Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 53 (2021) 726–734

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

49th SME North American Manufacturing Research Conference, NAMRC 49, Ohio, USA

Machine Fault Diagnosis of Fused Filament Fabrication Process with Physics-Constrained Dictionary Learning

Yanglong Lu[a], Yan Wang[a]*

[a]Georgia Institute of Technology, 813 First Drive NW, Atlanta, GA 30332-0405, USA

* Corresponding author. Tel.: +1-404-894-4714 ; fax: +1-404-894-9342. *E-mail address:* yan-wang@gatech.edu

Abstract

Sensors have been widely applied in modern manufacturing systems to monitor the processes and machine health conditions in order to control product quality. Processing a large amount of sensor data becomes a new challenge on the efficiency of diagnosis. In this paper, a novel physics-constrained dictionary learning approach is proposed to simultaneously improve the efficiency of data collection with compressed sensing (CS) and perform diagnosis with the classification of sensor data. Two-stage optimization is performed. At the first stage, measurement matrix is optimized to determine the time stamps of collected data points with a fixed basis matrix. This is solved based on a constrained FrameSense algorithm. At the second stage, the basis and classification matrices are optimized with the fixed measurement matrix based on the K-SVD algorithm. The above two optimization steps are repeated until the optimal measurement, classification, and basis matrices converge without further improvement. The recovered signals can be classified more accurately based on the learned classification matrix for specific data. The proposed approach for machine fault diagnosis is demonstrated with acoustic emission signals collected in the fused filament fabrication process.

Keywords: Dictionary learning; Compressed sensing; Sparse coding; Machine health monitoring; Additive manufacturing

1. Introduction

Sensors have been widely applied in modern manufacturing systems to monitor the processes and machine health conditions in order to control product quality. For monitoring and diagnostics, data are collected by sensors and transmitted to processing stations, where necessary signal processing is done. Important features in time and/or frequency domains are then extracted from the data and form the feature vector. Statistical machine learning tools are built and trained with feature vectors as the input to predict the system states. Given the large amount of sensor data collection with the latest sensing systems, processing them for real-time diagnostics becomes challenging [1,2]. Despite significant computational effort is spent on data collection, transmission, and storage, the diagnosis is only based on features that are extracted from raw data [3, 4]. The large amount of collected data often contain excessively redundant information which reduces the sensing efficiency [5]. Another challenge is the reliability of storage systems as the volume of data is growing faster than the storage capacity. Thus the storage efficiency [6] and energy efficiency [7] need to be improved. Bandwidth efficiency which is proportional to the amount of data in transmission also needs to be improved without sacrificing the performance and energy efficiency [8]. Therefore, more efficient schemes for data collection and usage are needed to improve the efficiency of machine health monitoring.

To improve the efficiency of data collection, compressed sensing (CS) and dictionary learning approaches have been developed in the recent decade. CS [9,10] allows us to reduce the amount of data collection by taking advantage of the sparsity of coefficient vector in the reciprocal space. If the original signal has a sparse representation with respect to a basis or transformation matrix, the sparse coefficient vector can be recovered with a few collected samples and the original signal is obtained as the linear combination of the basis matrix and the coefficient vector. When 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 $\boldsymbol{s} =$

10.1016/j.promfg.2021.06.071

$\mathbf{\Psi}\boldsymbol{\gamma}$, where $\mathbf{\Psi} \in \mathbb{R}^{N\times N}$ 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 $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ as $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{s}$, the original signal $\boldsymbol{s}$ can be recovered from the measured data $\boldsymbol{y}$ by solving the inverse problem based on the linear equations $\boldsymbol{y} = \mathbf{\Phi}\boldsymbol{s} = \mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\gamma}$. The basis matrix $\mathbf{\Psi}$ is usually predefined from some known transformation such as discrete cosine transformation, wavelet transformation, or some random matrices that satisfy the restricted isometry property. Dictionary learning methods have been developed to further improve the sparsity of coefficient vectors and the compression ratio of CS by customizing the basis matrix according to specific signal type. The learning process is formulated as an optimization problem where the entries of measurement and basis matrices are optimized to minimize the CS recovery error.

Recently, a physics-constrained dictionary learning method was proposed to compress roller bearing vibration signals [11]. The measurement and basis matrices were optimized simultaneously. There is only one non-zero entry in each row of the measurement matrix and the index of the non-zero entry indicates the time stamp to store the data point. The physical constraint to indicate the maximum sampling rate was also incorporated to minimize the redundant information collection. With a few stored data points, the original vibration signals can be reconstructed.

In this paper, the physics-constrained dictionary learning method is extended with a new formulation to solve classification problems. The measurement, basis and classification matrices are optimized simultaneously. Therefore, with a few collected data points, the proposed physics-constrained dictionary learning method can be applied to classify the signal based on only a few samples. The sensing efficiency and classification accuracy are significantly improved with the proposed method. The amount of data collection can be further reduced with the additional physical constraint to indicate the maximum sampling rate. We demonstrate this new approach in the fused filament fabrication (FFF) process monitoring, where acoustic emission (AE) signals are collected to identify machine states.

In the remainder of the paper, the background of AE technique applied in additive manufacturing process monitoring and dictionary learning methods are introduced in Section 2. The proposed physics-constrained dictionary learning method for classification is described in Section 3. The demonstration of its application to classify AE signals base on different machine conditions, and experimental results are given in Section 4.

2. Background

In this section, the applications of acoustic emission technique in additive manufacturing process monitoring are reviewed. The relate work of dictionary learning methods is introduced.

2.1. Acoustic emission applications in additive manufacturing process monitoring

AE is a non-destructive technique to evaluate mechanical performance of materials [12, 13] and monitor structural and machine health [14, 15] by detecting ultrasonic stress waves originated from some localized sources. Detection and analysis of AE signals can provide valuable information to characterise source mechanisms such as crack, friction, and deformation.

There has been research to monitor additive manufacturing (AM) processes with AE techniques [16]. Different AM processes are available. FFF or material extrusion process is commonly used for polymers such as acrylonitrile butadiene styrene, polylactide, and nylon. Metal AM processes including selective laser melting (SLM), electron beam melting, and direct energy deposition are used for metallic materials such as titanium alloys, stainless steel, and aluminium alloys. Shevchik *et al.* [17, 18] collected AE signals in the SLM process and applied convolution neural networks and spectral convolution neural networks to classify AE features from processes with different build qualities. Wu *et al.* employed the AE technique to identify normal and abnormal states of machine conditions in the FFF process based on support vector machine (SVM) [19, 20], semi-hidden Markov model [21], and self-organizing map [22]. Liu *et al.* [23] proposed a fault diagnosis approach based on linear discriminant analysis (LDA) and the clustering by fast search and find of density peaks (CFSFDP) approach to classify machine faults based on a high dimensional feature space. The classification performances from different unsupervised and supervised approaches were also compared.

2.2. Dictionary learning

The recovery accuracy of CS is affected by the sparsity level of the coefficient vector. In conventional CS, the basis matrix is usually predefined as some known transformation matrices such as Fourier transformation, discrete cosine transformation, wavelet transformation, or some random matrices that have the restricted isometry property. The sparsity level of the original signal is usually low with respect to the predefined basis matrix because it is not directly related to the observed signals. With dictionary learning methods, the basis matrix can be trained based on the collected data. The sparsity level of the coefficient vector is significantly improved because the basis matrix is customized according to the specific data type. The recovery accuracy of CS can also be improved with the trained basis matrix.

Various dictionary learning methods [24] 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 (MOD) [25], K-SVD [26], the online dictionary learning [27] and others. The training process was also based on the maximum likelihood [28], least-square error [29, 30], and hidden Markov model [31].

Particularly, K-SVD is an efficient algorithm for dictionary learning. The algorithm starts with an initial guess of dictionary $\mathbf{D}$. Then the coefficient matrix $\mathbf{\Upsilon}$ is calculated with the orthogonal matching pursuit (OMP) [32] algorithm. With the coefficient matrix $\mathbf{\Upsilon}$ fixed, the basis matrix is updated one column at a time. The update is to minimize the discrepancy between the subspace of this particular column for the training data and the subspace for the recovered data based on the current dictionary, given that all other columns are fixed. The eigenvectors of the subspace subject to the sparsity constraint are obtained by applying the singular value decomposition (SVD). The first left-singular vector is taken as the updated column of dictionary $\mathbf{D}$, whereas the first right-singular vector multiplied by the first singular value is taken as the corresponding row of coefficient matrix $\boldsymbol{\Upsilon}$. Each column of dictionary $\mathbf{D}$ is updated sequentially. With the most recent dictionary, the coefficient matrix $\boldsymbol{\Upsilon}$ needs to be updated again with the OMP algorithm. The above iterations of updates continue until the convergence of the dictionary is reached.

Dictionary learning methods have been applied in combination with CS. Chen et al. [28] applied the dictionary learning method to improve the performance of CS in extracting impulse components from noisy vibration signals. Lorintiu et al. [33] 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 (MRIs) [34, 35, 36, 37], videos [38] and electrocardiogram (ECG) signals [39], and image denoising [40, 41, 42].

Instead of learning the dictionary, which is the combination of the measurement matrix and the basis matrix, approaches to design the measurement matrix and the basis matrix individually were also developed. Duarte-Carvajalino and Sapiro [43] simultaneously optimized the measurement matrix and basis matrix with a new scheme called coupled-KSVD. The incoherence between the measurement matrix and basis matrix is improved which results in the better reconstruction performance. Bai et al. [44] 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 are improved for reconstructing natural images.

Dictionary learning methods have also been applied for classification and clustering [45, 46]. Zhang and Li [47] developed a discriminative K-SVD (D-KSVD) method for face recognition. The D-KSVD method is implemented by adding a discriminative term into the objective function of the original K-SVD algorithm. The D-KSVD method outperforms other existing methods such as the SRC algorithm [48]. Ptucha and Savakis [49] proposed a linear extension of graph embedding K-means-based singular value decomposition (LGE-KSVD) method to solve facial and activity recognition problems. LGE-KSVD utilized variants of the LGE to optimize the K-SVD problem. Other dictionary learning methods for classification include label consistent K-SVD [50], discriminative Bayesian dictionary learning [51], and task-driven dictionary learning [52].

The proposed framework in this paper is developed to solve classification problems by optimizing the measurement, basis, and classification matrices simultaneously. Therefore, instead of collecting the complete signal, only a few data points at the time stamps determined from the measurement matrix are needed to recover the sparse coefficient vectors with the optimal basis matrix and CS. The signal can be classified by multiplying the classification matrix with the recovered coefficient vectors. The measurement matrix is designed to determine the time stamps of stored data points in one-dimensional signals in physical experiments, which typically requires that there is only one non-zero entry in each row of the measurement matrix. The considerations of physical constraints and interpretations in measurement conditions are important in engineering applications.

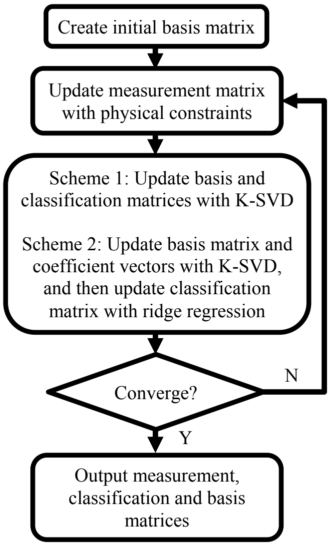

Fig. 1. Physics-constrained dictionary learning scheme

3. Methodology

The physics-constrained dictionary learning scheme in [11] has been developed to optimize the measurement and basis matrices simultaneously to reconstruct the original signal with a few data points. For machine fault diagnosis, the physics-constrained dictionary learning scheme is modified, so that the classification errors can also be minimized. It is to solve

$$\min_{\Phi,\Psi,W,\Upsilon}\begin{pmatrix}\alpha\|\mathbf{S}-\boldsymbol{\Psi\Upsilon}\|_F^2+\|\boldsymbol{\Phi}\mathbf{S}-\boldsymbol{\Phi\Psi\Upsilon}\|_F^2\\+\beta\|\mathbf{L}-\mathbf{C}\boldsymbol{\Upsilon}\|_F^2\end{pmatrix} \tag{1}$$

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

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

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

where F denotes the Frobenius norm, $\mathbf{S}=[\boldsymbol{s}_1,\boldsymbol{s}_2\dots\boldsymbol{s}_P]\in\mathbb{R}^{N\times P}$ contains P sets of training data and each data set $\boldsymbol{s}_j$ has the length of N. $\boldsymbol{\Psi}\in\mathbb{R}^{N\times W}$ is the basis matrix with $N<W$ and $W\ll P$. $\boldsymbol{\Upsilon}=[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\dots\boldsymbol{\gamma}_P]\in\mathbb{R}^{W\times P}$ contains the sparse coefficients that represent the training data in $\mathbf{S}$ with respect to the basis matrix. $\mathbf{C}\in\mathbb{R}^{N\times W}$ is the classification matrix and $\mathbf{L}=[\boldsymbol{l}_1,\boldsymbol{l}_2\dots\boldsymbol{l}_P]\in\mathbb{R}^{N\times P}$ is the class label of the training signals. $\boldsymbol{l}_i=[0,1,0\dots0]$ where the index of the non-zero value in $\boldsymbol{l}_i$ indicates the class. Different classes indicate different machine states. Lagrange multipliers α and β are applied to control the relative contribution of each term in Eq.(1). The constraint in Eq.(2) indicates the training sequence. That is, in the iterative optimization procedure, the basis matrix $\boldsymbol{\Psi}$ is fixed first, measurement matrix $\boldsymbol{\Phi}$ can be optimized based on the fixed $\boldsymbol{\Psi}$. Then the basis and classification matrices are optimized next. The constraint in Eq.(3) is the upper limit of the sparsity level, where $\boldsymbol{\gamma}_i$ is the *i*-th column of the 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 limitation of collected and stored data points, such

as the time interval between collected and stored data points in 1D signals need to be larger than a threshold value r. Other physical constraints can be added similarly.

Here, two different physics-constrained dictionary learning schemes are used to classify different machine faults based on a few collected data points as shown in Fig. 1. In the first scheme, the measurement, basis and classification matrices are optimized simultaneously. The learning procedure starts with an initial guess of the basis matrix. Two stages are performed in each iteration. At stage one, the measurement matrix is optimized to determine the time stamps of collected data points with the fixed basis matrix. This can be solved based on the constrained FrameSense algorithm [11]. At stage two, the basis and classification matrices can be optimized with the fixed measurement matrix based on the K-SVD algorithm. The above two optimization steps are repeated until both the optimal measurement, classification and basis matrices converge without further improvement. In the second scheme, only the basis matrix is optimized with the K-SVD algorithm. The classification matrix is computed separately with the coefficient vectors for the training dataset based on the ridge regression. Classification results from the two different schemes are compared.

3.1. Stage one optimization

At stage one, with the basis matrix $\mathbf{\Psi}$ fixed, the measurement matrix $\mathbf{\Phi}$ is optimized to determine the time stamps of collected data points. Determining the optimal time stamps from all available ones is often NP-hard if the amount of data collection is large. Therefore, a greedy algorithm called constrained FrameSense [11] is used to determine the near-optimal time stamps of collected data points as shown in Table 1. Given all available time stamps $\mathcal{N} = \{1, \dots, N\}$, 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 [53]

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

where $\mathrm{H}(\mathbf{\Psi})$ is the frame potential and represented as

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

where λ_i is the i-th largest eigenvalue of $\mathbf{\Psi}^*\mathbf{\Psi}$ and $\mathbf{\Psi}^*$ is the conjugate transpose of $\mathbf{\Psi}$. $\mathbf{\Psi}_{\mathcal{N}\backslash\mathcal{T}}$ is a sub-matrix of $\mathbf{\Psi}_{\mathcal{N}}$ 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

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

where the column index of the value of 1 in each row indicates the time stamp of each collected data point.

The physical constraint in Eq.(4) is also incorporated in the constrained FrameSense algorithm. If the time interval between any of two collected data points 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 that is determined by the resolution of sensors.

Table 1. The constrained FrameSense algorithm

1. Initialize time stamps of collected data points $\mathcal{L}$, all available time stamps $\mathcal{N}$, and desired number of collected data points m_t
2. Determine the first two removed rows in $\mathbf{\Psi}$ by solving $\mathcal{T} = \mathrm{argmax}_{i,j\in\mathcal{N}} |< \varphi_i, \varphi_j >|^2$ 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 $i^* = \mathrm{argmax}_{i\in\mathcal{L}} \mathrm{F}(\mathcal{T} \cup \{i\})$, where $\mathrm{F}(\mathcal{T} \cup \{i\})$ is the function in Eq.(5)
   - Update unsuitable time stamps of collected data points as $\mathcal{T} = \mathcal{T} \cup \{i^*\}$
   - Update available time stamps of collected data point as $\mathcal{L} = \mathcal{L}\backslash\{i^*\}$

   END WHILE
4. FOR i = 1 to the length of $\mathcal{L}$
   - FOR j = 1 to length of $\mathcal{L}$
     - If $I_{ij}(\mathbf{\Phi}) = |t_i - t_j| \le r$, 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

The objective function in Eq.(1) can be converted to the form of

$$\min_{\Phi,\Psi,\mathbf{Y}} \left\| \begin{pmatrix} \alpha\mathbf{S} \\ \mathbf{\Phi S} \\ \beta\mathbf{L} \end{pmatrix} - \begin{pmatrix} \alpha\mathbf{\Psi} \\ \mathbf{\Phi\Psi} \\ \beta\mathbf{C} \end{pmatrix} \mathbf{Y} \right\|_F^2 \tag{8}$$

With $\mathbf{X} = \begin{pmatrix} \alpha\mathbf{S} \\ \mathbf{\Phi S} \\ \beta\mathbf{L} \end{pmatrix}$, $\mathbf{Z} = \begin{pmatrix} \alpha\mathbf{\Psi} \\ \mathbf{\Phi\Psi} \\ \beta\mathbf{C} \end{pmatrix}$ is optimized by solving Eqs. (8) and (3) with the K-SVD algorithm [5]. The sub-matrix $\mathbf{Z_1} = \begin{pmatrix} \alpha\mathbf{\Psi} \\ \mathbf{\Phi\Psi} \end{pmatrix}$ is used to obtain the basis matrix $\mathbf{\Psi}$ by solving

$$\mathbf{\Psi} = (\alpha^2\mathbf{I} + \mathbf{\Phi}^\mathbf{T}\mathbf{\Phi})^{-\mathbf{1}}[\,\alpha\mathbf{I} \quad \mathbf{\Phi}^\mathbf{T}]\mathbf{Z_1} \tag{9}$$

With the optimized sub-matrix $\mathbf{Z_2} = \beta\mathbf{C}$, the classification matrix can be obtained as $\mathbf{C} = \mathbf{Z_2}/\beta$.

The basis matrix is then normalized as

$$\mathbf{\Psi}' = [\varphi_1', \varphi_2', \dots, \varphi_W',] = \left[\frac{\varphi_1}{\|\varphi_1\|_2}, \frac{\varphi_2}{\|\varphi_2\|_2}, \dots, \frac{\varphi_W}{\|\varphi_W\|_2},\right] \tag{10}$$

and the corresponding classification matrix is

$$\mathbf{C}' = [c_1', c_2', \dots, c_W',] = \left[\frac{c_1}{\|\varphi_1\|_2}, \frac{c_2}{\|\varphi_2\|_2}, \dots, \frac{c_W}{\|\varphi_W\|_2},\right] \tag{11}$$

where φ_{i} and $c_{\mathrm{i}}$are each column in the original basis matrix $\mathbf{\Psi}$ and classification matrix $\mathbf{C}$. The normalized basis matrix $\mathbf{\Psi}'$ and the corresponding classification matrix $\mathbf{C}'$ are used for the classification of testing signals.

The proposed physics-constrained dictionary learning algorithm for classification is shown in Table 2.

Table 2. Physics-constrained dictionary learning for classification

Initialize the basis matrix $\mathbf{\Psi}$ and m=0;
WHILE $m <$ total number of iterations DO
1. Compute $\mathbf{\Phi}$ based on the constrained FrameSense algorithm
2. With $\mathbf{X} = \begin{pmatrix} \alpha\mathbf{S} \\ \mathbf{\Phi S} \\ \beta\mathbf{L} \end{pmatrix}$, $\mathbf{Z} = \begin{pmatrix} \alpha\mathbf{\Psi} \\ \mathbf{\Phi\Psi} \\ \beta\mathbf{C} \end{pmatrix}$ and $\mathbf{\Upsilon}$ are updated by solving Eq.(8) and E1.(3) with K-SVD method.
3. Update $\mathbf{\Psi}$ with Eq.(9) and $\mathbf{C} = \mathbf{Z_2}/\beta$
4. Normalize $\mathbf{\Psi}$ as $\mathbf{\Psi}'$ and obtain the corresponding $\mathbf{C}'$
5. $m = m + 1$
END WHILE

3.3. Optimization of classification matrix with ridge regression

In Section 3.2, the basis and classification matrices are optimized simultaneously with the fixed measurement matrix based on the K-SVD algorithm. The classification matrix can be obtained from the sub-matrix $\mathbf{Z_2}$. An alternate way to optimize the classification matrix is based on the ridge regression. The basis matrix and coefficient vectors can be optimized simultaneously by solving

$$\min_{\Phi,\Psi,\Upsilon} \left\| \begin{pmatrix} \alpha\mathbf{S} \\ \mathbf{\Phi S} \end{pmatrix} - \begin{pmatrix} \alpha\mathbf{\Psi} \\ \mathbf{\Phi\Psi} \end{pmatrix} \mathbf{\Upsilon} \right\|_F^2 \quad (12)$$

$\mathbf{Z_1} = \begin{pmatrix} \alpha\mathbf{\Psi} \\ \mathbf{\Phi\Psi} \end{pmatrix}$ and $\mathbf{\Upsilon}$ can be optimized with the K-SVD algorithm, and $\mathbf{\Psi}$ is computed based on Eq.(9). The classification matrix is then solved by the ridge regression model as

$$\mathbf{C} = (\mathbf{\Upsilon}^T\mathbf{\Upsilon} + \beta'\mathbf{I})^{-1}\mathbf{\Upsilon}\mathbf{L}^T \quad (13)$$

where $\beta' > 0$ is a ridge parameter. This approach can reduce the computational cost of K-SVD because the matrix with lower dimension is optimized. However, the classification accuracy can also be affected because the classification matrix is not optimized simultaneously with the basis matrix.

4. Experiments

The proposed physics-constrained dictionary learning scheme was applied to diagnose machine failures in the FFF process based on AE signal. The data acquired by Wu *et al.* [20] are used in this experiment, where AE signals from AE sensor are processed and AE hits are counted. To reduce the memory usage, time-domain features are stored in output files and used for analysis. The time-domain features of AE hits include amplitude, signal strength, counts, duration, average signal level (ASL), root mean square (RMS) and absolute energy, which can be used to identify machine faults. Among all features, ASL, RMS, and signal strength are selected for the machine fault diagnosis, because these features contain more information than other features such as amplitude, counts, and duration.

RMS, which is used to describe the strength of AE signal in the time domain, is defined as

$$AE_{\mathrm{RMS}} = \sqrt{\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} u(t)^2 dt} \quad (14)$$

where $u(t)$ is the output voltage of AE sensor. ASL is defined as the average of the AE signal amplitude in a logarithmic scale and expressed as

$$AE_{\mathrm{ASL}} = \sqrt{\frac{1}{t_2 - t_1}\int_{t_1}^{t_2} 20\log\left(\frac{u(t)}{u_{ref}}\right) dt} \quad (15)$$

where u_{ref} is the reference voltage. Signal strength is defined as

$$AE_{\mathrm{str}} = \int_{t_1}^{t_2} |u(t)| dt \quad (16)$$

A Hyrel3D printer was used in this experiment, and the AE sensor was attached on the side surface of the extruder to collect signals with different machine conditions such as normal operating condition, material loading, blocked material and running out of material. The AE sensor made by Mistragroup has the operating frequency response within the range of 100-900 kHz and the temperature range between -65 and 177°C. The original AE signal is conditioned and amplified by a PAC 2/4/6 preamplifier, and received by a PAC PCI-2 fast data acquisition (DAQ) system. The experimental setup can be found in Fig. 2. The sampling rate was 5 M samples per second. RMS, ASL and signal strength in the normal operation condition and extruder blockage condition are shown in Fig. 3. These AE features under other machine conditions such as material loading and running out of material are also generated similarly. The classification of machine conditions with the proposed physics-constrained dictionary learning method is based on the features of RMS, ASL, and signal strength.

The collected values for different features under different machine conditions are divided into two regions, as shown in Fig. 3. The left region which consists of 75% of data points is used as the training dataset, and the remaining data points are used as the testing dataset. For instance, for RMS, the training dataset contains 1500 segments for each machine condition and each segment contains 200 successive RMS values. For each segment, the first RMS value is randomly selected from the c-

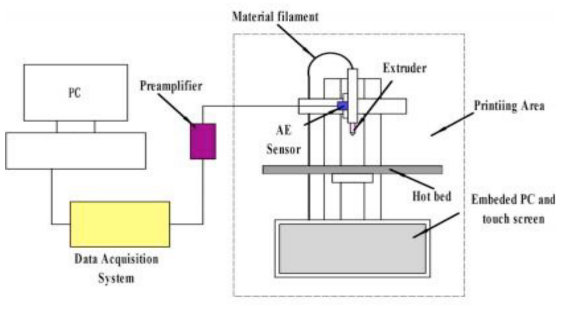

Fig. 2. Experimental setup [20]

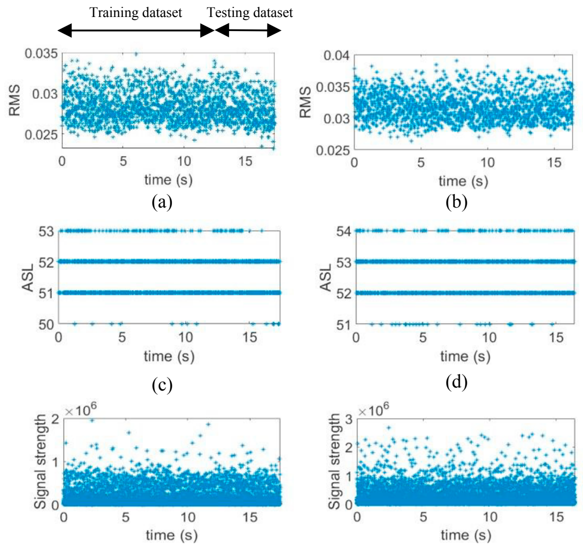

Fig. 3. RMS collected in time domain at (a) normal and (b) extruder blockage conditions; ASL collected at (c) normal and (d) extruder blockage conditions; Signal strength collected at (e) normal and (f) extruder blockage conditions.

omplete time period and the remaining 199 successive RMS values are selected accordingly. The testing dataset consisting of 300 segments for each machine condition is generated similarly. The collected RMS values at normal, blocked material, material loading and running out of material conditions are labelled as different classes. The training and testing datasets for ASL and signal strength are generated similarly.

4.1. Classification results without considering the physical constraint

With the training dataset corresponding to each class, the physics-constrained dictionary learning method in Table 2 is used to optimize the measurement, basis and classification matrices. The constraint in Eq.(4) that indicates the minimum sampling interval of collected samples is not considered. Therefore, the step 4 in Table 1 is eliminated. The size of the basis matrix $\mathbf{\Psi}$ is 200×600. The initial basis matrix is created by randomly selecting 600 columns in the training dataset. With the optimized measurement matrix $\mathbf{\Phi}$ and basis matrix $\mathbf{\Psi}$, the sparse coefficient vectors $\mathbf{Y_t}$ for the testing dataset can be recovered with CS based on OMP [32]. The time stamps of the reduced amount of data collection are indicated in the measurement matrix. With the optimized classification matrix $\mathbf{C}$, the class label $\mathbf{L_t}$ for the testing dataset is obtained by $\mathbf{CY_t}$. Therefore, with the proposed physics-constrained dictionary learning, the class label of the original signal can be determined with the reduced amount of data collection. The class label is then used to identify the machine condition. In Section 4.1.1, three machine conditions are identified based on different features of AE signal such as RMS, ASL and signal strength. In Section 4.1.2, four machine conditions are identified based on RMS values.

4.1.1. Classification of three machine conditions

RMS, ASL, and signal strength of AE signal are used to identify machine conditions such as normal operation condition, extruder blockage, and running out of material. The classification errors for each feature and machine condition are shown in Table 3. The classification error is computed as

$$e_c = \frac{N_i}{N_t} \times 100\% \qquad (17)$$

where N_i is the number of incorrect labels and N_t is the number of total labels. For RMS and ASL, the number of non-zero values in the coefficient vectors, T_0, is set to be 1 and 80 data points in each segment are collected. The compression ratio is 200/80=2.5. For signal strength, the similarity level of collected data points is high for different machine conditions. Therefore, more non-zero values in the coefficient vectors are required to identify different machine conditions and T_0 is set to be 15. As the sparsity level is low, more data points need to be collected for classification and 140 data points are collected in each segment for the signal strength dataset. The compression ratio is 200/140=1.4.

The results show that the proposed physic-constrained dictionary learning can be used to diagnose machine faults with different datasets. However, the classification performance varies among the three different features. The classification based on RMS has the smallest error and the amount of data collection is significantly reduced. The optimized indices for collected and stored 80 RMS values are shown in Fig. 4. Among 200 RMS values in each segment, only ones marked as stars are collected and unselected ones are marked as circles.

Sensitivity analysis is performed with different sparsity levels of each coefficient vector $\mathbf{Y} = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2 \dots \boldsymbol{\gamma}_P]$. The results are shown in Table 4. Among the 200 collected RMS values in each segment, the number of collected RMS values is set to be 80. It is found that as the maximum number of non-zeros values in each coefficient vector increases, the classification errors are increased because redundant information can be generated when more non-zero values in the coefficient vector are used. The important features of the original signal which are critical for the classification accuracy are represented by only a few non-zero values in the coefficient vector.

Table 3. Classification errors for different machine conditions based on different features.

	normal condition	extruder blockage	running out of material
RMS	4.3%	2%	1%
ASL	8%	3.3%	6%
Signal strength	9%	5%	0%

Table 4. Classification errors with different values of T_0.

T_0	normal condition	extruder blockage	running out of material
1	4.3%	2%	1%
3	4.3%	4.6%	9%
5	5.6%	6.6%	11.6%
8	5.6%	7%	13.6%
11	6.3%	7.3%	15%

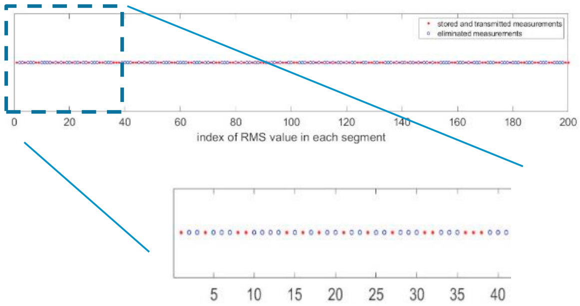

Fig. 4. Optimized indices of stored RMS values without considering the physical constraint

Another sensitivity analysis is performed with different numbers of collected and stored RMS values. The results are shown in Table 5. The maximum number of non-zero values in the coefficient vector is set to be 1. It is found that as more RMS values are collected, the classification errors are reduced, because more information can be used to recover the coefficient vector. However, the efficiency of data collection and storage is reduced.

Table 5. Classification errors with different numbers of stored RMS values.

number of stored values	normal condition	extruder blockage	running out of material
10	25.3%	19.3%	11%
40	2%	1.3%	5%
80	4.3%	2%	1%
120	3%	2.3%	2%
160	3%	2.3%	1.3%

Instead of optimizing measurement, basis and classification matrices simultaneously, the approach in Section 3.3 is also used to optimize the classification matrix separately. In each iteration, with the training dataset, the basis matrix and coefficient vectors are optimized based on the K-SVD algorithm. With the optimized coefficient vectors and the class label for the training dataset, the classification matrix $\mathbf{C}$ is computed based on ridge regression in Eq.(7). The coefficient vectors for the testing dataset can be recovered with OMP and the class label for the testing dataset is then obtained by $\mathbf{L_t} = \mathbf{CY_t}$. With 80 RMS values collected and $T_0 = 1$, the classification errors for normal operation condition, extruder blockage, and running out of material conditions are 3.3%, 3.3%, and 7.6% respectively. Compared to the results in Table 3 with RMS, the classification error is increased when the classification matrix is optimized separately.

4.1.2. Classification of four machine conditions

The proposed physics-constrained dictionary learning approach is also used to identify four machine conditions. Here, the feature of RMS is used. With 130 RMS values used in each segment and the compression ratio of 200/130=1.5, the classification errors for machine conditions of normal operating condition, material loading, extruder blockage, and running out of material are 4.6%, 0%, 0% and 2.3% respectively. Previously Liu *et al.* [23] proposed a fault diagnosis approach based on LDA and the CFSFDP approach to identify machine conditions. It outperforms other commonly used classification methods such as hidden Markov model, SVM, genetic algorithm-based back propagation neural network model and probabilistic neural network. The classification errors for four machine conditions in Ref. [23] based the same AE signal used here are 0%, 6%, 0% and 3% respectively. In comparison, the proposed physics-constrained dictionary learning method gives more accurate classifications of material loading, extruder blockage, and running out of material. Furthermore, the amount of data collection is also reduced with the proposed method.

4.2. Classification results with the physical constraint

In this scenario, the physical constraint in Eq.(4) to indicate the minimum time interval between collected RMS values is considered. Instead of determining the minimum time interval, the minimum difference between indices of collected RMS values is used because the time period for the computation of each RMS value can be different. The minimum difference is set to be 2 so that one of the two adjacent RMS values is eliminated and the redundant information can be minimized. With the original number of stored RMS values set to be 150, the physics-constrained dictionary learning method in Table 2 is used to optimize the measurement, basis, and classification matrices. The maximum number of non-zero values in the coefficient vector is set to be 1. The optimized indices for collected RMS values are shown in Fig. 5. Among 200 RMS values in each segment, only 85 values marked as stars are collected and unselected ones are marked as circles. Previously in Fig. 4, the collected information can be redundant as collected RMS values are too close to each other. Therefore, to minimize the redundant information, the close-by RMS values are eliminated with the additional physical constraints in Fig. 5. The classification errors for three machine conditions including normal operation condition, extruder blockage, and running out of material are 4%, 1%, and 1% respectively. Compared to the results in Table 3 with RMS, the classification error is reduced with the similar number of collected values by considering the physical constraint. When the minimum difference between indices of stored RMS values is set to be 3, the number of stored RMS values is further reduced to 62. The classification errors for three machine conditions are 5.2%, 2%, and 1.3% respectively. The classification errors are increased as the minimum difference between indices of collected RMS values increases as fewer data points are collected.

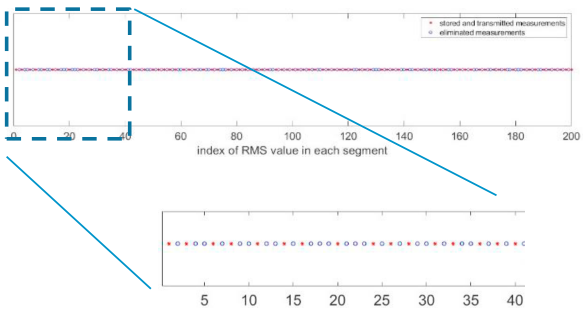

Fig. 5. Optimized indices of stored RMS values with the physical constraint

The proposed physics-constrained dictionary learning can be applied to identify different numbers of machine states with various features of AE signal. However, the classification performance also depends on the similarity of data points for different machine states. The input parameters such as the sparsity level of the coefficient vectors, the size of the basis matrix, and the number of collected data points can affect the classification performance. There is a trade-off between the accuracy and efficiency when the appropriate input parameters are selected. With the physical constraint to minimize the time interval between data points, the classification accuracy and sensing efficiency are improved. With more application-specific physical constraints included, the classification performance can be further improved.

5. Conclusion

In this paper, a new physics-constrained dictionary learning method for classification is proposed to diagnose machine faults. The physics-constrained dictionary learning method is implemented in two schemes. In the first scheme, a two-stage optimization is performed. At stage one, the measurement matrix is optimized with the fixed basis matrix based on the constrained FrameSense algorithm. At stage two, the basis and classification matrices are optimized simultaneously with the fixed measurement matrix based on the K-SVD algorithm. In the second scheme, the classification matrix is computed separately based on the ridge regression. Sensitivity analyses with different maximum numbers of non-zeros values in the coefficient vectors and different amounts of data collection are also performed. In addition to RMS, other features such as ASL and signal strength are also used to identify machine faults.

The proposed physics-constrained dictionary learning method can be used to classify machine states with only a few data points. Therefore, the required memory usage for data storage can be significantly reduced in monitoring machine conditions. It is shown that as few as 40% of the original data are required to successfully identify machine faults. With the physical constraint to minimize the redundant information collection, the classification error and the amount of stored data points can be further reduced.

The major challenge of the proposed physics-constrained dictionary learning comes from the training algorithm that is based on the K-SVD. The K-SVD can only find the local optima. The performance of CS recovery depends on the choices of the initial basis matrix and the recovery algorithm. In this work, the initial basis matrix is generated by randomly selecting a few columns of the training dataset. However, selected columns may not represent the characteristic of all categories since the training dataset consists of signals from different classes. Therefore, a better initialization method needs to be developed to improve the classification accuracy.

Here, the proposed physics-constrained dictionary learning method is applied to identify machine states based only on one feature of AE signal. More information can be obtained if several features are fused for classification. Therefore, the proposed method can be extended to solve classification problems in the high-dimensional feature space. The proposed method in combination with dimensionality reduction techniques such as principal component analysis and LDA can also be developed. In future work, the physics-constrained dictionary learning will also be used to classify higher dimensional signals 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 videos 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.

Acknowledgements

This work was supported in part by the George W. Woodruff Faculty Fellowship and the U.S. National Science Foundation under Grant CMMI-1663227.

References

[1] Dai, H. N., Wang, H., Xu, G., Wan, J., & Imran, M. (2020). Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. *Enterprise Information Systems*, *14*(9-10), 1279-1303.

[2] Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. *Sensors*, *18*(9), 2946.

[3] Lei, Y., Jia, F., Lin, J., Xing, S., & Ding, S. X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, *63*(5), 3137-3147.

[4] He, Q. P., & Wang, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. *Journal of Process Control*, *67*, 35-43.

[5] Cărbunar, B., Grama, A., Vitek, J., & Cărbunar, O. (2006). Redundancy and coverage detection in sensor networks. *ACM Transactions on Sensor Networks (TOSN)*, *2*(1), 94-128.

[6] Nachiappan, R., Javadi, B., Calheiros, R. N., & Matawie, K. M. (2017). Cloud storage reliability for big data applications: A state of the art survey. *Journal of Network and Computer Applications*, *97*, 35-47.

[7] Butt, A. R., Bhattacharjee, P., Wang, G., & Gniady, C. (2014). Exploring Trade-Offs between Energy Savings and Reliability in Storage Systems. *The Green Computing Book: Tackling Energy Efficiency at Large Scale*, *149*.

[8] Azoidou, E., Pang, Z., Liu, Y., Lan, D., Bag, G., & Gong, S. (2017). Battery lifetime modeling and validation of wireless building automation devices in thread. *IEEE Transactions on Industrial Informatics*, *14*(7), 2869-2880.

[9] E. J. Candes and T. Tao (2006) Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?. IEEE Transactions on Information Theory, vol. 52, no. 12, pp. 5406-5425.

[10] D. L. Donoho (2006) Compressed sensing. IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289-1306.

[11] Lu, Y., & Wang, Y. (2020) A Physics-Constrained Dictionary Learning Approach for Compression of Vibration Signals, *Mechanical Systems and Signal Processing*, *153*, 107434.

[12] Tragazikis, I. K., Dassios, K. G., Exarchos, D. A., Dalla, P. T., & Matikas, T. E. (2016). Acoustic emission investigation of the mechanical performance of carbon nanotube-modified cement-based mortars. *Construction and Building Materials*, *122*, 518-524.

[13] De Rosa, I. M., Santulli, C., & Sarasini, F. (2009). Acoustic emission for monitoring the mechanical behaviour of natural fibre composites: A literature review. *Composites part a: applied science and manufacturing*, *40*(9), 1456-1469.

[14] Manthei, G., & Plenkers, K. (2018). Review on in situ acoustic emission monitoring in the context of structural health monitoring in mines. *Applied Sciences*, *8*(9), 1595.

[15] Rabiei, M., & Modarres, M. (2013). Quantitative methods for structural health management using in situ acoustic emission monitoring. *International Journal of Fatigue*, *49*, 81-89.

[16] Wasmer, K., Kenel, C., Leinenbach, C., & Shevchik, S. A. (2017). In Situ and Real-Time Monitoring of Powder-Bed AM by Combining Acoustic Emission and Artificial Intelligence. In *International Conference on Additive Manufacturing in Products and Applications*, pp. 200-209. Springer, Cham

[17] Shevchik, S. A., Kenel, C., Leinenbach, C., & Wasmer, K. (2018). Acoustic emission for in situ quality monitoring in additive manufacturing using spectral convolutional neural networks. *Additive Manufacturing*, *21*, 598-604.

[18] Shevchik, S. A., Masinelli, G., Kenel, C., Leinenbach, C., & Wasmer, K. (2019). Deep learning for in situ and real-time quality monitoring in additive manufacturing using acoustic emission. *IEEE Transactions on Industrial Informatics*, *15*(9), 5194-5203.

[19] Wu, H., Yu, Z., & Wang, Y. (2016a). A new approach for online monitoring of additive manufacturing based on acoustic emission. In *ASME 2016 11th International Manufacturing Science and Engineering Conference*, pp. V003T08A013-V003T08A013.

[20] Wu, H., Wang, Y., & Yu, Z. (2016b). In situ monitoring of FDM machine condition via acoustic emission. *The International Journal of Advanced Manufacturing Technology*, **84**(5-8), 1483-1495.

[21] Wu, H., Yu, Z., & Wang, Y. (2017). Real-time FDM machine condition monitoring and diagnosis based on acoustic emission and hidden semi-Markov model. *The International Journal of Advanced Manufacturing Technology*, **90**(5-8), 2027-2036.

[22] Wu, H., Yu, Z., & Wang, Y. (2019). Experimental study of the process failure diagnosis in additive manufacturing based on acoustic emission. *Measurement*, *136*, 445-453.

[23] Liu, J., Hu, Y., Wu, B., & Wang, Y. (2018). An improved fault diagnosis approach for FDM process with acoustic emission. *Journal of Manufacturing Processes*, *35*, 570-579.

[24] Kreutz-Delgado, K., Murray, J. F., Rao, B. D., Engan, K., Lee, T. W., & Sejnowski, T. J. (2003). Dictionary learning algorithms for sparse representation. *Neural computation*, *15*(2), 349-396.

[25] Engan, K., Aase, S. O., & Husoy, J. H. (1999, March). Method of optimal directions for frame design. In *1999 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings. ICASSP99 (Cat. No. 99CH36258)* (Vol. 5, pp. 2443-2446). IEEE.

[26] Aharon, M., Elad, M., & Bruckstein, A. (2006). K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on signal processing*, *54*(11), 4311-4322.

[27] Mairal, J., Bach, F., Ponce, J., & Sapiro, G. (2010). Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, *11*(Jan), 19-60.

[28] Chen, X., Du, Z., Li, J., Li, X., & Zhang, H. (2014). Compressed sensing based on dictionary learning for extracting impulse components. *Signal Processing*, *96*, 94-109.

[29] Ding, X., & He, Q. (2016). Time–frequency manifold sparse reconstruction: A novel method for bearing fault feature extraction. *Mechanical Systems and Signal Processing*, *80*, 392–413.

[30] Guo, L., Gao, H., Li, J., Huang, H., & Zhang, X. (2015). Machinery vibration signal denoising based on learned dictionary and sparse representation. In *Journal of Physics: Conference Series* (Vol. 628, No. 1, p. 012124).

[31] Zhou, H., Chen, J., Dong, G., & Wang, R. (2016). Detection and diagnosis of bearing faults using shift-invariant dictionary learning and hidden Markov model. *Mechanical Systems and Signal Processing*, *72*, 65-79.

[32] Tropp, J. A., & Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on information theory*, *53*(12), 4655-4666.

[33] Lorintiu, O., Liebgott, H., Alessandrini, M., Bernard, O., & Friboulet, D. (2015). Compressed sensing reconstruction of 3D ultrasound data using dictionary learning and line-wise subsampling. *IEEE transactions on medical imaging*, *34*(12), 2467-2477.

[34] Huang, Y., Paisley, J., Lin, Q., Ding, X., Fu, X., & Zhang, X. P. (2014). Bayesian nonparametric dictionary learning for compressed sensing MRI. *IEEE Transactions on Image Processing*, *23*(12), 5007-5019.

[35] Ravishankar, S., & Bresler, Y. (2010). MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE transactions on medical imaging*, *30*(5), 1028-1041.

[36] Ravishankar, S., & Bresler, Y. (2013, April). Sparsifying transform learning for compressed sensing MRI. In *2013 IEEE 10th International Symposium on Biomedical Imaging* (pp. 17-20). IEEE.

[37] Zhan, Z., Cai, J. F., Guo, D., Liu, Y., Chen, Z., & Qu, X. (2015). Fast multiclass dictionaries learning with geometrical directions in MRI reconstruction. *IEEE Transactions on biomedical engineering*, *63*(9), 1850-1861.

[38] Song, Y., Yang, G., Xie, H., Zhang, D., & Xingming, S. (2017). Residual domain dictionary learning for compressed sensing video recovery. *Multimedia Tools and Applications*, *76*(7), 10083-10096.

[39] Polania, L. F., & Barner, K. E. (2013, August). Multi-scale dictionary learning for compressive sensing ECG. In *2013 IEEE Digital Signal Processing and Signal Processing Education Meeting (DSP/SPE)* (pp. 36-41). IEEE.

[40] Dong, W., Li, X., Zhang, L., & Shi, G. (2011, June). Sparsity-based image denoising via dictionary learning and structural clustering. In *CVPR 2011* (pp. 457-464). IEEE.

[41] Zhang, X., Feng, X., Wang, W., & Liu, G. (2013). Image denoising via 2D dictionary learning and adaptive hard thresholding. *Pattern Recognition Letters*, *34*(16), 2110-2117.

[42] Zhou, M., Chen, H., Paisley, J., Ren, L., Li, L., Xing, Z., ... & Carin, L. (2011). Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images. *IEEE Transactions on Image Processing*, *21*(1), 130-144.

[43] Duarte-Carvajalino, J. M., & Sapiro, G. (2009). Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization. IEEE Transactions on Image Processing, 18(7), 1395-1408.

[44] Bai, H., Li, G., Li, S., Li, Q., Jiang, Q., & Chang, L. (2015). Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing. *IEEE Transactions on Signal Processing*, *63*(6), 1581-1594.

[45] Ramirez, I., Sprechmann, P., & Sapiro, G. (2010, June). Classification and clustering via dictionary learning with structured incoherence and shared features. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 3501-3508). IEEE.

[46] Kong, S., & Wang, D. (2012, October). A dictionary learning approach for classification: separating the particularity and the commonality. In *European conference on computer vision* (pp. 186-199). Springer, Berlin, Heidelberg.

[47] Zhang, Q., & Li, B. (2010, June). Discriminative K-SVD for dictionary learning in face recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (pp. 2691-2698). IEEE.

[48] Wright, J., Yang, A. Y., Ganesh, A., Sastry, S. S., & Ma, Y. (2008). Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence*, *31*(2), 210-227.

[49] Ptucha, R., & Savakis, A. E. (2014). LGE-KSVD: Robust sparse representation classification. *IEEE Transactions on Image Processing*, *23*(4), 1737-1750.

[50] Jiang, Z., Lin, Z., & Davis, L. S. (2013). Label consistent K-SVD: Learning a discriminative dictionary for recognition. *IEEE transactions on pattern analysis and machine intelligence*, *35*(11), 2651-2664.

[51] Akhtar, N., Shafait, F., & Mian, A. (2016). Discriminative Bayesian dictionary learning for classification. *IEEE transactions on pattern analysis and machine intelligence*, *38*(12), 2374-2388.

[52] Bahrampour, S., Nasrabadi, N. M., Ray, A., & Jenkins, W. K. (2015). Multimodal task-driven dictionary learning for image classification. *IEEE transactions on Image Processing*, *25*(1), 24-38.

[53] Ranieri, J., Chebira, A., & Vetterli, M. (2014). Near-optimal sensor placement for linear inverse problems. *IEEE Transactions on signal processing*, *62*(5), 1135-1146.