

Application of Manifold Learning to Machinery Fault Diagnosis

Jiangping Wang, Tengfei Duan, Lujuan Lei

▶ **To cite this version:**

Jiangping Wang, Tengfei Duan, Lujuan Lei. Application of Manifold Learning to Machinery Fault Diagnosis. 9th International Conference on Intelligent Information Processing (IIP), Nov 2016, Melbourne, VIC, Australia. pp.41-49, 10.1007/978-3-319-48390-0_5 . hal-01614986

HAL Id: hal-01614986

https://inria.hal.science/hal-01614986

Submitted on 11 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Application of Manifold Learning to Machinery Fault Diagnosis

Jiangping Wang, Tengfei Duan, Lujuan Lei

School of Mechanical Engineering,Xi'an Shiyou University
Xi'an,Shanxi 710065,China
jpwang@xsyu.edu.cn
tfduan@163.com

Abstract. The essence of machinery fault diagnosis is pattern recognition. Extracting the fault pattern contained in the vibration signal is the frequently used method to diagnose mechanical fault. Manifold Learning is widely used to extract the non-linear structure within the data and could do the dimensionality reduction of high-dimensional signal. Therefore manifold learning is employed to diagnose the machinery fault. The feature space is constructed by characters in time-frequency domain of vibration signal firstly, and then the manifold learning method named as sparse manifold clustering and embedding is used to extract the essential nonlinear structure of feature space. Afterwards, the fault diagnosis is implemented with spectral clustering and support vector machine. The experiment demonstrates that the approach can effectively diagnose the fault of Machinery.

Keywords: Manifold Learning; Fault Diagnosis; Bearing

1 Introduction

With the increment of precision and operation complexity in industry, the techniques of machinery fault diagnosis, which is used for the safe guarantee in operation of mechanical equipments, has gained more and more attention. At present, data-driven methods play a important role in the mechanical fault diagnosis. The vibration signal contains a wealth of information which indicate the condition of machinery, therefore it is widely used in fault diagnosis that the method based on vibration signal processing [1].The characters of vibration signal in time-frequency domain that is used to establish the initial feature space can be extracted by methods such as Fourier transform or wavelet transform etc. And then, the classifier is constructed with the further processing of initial feature space [1]. The dimensionality reduction method is taken to get a low-dimensional feature representation since the initial feature space is generally high-dimensional. The frequently used dimensional reduction methods, such as principal component analysis [2] or linear discriminant analysis [3], is the linear dimensional reduction method. For the nonlinear signal, the linear method is not the best choice.

Manifold learning has been applied in many fields since three articles published in Science [4], [5], [6]. Manifold learning is a nonlinear data dimensional reduction method, which is used to extract low-dimensional manifold structure embed in high-dimensional space. Therefore there is a new way to diagnose the machinery fault. Generally speaking, the operation data of machinery with the same condition lie on the same manifold and the different condition lie on different manifold [7]. Based on this setting, manifold learning can be applied to machinery fault diagnosis.

A typical manifold learning method includes Local Linear Embedding (LLE) [5], Isometric Mapping(ISOMAP) [6] etc. A proper choice of neighborhood size is critical in manifold Learning. Using a fixed neighborhood size is inappropriate, such as in LLE and ISOMAP, because the curvature of the manifold and the density of the sample points may be different in different regions of the manifold in the engineering applications. Sparse Manifold Clustering and Embedding (SMCE) is proposed in [8]. In SMCE, the automatic choice of the neighborhood size is done by solving a sparse optimization problem. Thus SMCE is more suitable for engineering applications. In this paper, we briefly introduce the principle of SMCE and discuss its application in feature extraction and diagnosis of mechanical fault.

2 The principle of SMCE

SMCE can be used for manifold clustering and dimensionality reduction for multiple nonlinear manifold simultaneously. The key difference between SMCE and other manifold learning methods is that SMCE can automatically search neighbors by solving a sparse optimization problem and weight matrix is established by calculating weights between each points and its neighbors. And then, spectral clustering [9] and Laplacian eigenmap [10] can be used for manifold clustering and dimensional reduction respectively.

Given a set of N samples $\{\boldsymbol{x}_i \in \mathbb{R}^D\}_{i=1}^N$ lying in n different manifolds $\{\mathcal{M}_l\}_{l=1}^n$ with intrinsic dimensions $\{d_l\}_{l=1}^n$. For each sample point $\boldsymbol{x}_i \in \mathcal{M}_l$ consider the smallest ball $\mathcal{B}_i \subset \mathbb{R}^D$ contains that the $d_l + 1$ nearest neighbors of $\boldsymbol{x}_i$. Namely, the affine subspaces with intrinsic dimensions d_l is spanned by the $d_l + 1$ neighbors. This is the fundamental assumption of SMCE. That is,

$$|| \sum_{j \in \mathcal{B}_i} k_{ij}(\boldsymbol{x}_j - \boldsymbol{x}_i)||_2 \leqslant \epsilon \quad and \quad \sum_{j \in \mathcal{B}_i} k_{ij} = 1 \tag{1}$$

Where $\boldsymbol{x}_j \in \mathcal{B}_i$ and $j \neq i$, $\epsilon(\geqslant 0)$ is the upper bound of error, k_{ij} is coefficient. It is hard to know the diameter of $\mathcal{B}_i$, therefore the equation (1) can not be solved directly. The diameter of $\mathcal{B}_i$ is selected according to empirical rules in the LLE and ISOMAP which are local and global manifold learning algorithm respectively. In SMCE, this challenge is to be solved by a sparse optimization problem. Consider a sample point $\boldsymbol{x}_i$ and a sample set $\{\boldsymbol{x}_j | j \neq i, j = 1, 2, \cdots, N\}$, the column vector $\boldsymbol{c}_i$ with dimension $N - 1$ is obtained by solved Eq.(2).

$$||[\boldsymbol{x}_1 - \boldsymbol{x}_i, \cdots, \boldsymbol{x}_N - \boldsymbol{x}_i]\boldsymbol{c}_i||_2 \leqslant \epsilon \quad and \quad \mathbf{1}^T \boldsymbol{c}_i = 1 \tag{2}$$

Where the solution $\boldsymbol{c}_i$ is sparse and the non-zero elements are corresponding to several sample points lying in the same manifold that $\boldsymbol{x}_i$ belongs to. It is the key difference between SMCE and other manifold learning algorithm.

In the case of densely sampled set, the affine subspace coincides with the d_l dimensional tangent space of $\mathcal{M}_l$ at $\boldsymbol{x}_i$. In other words, the sample point corresponding to non-zeros elements of $\boldsymbol{c}_i$ may no longer be the closest points to $\boldsymbol{x}_i$ in $\mathcal{M}_l$. Therefore, the vectors $\{\boldsymbol{x}_j - \boldsymbol{x}_i\}_{j\neq i}$ is normalized and let

$$\boldsymbol{X}_i \coloneqq \left[\frac{\boldsymbol{x}_1 - \boldsymbol{x}_i}{||\boldsymbol{x}_1 - \boldsymbol{x}_i||_2}, \cdots, \frac{\boldsymbol{x}_N - \boldsymbol{x}_i}{||\boldsymbol{x}_N - \boldsymbol{x}_i||_2}\right] \in \mathbb{R}^{D\times(N-1)} \tag{3}$$

Thus the (2) has the following form:

$$||\boldsymbol{X}_i\boldsymbol{c}_i||_2 \leqslant \epsilon \quad and \quad \boldsymbol{1}^T\boldsymbol{c}_i = 1 \tag{4}$$

In this way, among all the solutions of (4), the one that uses a few closest neighbors of $\boldsymbol{x}_i$ is searched by considering the following ℓ^1 optimization problem.

$$min\ ||\boldsymbol{Q}_i\boldsymbol{c}_i||_1 \quad s.t.\ ||\boldsymbol{X}_i\boldsymbol{c}_i||_2 \leqslant \epsilon, \quad \boldsymbol{1}^T\boldsymbol{c}_i = 1 \tag{5}$$

Where the proximity inducing matrix $\boldsymbol{Q}_i$, which is a positive definite diagonal matrix, is defined as

$$\boldsymbol{Q}_i = \frac{||\boldsymbol{X}_j - \boldsymbol{X}_i||_2}{\sum_{t\neq i}||\boldsymbol{X}_t - \boldsymbol{X}_i||_2} \in (0, 1] \tag{6}$$

Another optimization problem which is related to (5) by the method of Lagrange multipliers is

$$min\ \lambda||\boldsymbol{Q}_i\boldsymbol{c}_i||_1 + \frac{1}{2}||\boldsymbol{X}_i\boldsymbol{c}_i||_2^2 \quad s.t. \quad \boldsymbol{1}^T\boldsymbol{c}_i = 1 \tag{7}$$

Where the parameter λ sets the trade-off between the sparsity of the solution and the affine reconstruction error. By solving the optimization problem above for each sample point, it is obtained that the solution $\boldsymbol{c}_i^T \coloneqq [c_{i1}, \cdots, c_{iN}]$. Thus, the weight vector $\boldsymbol{w}_i^T = [w_{i,1}, \cdots, w_{iN}] \in \mathbb{R}^N$ is defined as

$$w_{ii} = 0, \quad w_{ij} \coloneqq \frac{c_{ij}/||\boldsymbol{X}_j - \boldsymbol{X}_i||_2}{\sum_{t\neq i} c_{it}/||\boldsymbol{X}_t - \boldsymbol{X}_i||_2}, \quad j \neq i \tag{8}$$

The similarity graph $G = (V, E)$ whose nodes represent sample points is built. Node i, corresponding to $\boldsymbol{x}_i$, connects the node j, corresponding to $\boldsymbol{x}_j$, with an edge whose weight is equal to $|w_{ij}|$. Each node i connects only a few other nodes named sparse neighbors that correspond to the neighbors of $\boldsymbol{x}_i$ in the same manifold.Hence, the similarity matrix $\boldsymbol{W}$ is constructed with weight vector $\boldsymbol{w}_i$ and reflect the distance from sparse neighbors to $\boldsymbol{x}_i$. The samples are clustered by applying spectral clustering to $\boldsymbol{W}$ [9],or by applying dimensionality reduction for original samples with Laplacian eigenmap [10], and then the classifier is constructed effectively.

3 The method of feature extraction based on SMCE

The mathematical principles of SMCE adapt to engineering applications because the high-dimensional Euclidean space in SMCE is corresponding to the feature space that is spanned by feature vectors of machinery. In practice, the manifolds embedded in feature space have the following characters: the density of sample points in manifold is different in different regions in the manifold, and curvature of each point in manifold is also different. Therefore, it is inappropriate that the fixed neighborhood size is used to extract the manifold structure within mechanical operation data.

The initial feature space should be constructed firstly before that the nonlinear dimensionality reduction is done by manifold learning. In the engineering application, the initial feature space can be spanned by feature vectors that should be composed by characters in time domain or characters in frequency domain, or characters in both time and frequency domain of vibration signal [1]. In this paper, we combine characters in time domain and frequency domain of vibration signals as an initial feature vector. The term of characters in time domain is shown in Tab.1, and the sub-band energy method is employed to establish frequency feature vector.

Table 1. Characters in time domain

k	$\boldsymbol{y}_k$	k	$\boldsymbol{y}_k$	k	$\boldsymbol{y}_k$	k	$\boldsymbol{y}_k$
1	Peak value	4	Root mean square value	7	Kurtosis	10	Impulsion index
2	Peak to peak value	5	Variance	8	Peak index	11	Wave index
3	Absolute mean value	6	Degree of skewness	9	Kurtosis index	12	Tolerance index

The energy distribution in different frequency sub-bands of the vibration signals indicates the operation condition of machinery. Thus the sub-band energy vector could regard as the frequency feature vector. Divide vibration signal into m sub-bands with a constant bandwidth and the frequency range of i-th sub-band is shown in (9).

$$f_i = f_{i-1} + \Delta f \tag{9}$$

Where $i = 1, 2, \cdots, m$, f_i and f_{i-1} is the upper and low limiting frequency of i-th sub-band, Δf is the bandwidth of i-th sub-band. Through Fast Fourier Transform(FFT), the number of frequency lines n_i in each sub-band is equal, that is

$$n_i = \frac{f_i - f_{i-1}}{\xi} \tag{10}$$

Where ξ is frequency resolution. Hence, all frequency lines can be divided into m equal partitions related to m sub-bands. Suppose that the sequence index of frequency lines in i-th sub-band is corresponding to k_{i-1} and k_i, the energy of

i-th sub-band is obtained by (11) according to Parseval's theorem.

$$E_i = \sum_{k=k_{i-1}}^{k_i} |A_k|^2 \Delta f \tag{11}$$

Where A_k is the amplitude of frequency line whose sequence index is k. In this paper, the vibration signal is divided into 16 equal sub-band, thus we can obtain a feature vector $\boldsymbol{y}_f$ of dimension 16. And $\boldsymbol{y}_f$ is normalized by (12). In a similar way, the feature vectors in time domain that is shown in Tab.1 can be normalized, and we obtained a normalized feature vector in time domain Y_t with dimension 12. And then, the feature vectors $Y = [Y_f, Y_t]$ with dimension 28 is used to span the initial feature space.

$$\boldsymbol{Y}_f = \frac{\boldsymbol{y}_f - min(\boldsymbol{y}_f)}{max(\boldsymbol{y}_f) - min(\boldsymbol{y}_f)} \tag{12}$$

The presented method of machinery fault diagnosis based on manifold learning is done with following stage.

(1)For each vibration signal, the feature vector with dimension 28 is constructed firstly.

(2)Using appropriate parameter λ to build similarity matrix W. Due to there is not mature theory to guide us to choose the best parameter λ, therefore the choice of best parameter λ is done by equal step length searching. Namely, the diagnosis accuracy is different under different λ, and we fixed the λ corresponding to best accuracy.

(3) Clustering of the samples data by applying spectral clustering[9] to W. The samples data from the same machinery condition is in one set and from the different condition is in different set. This is the unsupervised fault diagnosis approach. In addition, we apply laplacian eigenmap [10] to do dimensionality reduction and the low-dimensionality representation of original high-dimensionality feature space is obtained. And then, the support vector machine(SVM) is employed to do the classification of the low-dimensionality space. Because of SVM is one of supervised classifier, this approach of fault diagnosis is named supervised method of fault diagnosis.

4 Experiment

Bearing Date Center(BDC) [11] of Case Western Reserve University has been used by many researchers [12], [13]. The availability of BDC has been proved, and this database has become a standard database of vibration signal of bearing defect. In this paper, The deep groove ball bearings SKF 6205-2RS-JEM-SKF with 9 rolling elements made by Swedish is employed as the analysis object. In experiment, the shaft speed is 1750RPM and the sampling frequency is 48KHZ. The vibration signal that collected by a accelerometer is shown in Fig.1. Considering four condition of bearing, that is, normal, inner-race defect, ball defect,

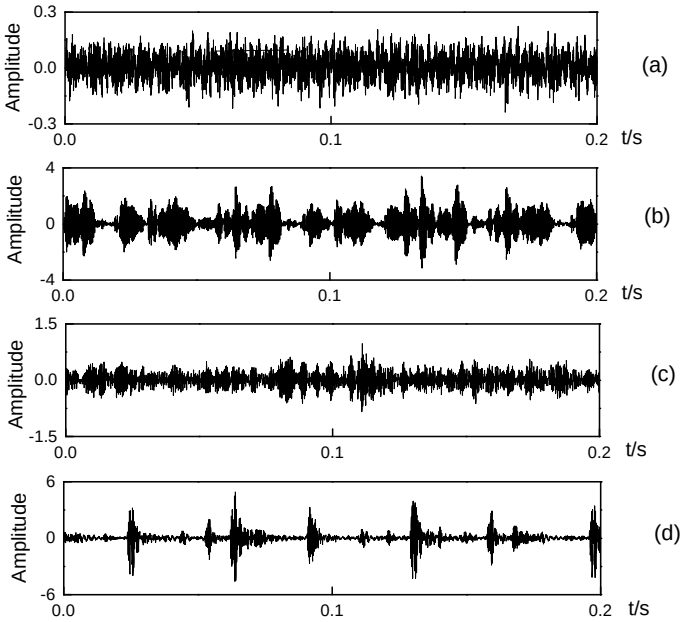


Fig. 1. The original signal of bearing.(a)Normal condition. (b) Inner-race defect. (c) Ball defect. (d)Outer-race defect.

outer-race defect, 200 samples were chosen for each condition. Thus there are 800 samples in all and the sample length is 2048 of all. The feature space is spanned by 800 feature vectors, denoted by $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_{800} \in \mathbb{R}^{28}$, and the manifold structure in the feature space is extracted by the method described previously. Similarity matrix $W \in \mathbb{R}^{800\times800}$ is obtained with fixed parameter λ. And then, the unsupervised fault diagnosis can be achieved by spectral clustering [9]. For the supervised fault diagnosis, Laplacian eigenmap [10] is applied to the W, and original feature space is mapped to three dimensional space. And then, the samples of each condition in three dimensional space is divided into 2 equal partition. One of the half samples are taken as training samples, and the remaining are testing samples. The problem we faced is to classify 4 conditions, but SVM can classify only 2 condition at a time. Generalizing method to solve this challenging is by decomposing the 4 class problem into 2 class problem. Each time take one class of the training samples as positive class and take remaining class as negative class, thus classification of 4 class is done by doing SVM 4 times. The parameter λ play a important role for diagnosis, we make the parameter λ to traverse [10,200] with step 5. Under different λ, the diagnosis accuracy of spectral clustering and support vector machine is shown in Fig.2.

As the Fig.2 show, the best accuracy of diagnosis with spectral clustering is up to 97.125% when the λ is set as 30. In the other way, three dimensional representation of origin feature space with dimension 28 is obtained under different

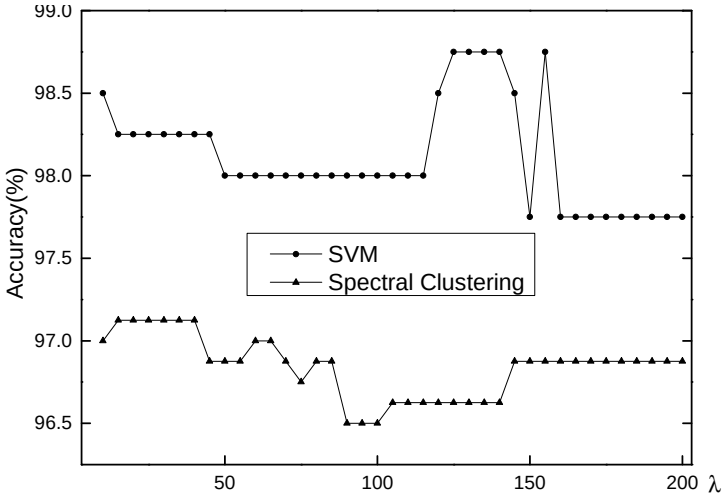


Fig. 2. The diagnosis accuracy with different λ

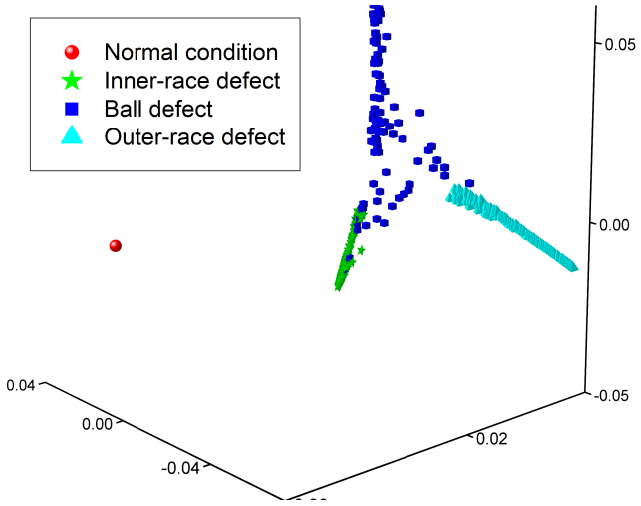


Fig. 3. Three dimensional representation of original feature space

λ. The best accuracy of diagnosis with SVM is up to 98.75% when the λ is set as 130. And the three dimensional space is shown in Fig.3. Spectral clustering do the classification using the geometric structure of space spanned by sample point and without the tag of samples. This is the reason that it is named unsupervised diagnosis. Conversely, SVM do the classification with tags of samples, thus the diagnosis based on SVM is named supervised diagnosis. From Fig.3, the diagnosis accuracy of SVM is higher than spectral clustering and the accuracy of both has remained stable under different λ. The accuracy difference between maximum and minimum is 0.625% and 1% with spectral clustering and SVM respectively. Namely, the parameter λ is easily fixed. It is demonstrated that the presented method is suitable for engineering applications. To reveal the outstanding characters of presented method, the fault diagnosis is done with LLE [5] instead of SMCE. The classifier is established with SVM. The accuracy of diagnosis is shown in Fig.4. It is illustrated that the diagnosis accuracy vary in a large range with different parameter. The comparison of diagnosis accuracy between SMCE and LLE is shown in Tab.2. The diagnosis method with SMCE is more stable than LLE. The satisfactory result is reached with nearly almost arbitrary parameter with SMCE, but the diagnosis result with LLE is more affected by the parameters. Therefore the convenience of parameter choice is one of the advantages of presented method.

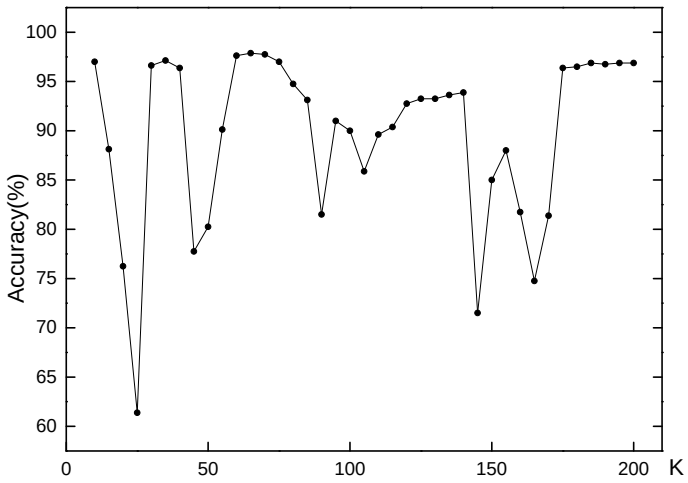


Fig. 4. The diagnosis accuracy with LLE and SVM

Table 2. The comparison of accuracy between SMCE and LLE

Diagnosis method	minimum	maximum	mean	standard variance
SMCE	97.75%	98.75%	98.12%	0.32%
LLE	61.38%	97.88%	89.66%	8.57%

5 Conclusion

(1)The method of feature extraction based on SMCE could extract low dimensional manifold structure that indicates the nature of mechanical condition embed in high-dimensional feature space.
(2)The feature vector that consist of time-domain characters and sub-band energy can be used to diagnose the condition of machinery.
(3)The diagnosis accuracy of presented method in this paper is less affected by the parameters, therefore it is more suitable for engineering applications thanks to the convenience of parameter choice.

References

[1] Jiangping Wang. *The fault diagnosis technology of mechanical equipment and its application*. Northwestern Polytechnical University Press, 2001.
[2] Svante Wold, Kim Esbensen, and Paul Geladi. Principal component analysis. *Chemometrics and intelligent laboratory systems*, 2(1-3):37–52, 1987.
[3] Leo H. Chiang, Mark E. Kotanchek, and Arthur K. Kordon. Fault diagnosis based on fisher discriminant analysis and support vector machines. *Computers and Chemical Engineering*, 28(8):1389–1401, 2004.
[4] H. Sebastian Seung and Daniel D. Lee. The manifold ways of perception. *Science*, 290:2268–2269, 2000.
[5] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–6, 2000.
[6] J. B. Tenenbaum, Silva V De, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–23, 2000.
[7] Quansheng Jiang, Minping Jia, Jianzhong Hu, and Feiyun Xu. Machinery fault diagnosis using supervised manifold learning. *Mechanical Systems and Signal Processing*, 23(7):2301–2311, 2009.
[8] E. Elhamifar and R. Vidal. Sparse manifold clustering and embedding. *Advances in Neural Information Processing Systems*, pages 55–63, 2011.
[9] Andrew Y. Ng, Michael I. Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in Neural Information Processing Systems*, 14:849–856, 2002.
[10] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. *Advances in Neural Information Processing Systems*, 14(6):585–591, 2002.
[11] K.A.Loparo. Bearing vibration data set. `http://csegroups.case.edu/bearingdatacenter/home`.
[12] Lijun Zhang, Jinwu Xu, Jianhong Yang, Debin Yang, and Dadong Wang. Multiscale morphology analysis and its application to fault diagnosis. *Mechanical Systems and Signal Processing*, 22(3):597–610, 2008.

[13] Hongyu Yang, Joseph Mathew, and Lin Ma. Fault diagnosis of rolling element bearings using basis pursuit. *Mechanical Systems and Signal Processing*, 19(2):341–356, 2005.