots\boldsymbol{\gamma}_P]\in\mathbb{R}^{W\times P}$ contains the sparse coefficients that represent the training signals in $\mathbf{S}$ with respect to the basis matrix. A Lagrange multiplier α is applied to combine the objectives of recovery accuracy and measurement accuracy. A small value of α such as 0.01 is used in practice because a relatively larger control weight of the error term $\|\boldsymbol{\Phi}\mathbf{S}-\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\Upsilon}\|_F^2$ is necessary to design the measurement matrix to minimize the reconstruction error. The constraint in Eq.(2) indicates the training sequence, where basis matrix $\boldsymbol{\Psi}$ is updated before measurement matrix $\boldsymbol{\Phi}$. With the fixed basis matrix $\boldsymbol{\Psi}$, measurement matrix $\boldsymbol{\Phi}$ can be optimized based on $f(\boldsymbol{\Psi})$. The constraint in Eq.(3) is the upper limit of the sparsity level, where $\boldsymbol{\gamma}_i$ is the *i*-th column of coefficient matrix, and T_0 is the target number of non-zero values in the sparse vectors of coefficients. The constraint in Eq.(4) shows the physical limitations of sampling, which is the lower limit of the time interval I_{ij} between the i-th and j-th stored or transmitted measurements, for instance, between any two consecutive measurements. If the time interval between stored or transmitted measurements is too small, more redundant information is collected because of large similarities between temporally correlated measurements. Other physical constraints can be added similarly.

The physics-constrained dictionary learning problem is solved to optimize the measurement and basis matrices by two stages iteratively as shown in Figure 1. It starts with an initial guess of the basis matrix which can be some known transformation matrices such as discrete cosine transformation and wavelet transformation. In the first stage, with the basis matrix fixed, the measurement matrix is optimized by determining suitable measurements to be stored and transmitted from originally collected signals under the constraint that there is only one non-zero entry in each row to determine the time stamp of the measurement. This can be solved based on algorithms such as the FrameSense [23]. In the second stage, with the measurement matrix fixed, the basis matrix is then optimized. This can be done with dictionary learning algorithms such as the K-SVD. The above two optimization steps are repeated until both the optimal measurement and basis matrices converge without further improvement. Physical constraints such as the total number of collected data points and the minimum sampling interval can be incorporated.

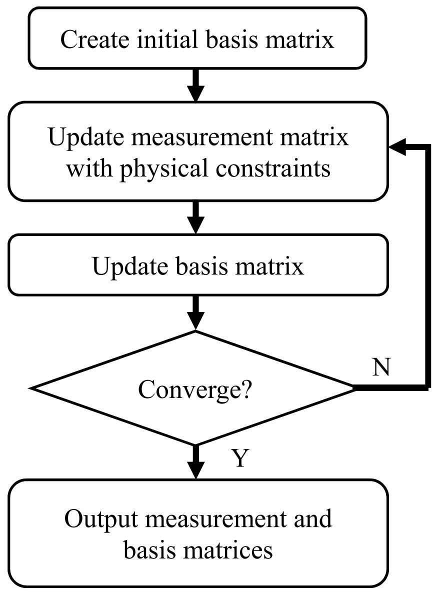


Figure 1. Two-stage optimization scheme.

3.1. Stage one optimization

At the stage one, the basis matrix $\mathbf{\Psi}$ is fixed, and the measurement matrix $\mathbf{\Phi}$ is optimized. Instead of directly searching for the optimal time stamps of stored and transmitted measurements which is often NP-hard if the number of measurements is large, the near-optimal time stamps of sampling can be obtained based on the greedy algorithm FrameSense in Table 1. Given all available time stamps $\mathcal{N} = \{1, \ldots, N\}$, an unsuitable set of time stamps $\mathcal{T}$ can be iteratively identified as the index of the row in the basis matrix $\mathbf{\Psi}$ by solving

$$\max_{\mathcal{T}} \mathrm{F}(\mathcal{T}) = \mathrm{H}(\boldsymbol{\Psi}) - \mathrm{H}(\boldsymbol{\Psi}_{\mathcal{N}\backslash\mathcal{T}}) \tag{5}$$

where $\mathrm{H}(\boldsymbol{\Psi})$ is the so-called frame potential and represented as

$$\mathrm{H}(\boldsymbol{\Psi}) = \sum_{i=1}^{N} |\lambda_i|^2 \tag{6}$$

where λ_i is the i-th largest eigenvalue of $\boldsymbol{\Psi}^*\boldsymbol{\Psi}$ and $\boldsymbol{\Psi}^*$ is the conjugate transpose of $\boldsymbol{\Psi}$. $\boldsymbol{\Psi}_{\mathcal{N}} = \boldsymbol{\Psi}$ if all time stamps of measurements are used. $\boldsymbol{\Psi}_{\mathcal{N}\backslash\mathcal{T}}$ is a sub-matrix of $\boldsymbol{\Psi}_{\mathcal{N}}$ with rows corresponding to indices with the unsuitable ones excluded. After determining the unsuitable time stamps $\mathcal{T}$, the new available time stamps are updated as $\mathcal{N}\backslash\mathcal{T}$.

If M measurements are desirable, the time stamps of M measurements are optimized by excluding $(N - M)$ unsuitable time stamps iteratively. Eventually the time stamps of the desirable measurements can be identified in the optimized $M \times N$ measurement matrix in a form of

$$\boldsymbol{\Phi} = \begin{bmatrix} \mathbf{0} & \mathbf{1} & \mathbf{0} & & \mathbf{0} \\ \mathbf{1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & & \mathbf{0} \\ & \vdots & & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \end{bmatrix} \tag{7}$$

where the column index of the value of 1 in each row indicates the time of each stored and transmitted measurement.

Here, the original FrameSense algorithm is modified to check if the additional physical constraints in Eq.(4) are satisfied. If the time interval between any of two stored and transmitted measurements is less than the threshold value r, one of the measurements in the pair is eliminated. For one-dimensional signals, r indicates the minimum time interval between two adjacent data points.

Table 1. The constrained FrameSense algorithm

1. Initialize time stamps of stored and transmitted measurements $\mathcal{L}$, all available time stamps $\mathcal{N}$, and desired number of stored and transmitted measurements m_t
2. Determine the first two removed rows in $\mathbf{\Psi}$ by solving $\mathcal{T} = \text{argmax}_{i,j\in\mathcal{N}} \left|< \boldsymbol{\psi}_i, \boldsymbol{\psi}_j >\right|^2$ and update remaining time stamps $\mathcal{L} = \mathcal{N}\backslash\mathcal{T}$ by excluding $\mathcal{T}$
3. WHILE the length of $\mathcal{L} < m_t$ DO
   Find the i^*-th row in $\mathbf{\Psi}$ to eliminate by solving $i^* = \text{argmax}_{i\in\mathcal{L}} \text{F}(\mathcal{T} \cup \{i\})$, where $\text{F}(\mathcal{T} \cup \{i\})$ is the function in Eq.(5)
   Update unsuitable time stamps of stored and transmitted measurements as $\mathcal{T} = \mathcal{T} \cup \{i^*\}$
   Update available time stamps of stored and transmitted measurements, $\mathcal{L} = \mathcal{L}\backslash\{i^*\}$
   END WHILE
4. FOR i = 1 to the length of $\mathcal{L}$
   FOR j = 1 to length of $\mathcal{L}$
   If $I_{\boldsymbol{ij}}(\mathbf{\Phi}) = \left|t_i - t_j\right| \leq r$, where t_i and t_j are time stamps for i-th and j-th data, $\mathcal{L} = \mathcal{L}\backslash\{j\}$.
   END FOR
   END FOR
5. Generate measurement matrix $\mathbf{\Phi}$ in the form of Eq.(7) with optimized time stamps $\mathcal{L}$

3.2. Stage two optimization

A simplified form of the objective function in Eq.(1) can be written as

$$\min_{\Phi,\Psi,\Upsilon} \left\| \begin{pmatrix} \alpha\mathbf{S} \\ \mathbf{\Phi S} \end{pmatrix} - \begin{pmatrix} \alpha\mathbf{I} \\ \mathbf{\Phi} \end{pmatrix} \mathbf{\Psi\Upsilon} \right\|_F^2 \quad (8)$$

At the stage two, $\mathbf{\Psi}$ and $\mathbf{\Upsilon}$ are optimized with the fixed measurement matrix $\mathbf{\Phi}$ obtained in the stage one subject to the constraint in Eq.(3). With new notations $\mathbf{X} = \begin{pmatrix} \alpha\mathbf{S} \\ \mathbf{\Phi S} \end{pmatrix}$ and $\mathbf{Z} = \begin{pmatrix} \alpha\mathbf{I} \\ \mathbf{\Phi} \end{pmatrix} \mathbf{\Psi}$, Eq.(8) can be rewritten, and $\mathbf{Z}$ and $\mathbf{\Upsilon}$ can be optimized by solving

$$\min_{\mathbf{Z},\mathbf{\Upsilon}} \|\mathbf{X} - \mathbf{Z\Upsilon}\|_F^2 \quad (9)$$

$$\text{subject to } \|\boldsymbol{\gamma}_i\|_{\mathbf{0}} \leq T_0, \ \forall i \tag{10}$$

The optimal $\mathbf{Z}$ and $\boldsymbol{\Upsilon}$ are found iteratively. With an initialized and fixed basis matrix $\boldsymbol{\Psi}$, the coefficient matrix $\boldsymbol{\Upsilon}$ can be obtained with the orthogonal matching pursuit (OMP) [24] algorithm in solving Eqs.(9) and (10). With the obtained coefficient matrix $\boldsymbol{\Upsilon}$ fixed, one column of the basis matrix is then updated each time by solving Eq.(9), which is re-written as

$$\min_{\mathbf{z}_k}\left\|(\mathbf{X}-\textstyle\sum_{\boldsymbol{j}\neq\boldsymbol{k}}^{W}\boldsymbol{z}_j\boldsymbol{\gamma}'_j)-\boldsymbol{z}_k\boldsymbol{\gamma}'_k\right\|_F^2=\min_{\mathbf{z}_k}\|\mathbf{E}_k-\boldsymbol{z}_k\boldsymbol{\gamma}'_k\|_F^2 \tag{11}$$

where $\boldsymbol{z}_j$ is the *j*-th column of $\mathbf{Z}$, $\boldsymbol{\gamma}'_j$ indicates the *j*-th row of $\boldsymbol{\Upsilon}$, and $\mathbf{E}_k$ represents the errors of all sample signals except *k*-the atom. Here, $\boldsymbol{z}_k$ as the *k*-th column of $\mathbf{Z}$ is updated iteratively. In order to satisfy the sparsity constraint in Eq. (10), additional modification of the objective function is needed. We define

$$\omega_k = \{i|1 \leq i \leq W, \boldsymbol{\gamma}'_k(i) \neq 0\} \tag{12}$$

as the set of indices where training signals $\{\boldsymbol{s}_i\}_{i=1}^{P}$ rely on atom $\boldsymbol{z}_k$. Eq. (11) is further written as

$$\min_{\mathbf{z}_k}\left\|\mathbf{E}_k^{\mathrm{R}}-\boldsymbol{z}_k\boldsymbol{\gamma}_k^{\mathrm{R}}\right\|_F^2 \tag{13}$$

where $\mathbf{E}_k^{\mathrm{R}} = \mathbf{E}_k\boldsymbol{\Omega}_k$ and $\boldsymbol{\gamma}_k^{\mathrm{R}} = \boldsymbol{\gamma}'_j\boldsymbol{\Omega}_k$. $\boldsymbol{\Omega}_k$ is a matrix of size $P\times|\omega_k|$ with ones on the $(\omega_k(i), i)$ entries and zero elsewhere. Following the K-SVD algorithm, $\mathbf{E}_k^{\mathrm{R}}$ can be decomposed to $\mathbf{U}\boldsymbol{\Delta}\mathbf{V}^{\mathrm{T}}$ with the singular value decomposition (SVD). Then updated $\boldsymbol{z}_k$ is the first column of $\mathbf{U}$ and the coefficient vector $\boldsymbol{\gamma}_k^{\mathrm{R}}$ is the first column of $\mathbf{V}$ multiplied by $\boldsymbol{\Delta}(1,1)$. After each column of $\mathbf{Z}$ is updated, basis matrix $\boldsymbol{\Psi}$ can be obtained with the pseudo-inverse as

$$\boldsymbol{\Psi} = \left(\alpha^2\mathbf{I} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-\mathbf{1}}[\ \alpha\mathbf{I} \quad \boldsymbol{\Phi}^{\mathrm{T}}]\mathbf{Z} \tag{14}$$

In this paper, a modified K-SVD algorithm is developed. Different from the conventional K-SVD which requires a prefixed number of non-zero values in the sparse vectors, the modified K-SVD can adaptively determine the most appropriate sparsity level T_0 in order to reduce the reconstruction error. Depending on different training datasets, initial basis matrices, and the number of available measurements, the optimal sparsity level T_0 can be different. In the adaptive K-SVD algorithm listed in Table 2, $\mathbf{S}_n$ represents n

randomly selected training data from $\mathbf{S}$. With optimized $\mathbf{\Phi}$ and $\mathbf{\Psi}$, coefficient vectors $\mathbf{\Upsilon}_{\mathrm{n}}$ can be recovered by solving $\mathbf{\Phi S}_{\mathrm{n}} = \mathbf{\Phi \Psi \Upsilon}_{\mathrm{n}}$ with OMP. The average of reconstruction errors is computed as $e_{ave} = 1/(N \times n) \cdot \sum_i^N \sum_j^n \big(([\mathbf{S}_n]_{ij} - [\mathbf{\Psi \Upsilon}_{\mathrm{n}}]_{ij})/[\mathbf{S}_{\boldsymbol{n}}]_{\boldsymbol{ij}} \times 100\%\big)$. T_0 is determined adaptively by finding the minimum e_{ave}.

Table 2. Adaptive K-SVD algorithm

Input:	Training signals $\mathbf{S}$; initial basis matrix $\mathbf{\Psi}_0$; measurement matrix $\mathbf{\Phi}$, initial number of non-zero values T_0; sparsity adjustment step size ΔT_0; weight of error α; number of selected training data used to compute the reconstruction error n; maximum number of iterations C; target training error e_t
Output:	Estimated sparse coefficients or parameters $\mathbf{\Upsilon}$; basis matrix $\mathbf{\Psi}$; desired sparsity level T_0

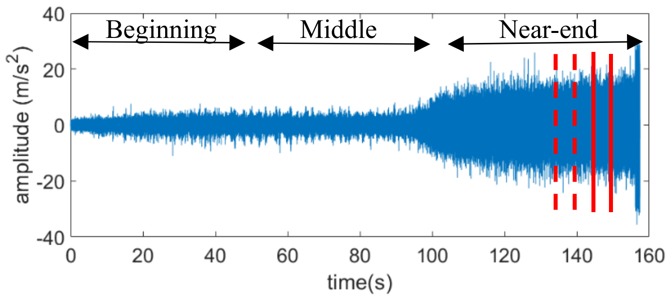
Procedure:	Initialize the average of reconstruction errors e_{ave}^0 and e_{ave}^1 with $e_{ave}^1 > e_{ave}^0$, and $m = 0$.
	WHILE $m < C$ and $e_{ave}^1 < e_{ave}^0$ DO
	    $e_{ave}^0 = e_{ave}^1$
	    WHILE $\Vert\mathbf{X} - \mathbf{Z\Upsilon}\Vert_F^2 > e_t$ DO
	        Compute $\mathbf{\Upsilon}$ by solving Eqs.(9) and (10) with OMP.
	        FOR k=1 to W
	            Update ω_k with Eq.(12)
	            Compute $\mathbf{E}_k$, $\mathbf{E}_k^{\mathrm{R}}$ and $\boldsymbol{\gamma}_k^{\mathrm{R}}$
	            Apply SVD to $\mathbf{E}_k^{\mathrm{R}} = \mathbf{U\Delta V}^{\mathrm{T}}$,
	            Update $\boldsymbol{z}_k$ as first column of $\mathbf{U}$
	            Update $\boldsymbol{\gamma}_k^{\mathrm{R}}$ as first column of $\mathbf{V}$ multiplied by $\mathbf{\Delta}(1{,}1)$.
	        END FOR
	        Update $\mathbf{\Psi}$ with Eq.(14).
	    END WHILE
	    Compute e_{ave}^1 with optimized $\mathbf{\Phi}$, $\mathbf{\Psi}$ and $\mathbf{S}_n$.
	    $m = m + 1$
	    $T_0 = T_0 + \Delta T_0$
	END WHILE

4. Experiments

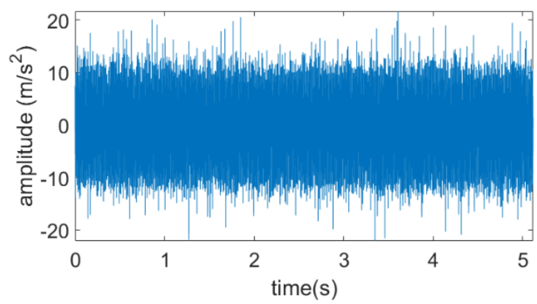
The proposed physics-constrained dictionary learning scheme was applied to compress roller bearing vibration signals to reduce the storage space and communication cost. The run-to-failure data from accelerated degradation tests of bearings were acquired by Wang *et al.* [25]. The major components of the test bed include digital force display, motor speed controller, support bearings, AC motor, hydraulic loading, tested bearing and accelerometers. Two PCB 352C33 accelerometers were used to acquire the run-to-failure data of the tested bearings. One of them was placed on the vertical axis and the other one was on the horizontal axis. The vibration signals were collected with a sampling rate of 25.6 kHz. For every minute, only the first 32768 data points (i.e. 1.26s) were recorded.

To test the proposed scheme, the vibration signals in both of horizontal and the vertical axes were used and 115000 data points were extracted from the run-to-failure data in each axis. These data points were divided into 1150 segments, and each segment contains 100 data points. After reconstructing each segment with a few data points, the complete signal can be obtained by combing all segments. Therefore, the number of data points in each segment is not critical, whereas the compression ratio is important for the reconstruction performance. Therefore, different compression ratios are used when the original signal is reconstructed. These 1150 segments were used as the training dataset. For the testing dataset, 115000 data points were extracted from a different period of time. 5000 consecutive data points is then randomly selected from 115000 data points and further divided into 50 segments as the testing dataset. Training and testing datasets are extracted from different sections of the run-to-failure data such as the beginning (i.e. 0~50s), middle (i.e. 50~100s), and near end (i.e. 100~160s) of life to demonstrate the robustness of the proposed scheme. The complete and different sections of the run-to-failure data in horizontal axis are seen in Figure 2 (a). The x axis in Figure 2 (a) shows the time period of the recorded signals. Both vibration and noise levels increase as bearings reach their later stage of lives.

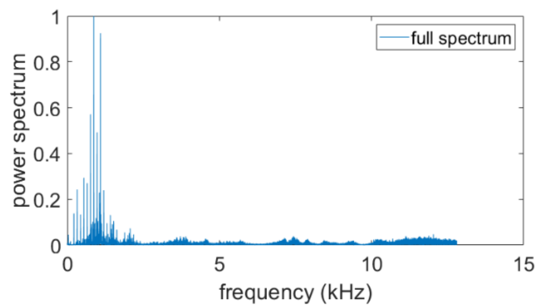
Dictionary learning tends to perform better for more uniform signals in a narrower bandwidth. Therefore Fourier transform is applied to decompose the original signal into signals with different bandwidths. The proposed dictionary learning scheme is applied to decomposed signals individually. The vibration signal between the vertical dash lines in Figure 2 (a) near the end of life is extracted and shown in Figure 2 (b). The extracted signal in the frequency domain after Fourier transform is shown in Figure 2 (c). The original vibration signal in Figure 2 (b) is decomposed into signals with six different bandwidths shown in Figure 3. The testing signal is extracted near the end of life data but from a different period of time, as shown in the region between two solid lines in Figure 2 (a). The extracted testing signal is shown in Figure 4 (a). The testing signal is similarly decomposed into six different bandwidths, as shown in Figure 4 (b). With the proposed physics-constrained dictionary learning scheme, one basis matrix is optimized for each decomposed training signal within the individual bandwidth. Six optimal basis matrices are obtained. Only one measurement matrix is generated, because the same time stamps are physically required during the sampling for storage and transmission for all decomposed signals. The training of the measurement matrix is based on the signal with a bandwidth of 0 ~ 2.5 kHz.



(a) The complete run-to-failure data in horizontal axis



(b) The signal extracted from the near-end of life-time



(c) The extracted signal in frequency domain

Figure 2. Extract the training signal from the complete run-to-failure data.

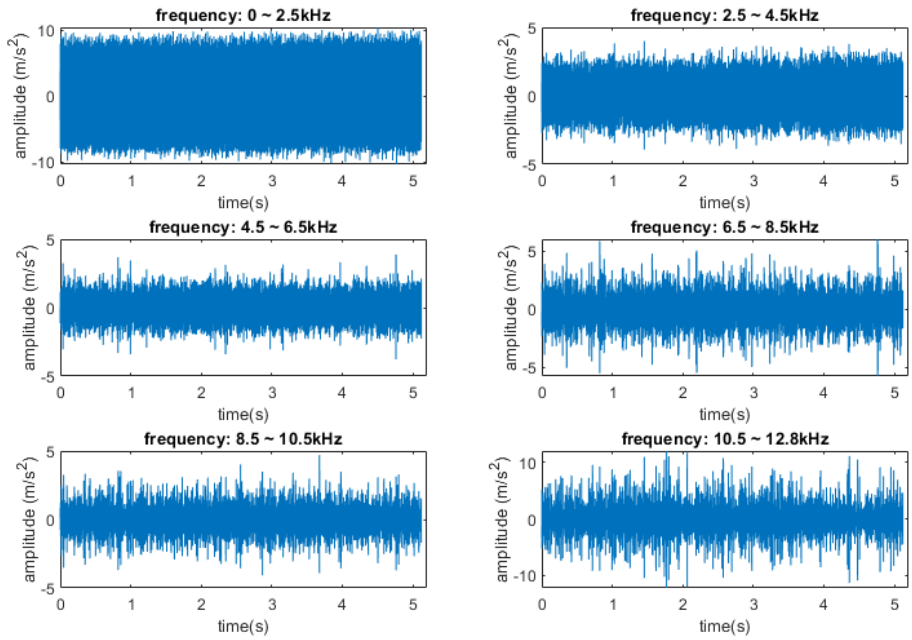
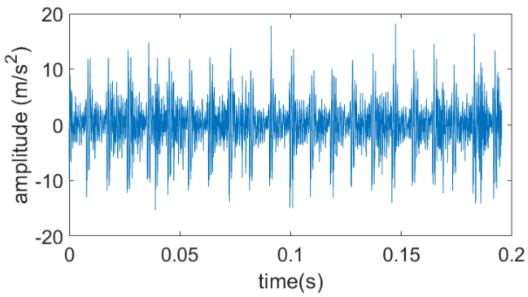
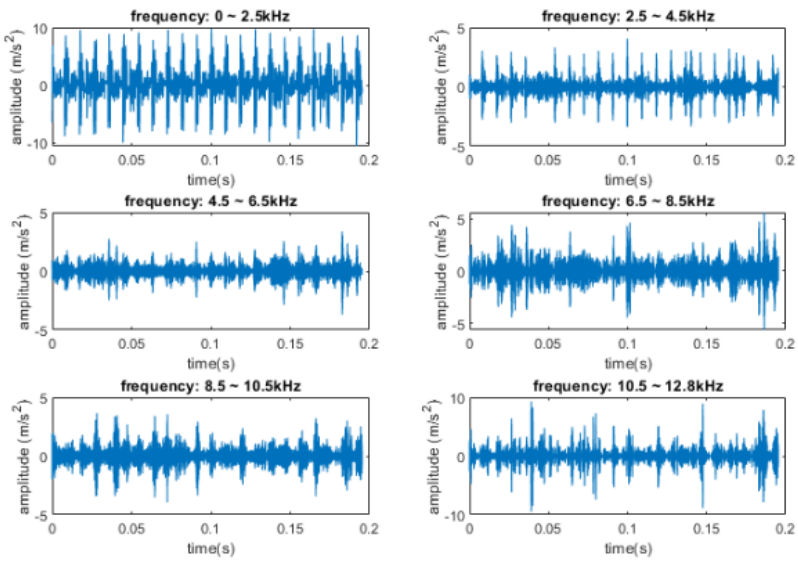


Figure 3. Decomposed training signals with different bandwidths.



(a) Original testing signal



(b) Decomposed testing signals with different bandwidths

Figure 4. Testing signal and the compositions with different bandwidths

4.1. Experimental results without considering the physical constraint

In the first scenario, the constraint in Eq.(4) that indicates the minimum sampling interval is not considered. The measurement matrix $\mathbf{\Phi}$ and the basis matrix $\mathbf{\Psi}$ are trained with the 1150 training segments. The size of the basis matrix $\mathbf{\Psi}$ is 100×300. The maximum number of non-zero values in each coefficient vector $\mathbf{\Upsilon} = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \dots \boldsymbol{\gamma}_P]$ is determined by the modified K-SVD with adaptive T_0. Since a high sparsity level in the coefficient vectors is preferred, initial T_0 is set to be 1 and ΔT_0 is 1. The desired T_0 can be found by gradually increasing its magnitude. Among the 100 collected data points in a segment, the number of stored and transmitted measurements was set to be 35. The time stamps for these 35 measurements are optimized with the physics-constrained dictionary learning scheme. The initial basis matrix $\mathbf{\Psi}$ is the discrete cosine transformation matrix. For the decomposed training and testing signals with a bandwidth of 0 ~ 2.5 kHz, the average reconstruction error for each iteration is shown in Figure 5. It is seen that the reconstruction converges after 10 iterations. In each segment, there are 100 data points and the time period is 0.0039s. The optimal 35 time stamps for samplings are selected from 100 available time stamps and marked as stars in Figure 6 (a), where unselected time stamps are marked as circles. The compression ratio is 100/35=2.86. With 35 measurements in each testing segment, all data points in all testing segments can be reconstructed with very small errors. The reconstruction errors of decomposed signals from 50 testing segments are shown in Figure 7 (a). The reconstructed signals with different bandwidths are combined and compared with the original signal in Figure 4 (a). The reconstruction errors of the combined signal are seen in Figure 7 (b). The maximum reconstruction error of the combined signal is 1.159%, the average error is 0.0024%, and the standard deviation (STD) of errors is 0.0267%.

Sensitivity analysis is also performed with different number of stored and transmitted measurements. With 40 and 45 measurements in each segment, the maximum reconstruction errors of combined signals are 0.4256% and 0.0899% respectively, the average errors are 0.000272% and 0.000117% respectively, and the STD of errors are 0.0063% and 0.0015% respectively. When more measurements are used, reconstruction errors can be reduced. It was found that if the number of measurements in each segment is less than 35, the reconstruction error can become large quickly. Therefore, there is a lower limit on the number of stored and transmitted measurements.

Three different bearing datasets collected in [25] are used to demonstrate the robustness of the proposed framework. Training and testing signals from both of horizontal and vertical axes in the beginning, middle and near end of the life period are used, and the average reconstruction errors are compared in Table 3. It is found that the reconstruction errors for all periods and both axes in three bearing datasets are very small.

The proposed framework is compared with the traditional dictionary learning with K-SVD algorithm by randomly selecting time stamps for stored and transmitted measurements. With a total of 35 measurements in each segment, the reconstruction errors of the combined signal with the traditional K-SVD algorithm can be found in Figure 8. The maximum reconstruction error is 6.09%, the average error is 0.0196%, and the STD of errors is 0.1737%. It is seen that the proposed physics-constrained dictionary learning can significantly improve the reconstruction performance by optimizing the measurement matrix and the basis matrix separately.

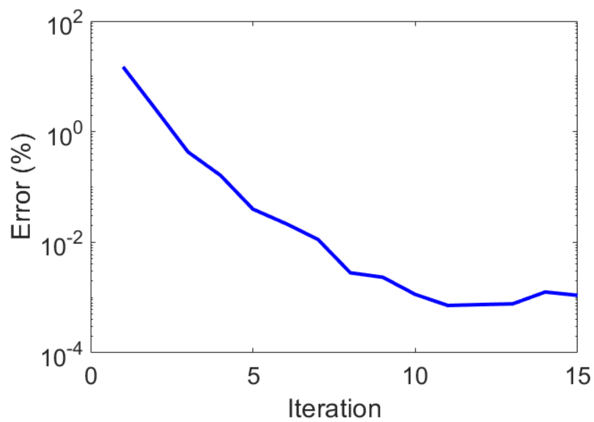


Figure 5. Convergence history of the reconstruction

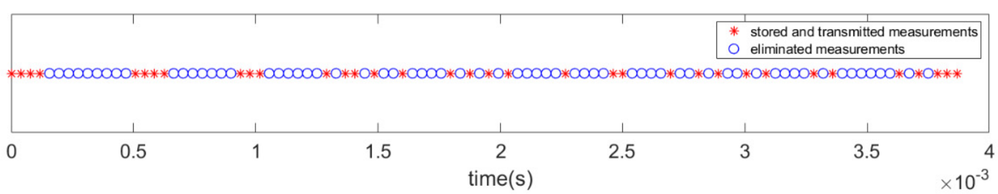
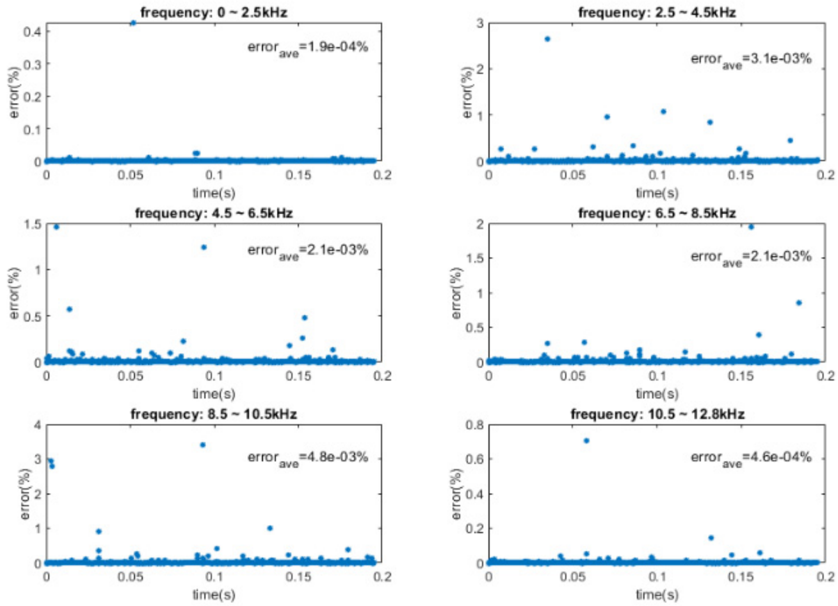
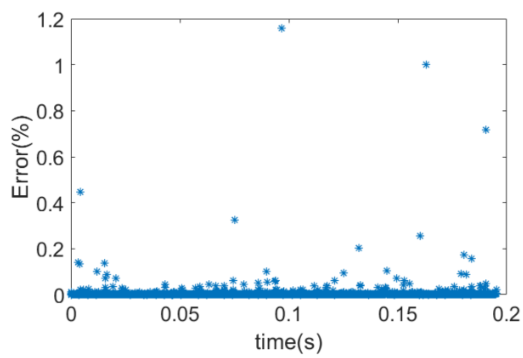


Figure 6. Optimized time stamps for data storage and transmission

This document is the unedited Author's version of a Submitted Work that was subsequently accepted for publication in Mechanical Systems & Signal Processing, copyright © Elsevier. The final edited and published work is at https://doi.org/10.1016/j.ymssp.2020.107434



(a) Reconstruction errors of decomposed signals



(b) Reconstruction errors of the combined signal

Figure 7. Reconstruction error of 50 testing segments

Table 3. Comparison of average reconstruction errors of signals collected for two axes during three different life periods in three datasets

Bearing dataset #	Axis	Beginning	Middle	Near End
1	Horizontal	0.0033%	0.0034%	0.0024%
	Vertical	0.0034%	0.0026%	0.0025%
2	Horizontal	0.0039%	0.0018%	0.0036%
	Vertical	0.0038%	0.001%	0.00075%
3	Horizontal	0.0041%	0.0028%	0.003%
	Vertical	0.0034%	0.0022%	0.00073%

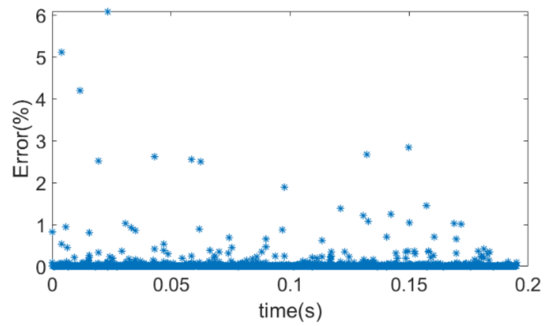


Figure 8. Reconstruction errors of the combine signal from 50 segments with traditional K-SVD

The target sparsity level T_0 in Eq.(3) can affect the training process and reconstruction performance. With a smaller T_0, the original signal is reconstructed with fewer measurements. However, the training error of $||\mathbf{S} - \mathbf{\Psi}\mathbf{Y}||_F^2$ can be large, because too few non-zero values in the coefficient vector can cause significant information loss and the original signal can no longer be represented with the basis matrix. Instead of assuming a constant T_0, the modified K-SVD can determine the desired T_0 adaptively. Here, the result of the adaptive K-SVD is verified with the physics-constrained dictionary learning based on the conventional K-SVD when different values of T_0 are assigned. The reconstruction errors from the conventional K-SVD with different T_0 values are listed in Table 4. For all tests, the number of measurements is 35. The size of the basis matrix $\mathbf{\Psi}$ is 100×300. It is seen in Table 4 that the reconstruction error has the lowest level when $T_0 = 5$. When $T_0 > 11$, the reconstruction error becomes very large, because the lower limit of the number of stored and transmitted measurements is not satisfied. More measurements are needed to reconstruct the original signal if $T_0 > 11$. The convergence to the optimal T_0 with the adaptive K-SVD algorithm is shown in Figure 9, which matches the trend in Table 4. The initial value of T_0 is 1 in the adaptive K-SVD algorithm and adjustment step size ΔT_0 is 1. The incremental process is shown with the solid line. In the second test of the adaptive K-SVD, the initial value of T_0 is 10. The decrease of sparsity level is shown as the dashed line. The average reconstruction error decreases as T_0 decreases and the optimal T_0 is also 5. Comparing results from conventional and modified K-SVD algorithms, the same desired T_0 can be found which is $T_0 = 5$. Since the computational cost is more expensive when T_0 is larger because more non-zeros values in coefficient vectors need to be determined and higher sparsity level is preferred during the reconstruction process, initial T_0 of 1 is suggested.

Table 4. Reconstruction errors with different sparsity levels

T_0	Max error (%)	Average error (%)	STD (%)
1	15.55	0.0072	0.2322
3	6.721	0.0029	0.0961
5	1.159	0.0024	0.0267
7	2.591	0.0035	0.0547
9	8.795	0.0369	0.2835
11	270.5	0.5521	6.0274

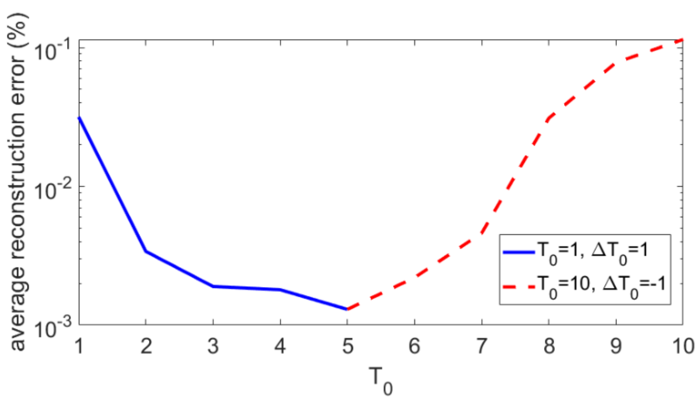


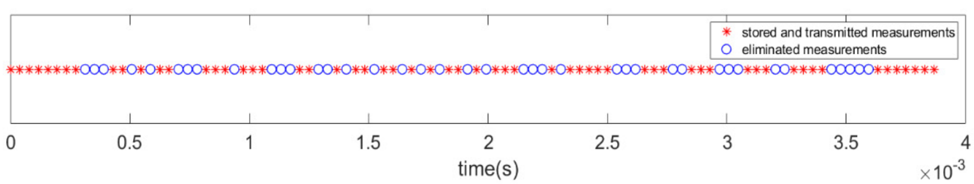
Figure 9. Convergence of T_0

4.2. Experimental results with the physical constraint considered

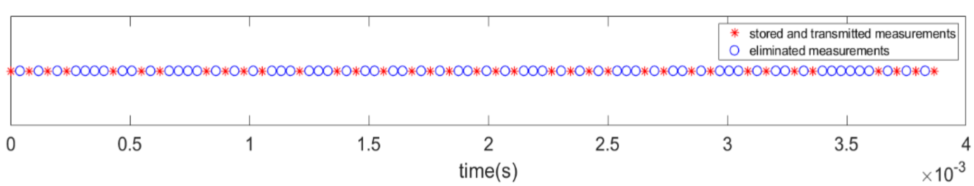
In the second scenario, the constraint in Eq.(4) with the minimum sampling interval is considered. The sampling rate of the training signal is 25.6 kHz, and the sampling interval between consecutive data points is $r_s = 1/25600$ s. In this scenario, different minimum sampling interval of stored and transmitted data points are tested to optimize the measurement matrix. The minimum sampling interval is set to be r_s and $2r_s$ respectively.

The initial number of stored and transmitted measurements in one segment is set to be 60 and optimized time stamps of measurements are indicated in Figure 10 (a) when the minimum sampling interval is r_s. It can be found that some stored and transmitted measurements are too close to each other. The collected information can be redundant. When the minimum sampling interval of $2r_s$ is used, some close-by measurements are eliminated as in Figure 10 (b). The 23 eliminated measurements are marked with circles and the 37 remaining data points to be stored and transmitted are marked as stars in Figure 10 (b).

The reconstruction errors with different minimum sampling intervals are shown in Figure 11. For the minimum sampling intervals of r_s and $2r_s$, the maximum reconstruction errors are 0.1499% and 0.9944%, the average errors are 0.0000773% and 0.0022%, and the STD of errors are 0.0022% and 0.0222% respectively. It is seen that the reconstruction errors increase when the sampling interval is increased, because fewer measurements are used. Compared to the reconstruction result in the first scenario in Section 4.1, where 35 measurements in each segment is used, the compression ratios between the two scenarios are similar. However, less redundant information is stored and transmitted with a higher level of reconstruction accuracy when the physical constraint is applied. The physical constraint of minimum sampling intervals is particularly useful when low cost sensors have limitations in sampling rates or there are bandwidth limitations in transmitting data in a distributed environment.



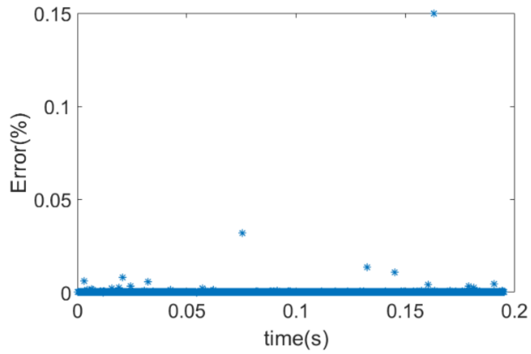
(a) Sampling interval is r_s



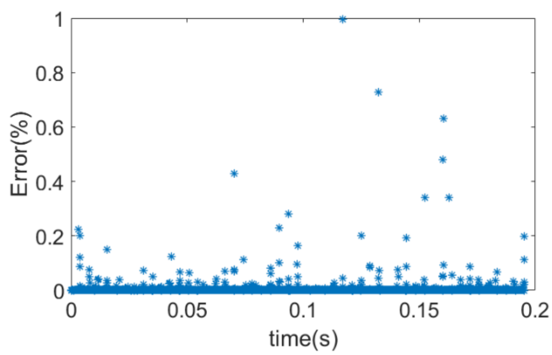
(b) Sampling interval is $2r_s$

Figure 10. Optimized measurements with different minimum sampling intervals as the physical

constraint



(a) Sampling interval is r_s



(b) Sampling interval is $2r_s$

Figure 11. Reconstruction errors with different minimum sampling intervals as the physical constraint

5. Conclusion

In this paper, a new technique is presented to reduce the amount of data in storage and communication to monitor machinery health with vibration signals. Instead of all collected signals, only a few data points are stored and transmitted, which can be used to reconstruct the complete signals. The energy consumption and memory usage in both data storage and communication can be saved. Different from other dictionary learning methods, the measurement matrix in the proposed physics-constrained dictionary learning formulation directly indicates the time stamps of collected data points or measurements. The basis matrix is optimized to sparsely represent the available data points. Compared to the conventional dictionary learning, the measurement and basis matrices are trained separately in the physics-constrained dictionary learning, which significantly reduces the reconstruction errors. The minimum sampling interval is applied as an additional physical constraint to minimize the redundant information of measurements. The constrained FrameSense algorithm allows us to impose physical constraints during the training of measurement matrix. An adaptive K-SVD algorithm is also developed to train the basis matrix and determine the desired number of non-zero values in coefficient vectors on the fly. The proposed approach can be used to customize the basis matrices that target at decomposed signals with different bandwidths. In this way, the information loss due to compression can be minimized.

The major challenge of the proposed physics-constrained dictionary learning is related to the optimization in the high-dimensional space formed by the measurement and basis matrices. One limitation of the current computational scheme comes from the training algorithm that is based on the K-SVD. The K-SVD can only find the local optima. The performance of reconstruction depends on the choices of the initial basis matrix and the reconstruction algorithm. Instead of using the K-SVD, other dictionary learning algorithms can be applied. For instance, the online dictionary learning is more efficient than the K-SVD algorithm when monitoring real-time systems, because it does not need to store and access the entire dataset. The performance of reconstruction also depends on the dimensions of the dataset. High-dimensional data usually exhibit more correlations along multiple dimensions. Basis matrices may not be able to capture all, which affects the accuracy of reconstruction.

In this paper, one-dimensional vibration signals are used to test the proposed approach. In future work, the physics-constrained dictionary learning will be used to reconstruct higher dimensional data such as images and videos. Application-specific physical constraints need to be considered in order to store and transmit signals more efficiently. For example, the constraints can be related to the similarity between frames of the video to minimize the redundant information stored. In large-scale sensor networks, the physical constraints can be designed based on the limitations of communication between sensors and the coverage of the sensor network.

Acknowledgement

This work was supported in part by National Science Foundation under grant CMMI-1663227.

References

[1] E. J. Candes and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," *IEEE Transactions on Information Theory,* vol. 52, no. 12, pp. 5406-5425, 2006.
[2] D. L. Donoho, "Compressed sensing," *Ieee Transactions on Information Theory,* vol. 52, no. 4, pp. 1289-1306, Apr 2006.
[3] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Computation,* vol. 15, no. 2, pp. 349-396, Feb 2003.
[4] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," *Icassp '99: 1999 Ieee International Conference on Acoustics, Speech, and Signal Processing, Proceedings Vols I-Vi,* pp. 2443-2446, 1999.
[5] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *Ieee Transactions on Signal Processing,* vol. 54, no. 11, pp. 4311-4322, Nov 2006.
[6] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online Learning for Matrix Factorization and Sparse Coding," *Journal of Machine Learning Research,* vol. 11, pp. 19-60, Jan 2010.
[7] X. F. Chen, Z. H. Du, J. M. Li, X. Li, and H. Zhang, "Compressed sensing based on dictionary learning for extracting impulse components," *Signal Processing,* vol. 96, pp. 94-109, Mar 2014.
[8] X. X. Ding and Q. B. He, "Time-frequency manifold sparse reconstruction: A novel method for bearing fault feature extraction," *Mechanical Systems and Signal Processing,* vol. 80, pp. 392-413, Dec 1 2016.
[9] L. Guo, H. L. Gao, J. Li, H. F. Huang, and X. C. Zhang, "Machinery vibration signal denoising based on learned dictionary and sparse representation," *11th International Conference on Damage Assessment of Structures (Damas 2015),* vol. 628, 2015.
[10] H. T. Zhou, J. Chen, G. M. Dong, and R. Wang, "Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model," *Mechanical Systems and Signal Processing,* vol. 72-73, pp. 65-79, May 2016.
[11] O. Lorintiu, H. Liebgott, M. Alessandrini, O. Bernard, and D. Friboulet, "Compressed Sensing Reconstruction of 3D Ultrasound Data Using Dictionary Learning and Line-Wise Subsampling," *Ieee Transactions on Medical Imaging,* vol. 34, no. 12, pp. 2467-2477, Dec 2015.
[12] Y. Huang, J. Paisley, Q. Lin, X. H. Ding, X. Y. Fu, and X. P. Zhang, "Bayesian Nonparametric Dictionary Learning for Compressed Sensing MRI," *Ieee Transactions on Image Processing,* vol. 23, no. 12, Dec 2014.
[13] S. Ravishankar and Y. Bresler, "MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning," *Ieee Transactions on Medical Imaging,* vol. 30, no. 5, pp. 1028-1041, May 2011.
[14] S. Ravishankar and Y. Bresler, "Sparsifying Transform Learning for Compressed Sensing Mri," *2013 Ieee 10th International Symposium on Biomedical Imaging (Isbi),* pp. 17-20, 2013.
[15] Z. F. Zhan, J. F. Cai, D. Guo, Y. S. Liu, Z. Chen, and X. B. Qu, "Fast Multiclass Dictionaries Learning With Geometrical Directions in MRI Reconstruction," *Ieee Transactions on Biomedical Engineering,* vol. 63, no. 9, pp. 1850-1861, Sep 2016.
[16] Y. Song, G. B. Yang, H. T. Xie, D. Y. Zhang, and X. M. Sun, "Residual domain dictionary learning for compressed sensing video recovery," *Multimedia Tools and Applications,* vol. 76, no. 7, pp. 10083-10096, Apr 2017.
[17] L. F. Polania and K. E. Barner, "Multi-scale dictionary learning for compressive sensing ECG," in *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)*, 2013, pp. 36-41.
[18] W. S. Dong, X. Li, L. Zhang, and G. M. Shi, "Sparsity-based Image Denoising via Dictionary Learning and Structural Clustering," *2011 Ieee Conference on Computer Vision and Pattern Recognition (Cvpr),* pp. 457-464, 2011.
[19] X. D. Zhang, X. C. Feng, W. W. Wang, and G. J. Liu, "Image denoising via 2D dictionary learning and adaptive hard thresholding," *Pattern Recognition Letters,* vol. 34, no. 16, pp. 2110-2117, Dec 1 2013.
[20] M. Y. Zhou *et al.*, "Nonparametric Bayesian Dictionary Learning for Analysis of Noisy and Incomplete

Images," *Ieee Transactions on Image Processing,* vol. 21, no. 1, pp. 130-144, Jan 2012.
[21] J. M. Duarte-Carvajalino and G. Sapiro, "Learning to Sense Sparse Signals: Simultaneous Sensing Matrix and Sparsifying Dictionary Optimization," *Ieee Transactions on Image Processing,* vol. 18, no. 7, pp. 1395-1408, Jul 2009.
[22] H. Bai, G. Li, S. Li, Q. W. Li, Q. R. Jiang, and L. P. Chang, "Alternating Optimization of Sensing Matrix and Sparsifying Dictionary for Compressed Sensing," *Ieee Transactions on Signal Processing,* vol. 63, no. 6, pp. 1581-1594, Mar 15 2015.
[23] J. Ranieri, A. Chebira, and M. Vetterli, "Near-Optimal Sensor Placement for Linear Inverse Problems," *Ieee Transactions on Signal Processing,* vol. 62, no. 5, pp. 1135-1146, Mar 2014.
[24] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Ieee Transactions on Information Theory,* vol. 53, no. 12, pp. 4655-4666, Dec 2007.
[25] B. Wang, Y. G. Lei, N. P. Li, and N. B. Li, "A Hybrid Prognostics Approach for Estimating Remaining Useful Life of Rolling Element Bearings," *Ieee Transactions on Reliability,* vol. 69, no. 1, pp. 401-412, Mar 2020.