2020 IEEE INTERNATIONAL WORKSHOP ON MACHINE LEARNING FOR SIGNAL PROCESSING, SEPT. 21–24, 2020, ESPOO, FINLAND

UNSUPERVISED ACOUSTIC CONDITION MONITORING WITH RIEMANNIAN GEOMETRY

Pavel Lifshits *Ronen Talmon*

Andrew and Erna Viterbi Faculty of Electrical Engineering
Technion – Israel Institute of Technology
Technion City, Haifa, Israel

ABSTRACT

In this paper, we present an unsupervised method for acoustic condition monitoring. Our method relies on the Riemannian geometry of symmetric and positive-definite (SPD) matrices. Specifically, SPD matrices enable us to build features for multi-channel data, which naturally encode the mutual relationships between the channels. By exploiting the Riemannian geometry of SPD matrices, we show that these features encompass informative comparisons. The proposed anomaly score is then based on a one-class SVM applied to the proposed features and their induced Riemannian distance. We test the proposed method on two benchmarks and show that it achieves state-of-the-art results. In addition, we demonstrate the robustness of the proposed method to noise and to low sampling rates.

Index Terms— Acoustic scene classification, Anomaly detection in audio, Condition monitoring, Riemannian geometry, SPD matrices, Unsupervised anomaly detection

1. INTRODUCTION

Industrial machinery is prone to failures and breakdowns, resulting in significant expenses, manufacturing delays, and product defects. Hence, there is a rising interest in machine condition monitoring using different sensors, such as vibration, temperature, and pressure [1–3]. As it might be costly to add an extra monitoring layer to production machines, solutions that are non-intrusive and require no changes to the existing infrastructure are favored. Using microphones opens the door to such solutions, since microphones are affordable and easy to deploy, thereby facilitating a modern stethoscope for machines. Indeed, anomaly detection in sound, or acoustic condition monitoring, is an active field of research with many promising studies [4–11].

Accomplishing accurate acoustic condition monitoring is challenging, because (i) the sound produced by machines might be highly non-stationary, (ii) different machines with various properties give rise to a broad variety of sounds, (iii) machines often operate in dynamic environments introducing multiple types of interferences and noise, and (iv) it is impossible to foresee and impractical to account for all possible anomalies, which are by definition rare and unknown a priori. Existing work has been mostly focused on devising hand-crafted acoustic features and models, mainly for single-channel recordings, e.g. [11]. Another approach is to detect anomalies in sound by using technologies for acoustic scene classification and event detection [12]. Recently, following the reemergence of deep learning, a new line of studies based on autoencoders (AEs) has been presented, achieving a significant increment in performance [7–10].

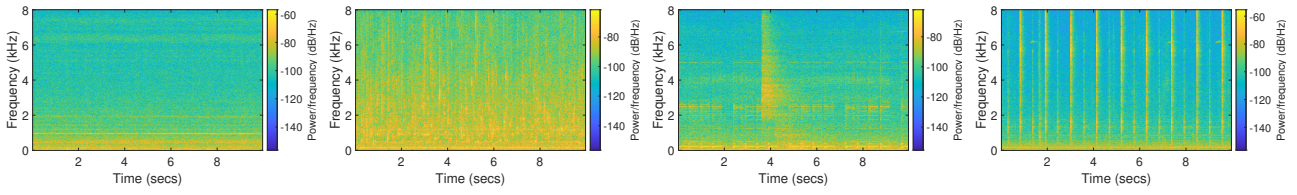
In this paper, we propose an algorithm for unsupervised anomaly detection from multi-channel acoustic recordings. Addressing the problem of *unsupervised* anomaly detection exhibits high flexibility since it does not require any labels. In this setup, only the intrinsic properties of the data are used to detect instances deviating from the majority of the data. Such an approach fits well to acoustic condition monitoring because often, while large amounts are available, the data lack labels. The proposed algorithm is based on sample covariance matrices as features. Covariance matrices have been shown to be powerful features in a broad range of applications, e.g., time series classification [13], trajectory anomaly detection [14], detection of network anomalies [15], brain computer interface [16], and domain adaptation [17], to name but a few. The covariance matrices naturally take into account the mutual-relationships between the channels, explicitly exploiting the availability of multiple channels. In this work, we make particular use of the known Riemannian geometry of covariance matrices [16, 18]. Broadly, the covariance matrices are symmetric and positive-definite (SPD). The collection of SPD matrices forms a Riemannian cone manifold equipped with a distance metric, facilitating informative comparisons between the features. These features and their corresponding Riemannian distance metric are then incorporated in a one-class SVM, giving rise to an unsupervised anomaly score.

We test the proposed algorithm on two benchmarks. The first benchmark was recently introduced by Purohit et al. [19]. This dataset consists of sound recordings of industrial machines under normal and anomalous operating conditions acquired in real factory environments. Koizumi et al. [20] recently introduced the second benchmark, where normal operating sounds of miniature machines (toys) were collected. Anomalous sound was collected by deliberately damaging the machines. We show that the proposed algorithm attains state-of-the-art results on both benchmarks.

2. PROPOSED METHOD

Our method of detecting anomalies comprises several steps. The first step consists of the computation of sample covariances from the data. These sample covariances serve as features, which entail two useful properties. First, the covariance matrix enables us to capture the connectivity between the channels. As we will show in the sequel, such information on the mutual relationships between the channels is critical for anomaly detection. Second, covariance matrices reside in a space with a known and highly useful geometry. At the second step, based on this known geometry, the features (covariance matrices) are embedded in a new Euclidean vector space. Finally, the third step involves the application of a one-class SVM, where the score of

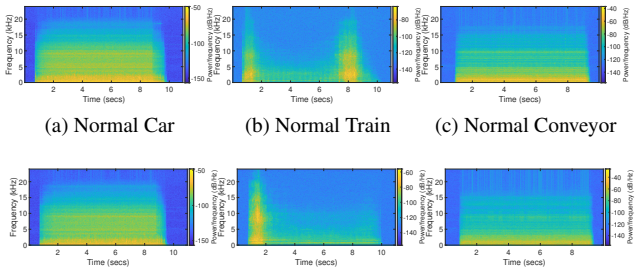
978-1-7281-6662-9/20/$31.00



(a) Normal Fan (b) Normal Pump (c) Normal Slide Rail (d) Normal Valve

(e) Abnormal Fan (f) Abnormal Pump (g) Abnormal Slide Rail (h) Abnormal Valve

Fig. 1. Spectrograms of sessions recorded from the four types of machines with real factory noise of 6 dB SNR, taken from MIMII dataset [19]. The top row illustrates normal operation and the bottom row illustrates abnormal conditions.



(a) Normal Car (b) Normal Train (c) Normal Conveyor

(d) Abnormal Car (e) Abnormal Train (f) Abnormal Conveyor

Fig. 2. Spectrograms of sessions from ToyADMOS dataset [20]. The top row depicts normal operation and the bottom row depicts abnormal conditions.

the trained SVM assigned to newly arrived data points constitutes the anomaly score.

A block diagram of the proposed algorithm is depicted in Fig. 3. The algorithm itself is presented in Algorithm 1, and the different steps are described in more detail in the remainder of this section.

2.1. Covariance matrices as data features

We consider repeated multi-channel recordings in time of a particular system of interest. Let $\mathbf{X}_i \in \mathbb{R}^{n\times k}$ denote the data acquired in the i-th session by n channels at k time samples. Namely, each row of $\mathbf{X}_i$ is a short time series of k samples, and each column of $\mathbf{X}_i$ consists of the signal acquired by n channels at the same time.

We propose to use the sample covariance matrix as the feature of each session. Let $\mathbf{C}_i$ denote the $n \times n$ sample covariance matrix of the i-th session, given by:

$$\mathbf{C}_i = \frac{1}{k-1}\mathbf{X}_i \cdot \mathbf{X}_i^T \tag{1}$$

where $\mathbf{X}_i^T$ is the matrix transpose of $\mathbf{X}_i$. For simplicity, we assume zero mean. Otherwise, the sample mean of the columns of $\mathbf{X}_i$ needs to be subtracted from each column.

The use of covariance matrices gives us several immediate benefits. In the typical case where $k > n$, representing a session by $\mathbf{C}_i \in \mathbb{R}^{n\times n}$ rather than by $\mathbf{X}_i \in \mathbb{R}^{n\times k}$ is more efficient, facilitating a natural dimension reduction. Furthermore, since covariance matrices are computed from data by averaging over time, they tend to be robust to noise. Arguably, these two benefits come at the expense of information loss, and therefore, their contribution could depend on the specific application. Therefore, perhaps the most important benefit stems from the fact that a covariance matrix $\mathbf{C}$ is an SPD matrix, namely, it is symmetric satisfying $\mathbf{C} = \mathbf{C}^T$, and positive-definite $\mathbf{C} \succ 0$, i.e., all its eigenvalues are strictly positive. As we describe in Section 2.2 and Section 2.3, this property gives us a useful interface for processing such features, which extends the naïve treatment of the feature vectors as vectors in a Euclidean (linear) space.

2.2. Riemannian metric and distance of SPD matrices

SPD matrices lie on a cone manifold with a known non-Euclidean geometry [21]. The definition of an SPD matrix entails that the collection of all SPD matrices constitutes a convex half-cone in the vector space of real $n \times n$ symmetric matrices. This cone forms a differentiable Riemannian manifold $\mathcal{M}$, with a Riemannian metric induced by the Euclidean metric of the space of symmetric matrices. Specifically, the inner product is given by

$$\langle \mathbf{S}_1, \mathbf{S}_2 \rangle_{\mathbf{C}} = \langle \mathbf{C}^{-\frac{1}{2}}\mathbf{S}_1\mathbf{C}^{-\frac{1}{2}}, \mathbf{C}^{-\frac{1}{2}}\mathbf{S}_2\mathbf{C}^{-\frac{1}{2}} \rangle \tag{2}$$

where $\mathbf{C}$ is an $n\times n$ SPD matrix, $\mathbf{S}_1$ and $\mathbf{S}_2$ are two $n\times n$ symmetric matrices in the tangent plane to the SPD cone at $\mathbf{C}$, and $\langle \mathbf{S}_1, \mathbf{S}_2 \rangle = \mathrm{Tr}\left(\mathbf{S}_1^T\mathbf{S}_2\right)$ is the standard inner product. Note that the tangent plane to the SPD cone at every point consists of the entire space of $n \times n$ symmetric matrices.

The cone manifold of SPD matrices is a Hadamard manifold, meaning that it is simply connected and it is a complete Riemannian manifold with non-positive sectional curvature. Manifolds with non-positive curvature have a unique geodesic path between any two points. The geodesic path between two SPD matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ is given by

$$\varphi(t) = \mathbf{C}_1^{\frac{1}{2}}(\mathbf{C}_1^{-\frac{1}{2}}\mathbf{C}_2\mathbf{C}_1^{-\frac{1}{2}})^t\mathbf{C}_1^{\frac{1}{2}},\ 0 \le t \le 1 \tag{3}$$

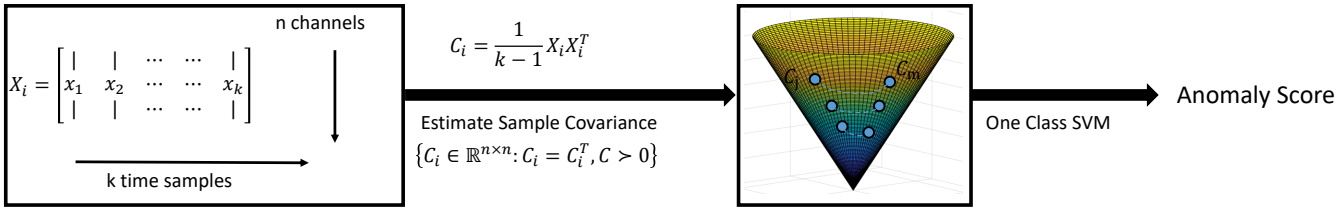


Fig. 3. Block diagram of the proposed anomaly detection method.

The arc length of the geodesic path defines the following Riemannian distance on the manifold

$$d_R^2(\mathbf{C}_1, \mathbf{C}_2) = \left\| \log(\mathbf{C}_2^{-\frac{1}{2}} \mathbf{C}_1 \mathbf{C}_2^{-\frac{1}{2}}) \right\|_F^2 = \sum_{l=1}^{n} \log^2(\lambda_l(\mathbf{C}_2^{-\frac{1}{2}} \mathbf{C}_1 \mathbf{C}_2^{-\frac{1}{2}})) \tag{4}$$

where $\|\cdot\|_F$ is the Frobenius norm, $\log(\mathbf{C})$ is the matrix logarithm, and $\lambda_l(\mathbf{C})$ is the l-th eigenvalue of $\mathbf{C}$.

2.3. Feature embedding in a Euclidean space

Seemingly, the Riemannian metric (2) and distance (4) give us a convenient "interface" for processing the covariance matrices features. One prominent shortcoming is that the Riemannian distance (4) requires the application of eigenvalue decomposition, which could be computationally inefficient in subsequent processing components that make use of pairwise distances, such as SVM. In this subsection, we offer a possible remedy. Let $\overline{\mathbf{C}}$ be the Riemannian mean of the set $\{\mathbf{C}_i\}$, given by the Fréchet mean:

$$\overline{\mathbf{C}} \triangleq \arg\min_{\mathbf{C}} \sum_i d_R^2(\mathbf{C}, \mathbf{C}_i) \tag{5}$$

Consider two SPD matrices $\mathbf{C}_i$ and $\mathbf{C}_j$. Let $\mathbf{S}_i$ and $\mathbf{S}_j$ be the projections of $\mathbf{C}_i$ and $\mathbf{C}_j$, respectively, to the linear tangent space to the SPD manifold at $\overline{\mathbf{C}}$, which are given explicitly by the following logarithmic map

$$\mathbf{S}_i = \overline{\mathbf{C}}^{-\frac{1}{2}} \log(\mathbf{C}_i) \overline{\mathbf{C}}^{-\frac{1}{2}}, \ \mathbf{S}_j = \overline{\mathbf{C}}^{-\frac{1}{2}} \log(\mathbf{C}_j) \overline{\mathbf{C}}^{-\frac{1}{2}} \tag{6}$$

Then, the Riemannian distance on the SPD cone in the neighborhood of $\overline{\mathbf{C}}$ is approximated by [22]:

$$d_R^2(\mathbf{C}_i, \mathbf{C}_j) \approx \|\mathbf{S}_i - \mathbf{S}_j\|_F^2 \tag{7}$$

Since the Frobenious norm on the right-hand side is given by

$$\|\mathbf{S}_i - \mathbf{S}_j\|_F = \|\mathbf{s}_i - \mathbf{s}_j\|_2 \tag{8}$$

where $\mathbf{s}_i = \text{vec}(\mathbf{S}_i)$ and $\mathbf{s}_j = \text{vec}(\mathbf{S}_j)$ are column stack vector representation of an upper triangle with an appropriate normalization scale. Consequently, the logarithmic map in (6) can be seen as an embedding of the covariance matrices features into a Euclidean vector space.

When the pairwise distances are required, the advantage of the embedding is that eigenvalue decomposition of each SPD matrix needs to be computed only once in (6). However, if the matrices do not reside in a local region but rather are spread over a large region of the manifold, then the Riemannian distance d_R in (4) can be used at the expense of additional computational cost.

2.4. Proposed anomaly score

Here, we use a one-class SVM [23] with an RBF kernel for unsupervised anomaly detection [24]. The main idea is to train a one-class SVM based on training data without labels, so that the resulting support vectors separate data from the origin in the high-dimensional kernel space. Analogously to binary classification, the objective is to separate normal data from abnormal data by minimizing:

$$0.5 \sum_{i,j=1}^{m} \alpha_i \alpha_j G(\mathbf{s}_i, \mathbf{s}_j), \text{ s.t.} \sum_{i=1}^{m} \alpha_i = m\nu, \ 0 \le \alpha_i \le 1$$

where m is the number of sessions, $G(\mathbf{s}_i, \mathbf{s}_j)$ is the Gram matrix of an RBF kernel given by $e^{-\gamma \|\mathbf{s}_i - \mathbf{s}_j\|^2}$, and ν is a hyperparameter.

A small value of ν leads to fewer support vectors, and therefore, to a smooth, crude decision boundary. A large value of ν leads to more support vectors, and therefore, to a curvy, flexible decision boundary. The optimal value of ν should be large enough to capture the data complexity and small enough to avoid overtraining. In this case, we set ν to a default value of 0.5, which was empirically shown to attain good results.

In our case, the data points are the vector representation $\mathbf{s}_i$ of the covariance matrices $\mathbf{C}_i$ of the multi-channel sessions $\mathbf{X}_i$, and the kernel space of the SVM is defined on the Euclidean embedded space of $\mathbf{s}_i$, as described in Section 2.3. Here, we use the approximation of the Riemannian distance, given in (8), in the RBF kernel rather than the Riemannian distance d_R for computational efficiency. Indeed, as we report in the sequel, this implementation yields satisfactory results. We remark that using an RBF kernel with the exact Riemannian distance is possible, at the expense of additional computational effort.

Finally, once the SVM is trained, the anomaly score of a new incoming data point $\mathbf{X}_i$ is defined as the distance of its respective feature $\mathbf{s}_i$ from the trained separating hyperplane.

The entire algorithm for computing the anomaly score is summarized in Algorithm 1.

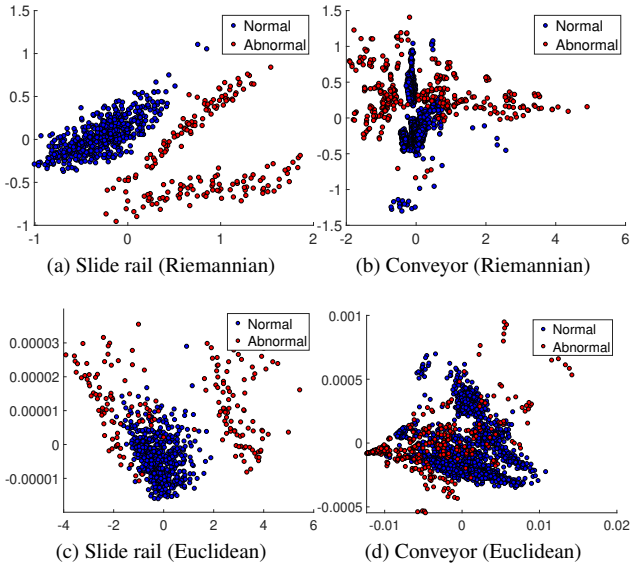
Algorithm 1 Computing the Anomaly Score

Training Stage:

1: Input: multi-channel recordings of sessions $\{\mathbf{X}_i\}$
2: Calculate the sample covariance matrices of the sessions $\{\mathbf{C}_i\}$
3: Compute the embedded features $\{\mathbf{s}_i\}$
4: Train one-class SVM classifier with RBF kernel on $\{\mathbf{s}_i\}$

Anomaly Detection Stage:

1: Input: multi-channel recordings of a new session $\mathbf{X}_j$
2: Calculate the sample covariance matrix of the session $\mathbf{C}_j$
3: Compute the embedded features $\mathbf{s}_j$
4: Compute the distance of $\mathbf{s}_j$ from the trained separating hyperplane



(a) Slide rail (Riemannian) (b) Conveyor (Riemannian)

(c) Slide rail (Euclidean) (d) Conveyor (Euclidean)

Fig. 4. 2D MDS of the features of a slide rail (MIMII) and a conveyor (ToyADMOS). Each point represents a session. Top row using Riemannian distances, and bottom row using Euclidean distances.

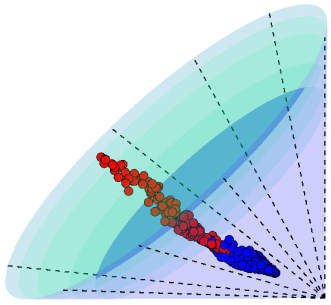


Fig. 5. The reduced 2×2 covariance matrices of sessions recorded from a slide rail (MIMII dataset) presented in $\mathbb{R}^3$ with respect to the cone SPD manifold. Each point represents one session. Normal sessions are colored in blue and abnormal sessions are colored in red. The dotted lines mark the boundaries of the SPD cone manifold.

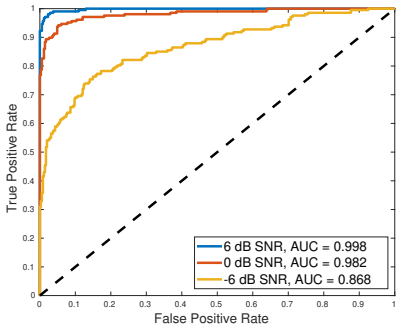


Fig. 6. ROC curve obtained for fan recordings (MIMII) under three noise levels.

3. EXPERIMENTAL RESULTS

3.1. Tested datasets

The MIMII dataset [19] contains the sound of four types of machines: (i) solenoid *valves* that are repeatedly opened and closed, (ii) water *pumps* that drain water from a pool and discharge water to the pool continuously, (iii) industrial *fans*, which provide a continuous flow of gas or air in factories, and (iv) linear *slide rails*, which consist of a moving platform and a stage base. For each type of machine, different real-life anomalous scenarios are considered, including contamination, leakage, rotating unbalance, and rail damage.

The dataset was collected by a TAMAGO-03 microphone array, which is a circular microphone array consisting of eight distinct microphones. The sounds of the machines in the different scenarios were recorded in separate sessions. The sound of the machine was sampled at 16 kHz in a reverberant environment. Apart from the target machine sound, background noise from several factories was recorded and later added to the target machine sound for mimicking real environments. The same microphone array was used for recording the target machine sound and the background noise. The entire dataset contains a total of $26,092$ sessions of normal operation and $6,065$ sessions of operation during anomalous conditions. Examples of normal and abnormal sessions recorded from the four types of machines are depicted in Fig. 1.

The ToyADMOS dataset [20] contains normal operating sounds of miniature machines. Since miniature machines can be installed in an acoustic laboratory, the recording conditions were controlled. The anomalous operating sounds were acquired by deliberately damaging the components of the miniature machines. The dataset contains three tasks: product inspection (toy car), fault diagnosis for a fixed machine (toy conveyor), and fault diagnosis for a moving machine (toy train). The toy car anomaly conditions include bent shaft, deformed or melted gears, coiled tires (with plastic or steel ribbon) and over/under voltage. The toy conveyor anomalies include excessive tension of the tension or tail pulley, missing tail pulley, attached metallic object to the belt, and over/under voltage. The toy train anomalies are chipped wheel axle of the first or last carriage, broken/disjointed/obstructed straight or curved railway track.

The sounds were recorded by four omnidirectional microphones (SHURE SM11-CN) at a 48 kHz sampling rate. Overall, the dataset contains $4,500$ sessions of normal sound and 875 sessions of abnormal sound. Fig. 2 depicts examples of sessions from the three tasks, demonstrating normal and abnormal operation.

3.2. Performance evaluation

We apply Algorithm 1 to the recordings from the two datasets described above. First, to demonstrate the anomaly detection capabilities of the covariance matrices as features, and particularly, the use of their Riemannian geometry, we show in Fig. 4 the 2D embedding of the sessions obtained by multi-dimensional scaling (MDS) [25] equipped with the Euclidean and Riemannian distances. Each point in the plots is a single session, where normal sessions are colored in blue and abnormal sessions are colored in red. We observe that in all the cases, the distances between the covariance matrices of the sessions exhibit a significant distinction between normal and abnormal operation. In addition, we also observe the clear benefit of using the Riemannian distance compared to the Euclidean distance.

To give further intuition to the Riemannian geometry of the SPD matrices, we use the dimension reduction method proposed in [26] to reduce the dimension of the 8×8 covariance matrices to 2×2. Any 2×2 symmetric matrix $\mathbf{C} = \left(\begin{smallmatrix} x & y \\ y & z \end{smallmatrix}\right)$ can be displayed in $\mathbb{R}^3$ by

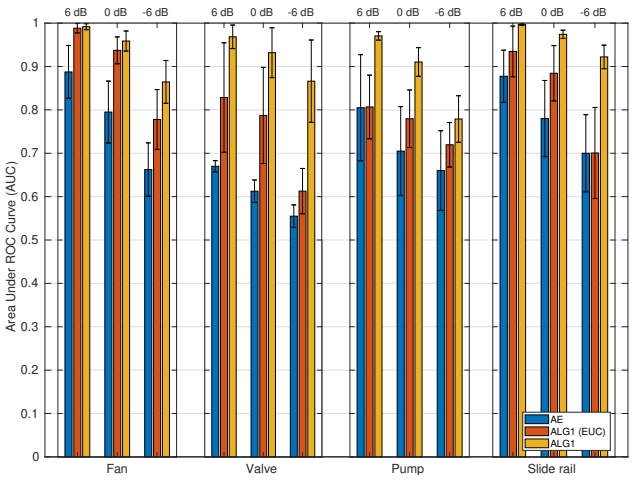


Fig. 7. Detection results (AUC) on the MIMII dataset.

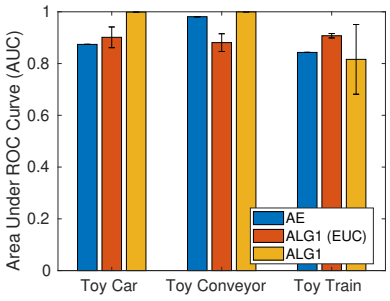


Fig. 8. Detection results (AUC) on the ToyADMOS dataset.

(x, y, z). Furthermore, the manifold of 2×2 SPD matrices can be displayed as a cone in $\mathbb{R}^3$, since $\mathbf{C}$ is positive-definite if and only if $x > 0, z > 0$ and $y^2 < xz$. Fig. 5 presents the reduced 2×2 covariance matrices in $\mathbb{R}^3$ of sessions from the MIMII dataset. Indeed, in this geometry, despite the very low dimension, the separation between normal and abnormal operations is evident.

To objectively evaluate the performance of the unsupervised anomaly detection, we follow common practice and rank the results according to the anomaly score. Then, we iteratively apply a varying threshold, where every threshold value results in a tuple of true positives and false positives. Collecting these tuples from all the thresholds forms a single receiver operator characteristic (ROC) curve. The area under the curve (AUC) is used as a detection performance measure. Figure 6 presents an example for such a ROC curve obtained by Algorithm 1 applied to sessions from the MIMII dataset recorded from a fan with different levels of noise.

We compare the results obtained by Algorithm 1 to two competitors. The first competitor is a variant of Algorithm 1 based on the covariance matrices as features, where the Riemannian distance (approximation) is replaced by the standard Euclidean distance. The second competitor is based on AE. For the MIMII dataset, the implementation is described in [19] based on the algorithm proposed in [9, 10]. For the ToyADMOS dataset, the implementation details appear in [20] based on [27].

In Fig. 7 we present the results for the MIMII dataset obtained by 50-fold cross-validation. We observe that the proposed method

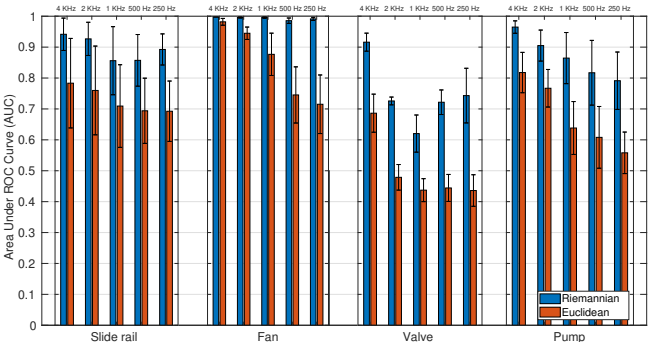


Fig. 9. Detection results (AUC) on the MIMII dataset for different sampling rates (6 dB SNR).

(ALG1) outperforms the competitor (AE), demonstrating state-of-the-art performance in all tested cases including all the machines and different noise levels. The results also imply that the covariance matrix is a powerful feature in this task. In addition, the benefit of using the Riemannian geometry is clearly evident.

Importantly, we remark that the performance obtained by the proposed method on the MIMII dataset exceeds the performance of the variational deep neural networks presented in [7], which was published few days before the submission date of the present work.

Fig. 8 presents the results for the ToyADMOS dataset obtained by 50-fold cross-validation. We observe that the proposed method outperforms the competitor for Toy Car sessions, obtains comparable results for Toy Conveyor sessions, and is inferior for Toy Train sessions. As implied by the large standard deviation, we remark that the relatively low performance for Toy Train sessions is due to poor detection of an abnormal operation of one case, whereas in the other three cases, the proposed method obtains comparable or better results compared to the competitor.

Kawaguchi et al. [9] suggested to lower the cost of acoustic monitoring by reducing the sampling rate. Following [28], we test the robustness of our method to the sampling rate. To this end, we resample the signals with an appropriate anti-aliasing lowpass filter. In Fig. 9 we present the detection results. We observe that using the Euclidean distance the performance degrades fast with the decrease in the sampling rate. Conversely, using the Riemannian distance leads to robustness, and for most machines even 1 kHz sampling rate still attains satisfactory performance.

4. CONCLUSIONS

In this paper, we proposed an unsupervised method for detecting anomalous sounds of industrial machines, facilitating affordable, easy to deploy, and accurate condition monitoring. While the vast majority of recent work involves deep learning solutions, our approach is based on the Riemannian geometry of SPD matrices. We demonstrate that a learning-based solution that makes explicit use of fundamental theoretical ideas, such as sensor connectivity and Riemannian geometry, may be advantageous over learning-only solutions in particular cases. Indeed, in the task at hand, where the typical datasets are relatively small and the data exhibit distinct structures, we show that our method achieves state-of-the-art results on two benchmarks. Future work will focus on the extension of this approach to positive kernels rather than covariance matrices, which are SPD as well.

Acknowledgements

This research was supported by the Technion Hiroshi Fujiwara cyber security research center and by the Pazy Foundation.

5. REFERENCES

[1] E. P. Carden and P. Fanning, "Vibration based condition monitoring: a review," *Structural Health Monitoring*, vol. 3, no. 4, pp. 355–377, 2004.

[2] G. Lodewijks, W. Li, Y. Pang, and X. Jiang, "An application of the IoT in belt conveyor systems," *International Conference on Internet and Distributed Computing Systems*, pp. 340–351, 2016.

[3] R.F. Salikhov, Y.P. Makushev, G.N. Musagitova, L.U. Volkova, and R.S. Suleymanov, "Diagnosis of fuel equipment of diesel engines in oil-and-gas machinery and facilities," in *AIP Conference Proceedings*, 2019, vol. 2141, p. 050009.

[4] Y. Koizumi, S. Murata, N. Harada, S. Saito, and H. Uematsu, "Sniper: Few-shot learning for anomaly detection to minimize false-negative rate with ensured true-positive rate," *Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP)*, pp. 915–919, 2019.

[5] Y. Kawachi, Y. Koizumi, S. Murata, and N. Harada, "A two-class hyper-spherical autoencoder for supervised anomaly detection," *Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP)*, pp. 3047–3051, 2019.

[6] Y. Kawaguchi, R. Tanabe, T. Endo, K. Ichige, and K. Hamada, "Anomaly detection based on an ensemble of dereverberation and anomalous sound extraction," *Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP)*, pp. 865–869, 2019.

[7] K. Suefusa, T. Nishida, H. Purohit, R. Tanabe, T. Endo, and Y. Kawaguchi, "Anomalous sound detection based on interpolation deep neural network," *Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP)*, pp. 271–275, 2020.

[8] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi, and N. Harada, "Unsupervised detection of anomalous sound based on deep learning and the neyman–pearson lemma," *IEEE/ACM Trans Audio Speech Lang Process*, vol. 27, no. 1, pp. 212–224, 2018.

[9] Y. Kawaguchi and T. Endo, "How can we detect anomalies from subsampled audio signals?," *IEEE Int Workshop Mach Learn Signal Process (MLSP)*, pp. 1–6, 2017.

[10] E. Marchi, F. Vesperini, F. Eyben, S. Squartini, and B. Schuller, "A novel approach for automatic acoustic novelty detection using a denoising autoencoder with bidirectional LSTM neural networks," *Proc IEEE Int Conf Acoust Speech Signal Process (ICASSP)*, pp. 1996–2000, 2015.

[11] P. Foggia, N. Petkov, A. Saggese, N. Strisciuglio, and M. Vento, "Audio surveillance of roads: A system for detecting anomalous sounds," *IEEE trans Intell Transp Syst*, vol. 17, no. 1, pp. 279–288, 2015.

[12] D. Barchiesi, D. Giannoulis, D. Stowell, and M. D. Plumbley, "Acoustic scene classification: Classifying environments from the sounds they produce," *IEEE Signal Process Mag*, vol. 32, no. 3, pp. 16–34, 2015.

[13] H. Ergezer and K. Leblebicioğlu, "Time series classification with feature covariance matrices," *Knowl Inf Syst*, vol. 55, no. 3, pp. 695–718, 2018.

[14] H. Ergezer and K. Leblebicioğlu, "Anomaly detection and activity perception using covariance descriptor for trajectories," *Comput Vis ECCV*, pp. 728–742, 2016.

[15] S. Jin, D. S. Yeung, X. Wang, and E. C. Tsang, "A covariance matrix based approach to internet anomaly detection," in *Advances in Machine Learning and Cybernetics*, pp. 691–700. Springer, 2006.

[16] A. Barachant, S. Bonnet, M. Congedo, and C. Jutten, "Classification of covariance matrices using a Riemannian-based kernel for BCI applications," *Neurocomputing*, vol. 112, pp. 172–178, 2013.

[17] O. Yair, M. Ben-Chen, and R. Talmon, "Parallel transport on the cone manifold of SPD matrices for domain adaptation," *IEEE Trans Signal Process*, vol. 67, no. 7, pp. 1797–1811, 2019.

[18] X. Pennec, P. Fillard, and N. Ayache, "A Riemannian framework for tensor computing," *Int J Comput Vis*, vol. 66, no. 1, pp. 41–66, 2006.

[19] H. Purohit, R. Tanabe, T. Ichige, T. Endo, Y. Nikaido, K. Suefusa, and Y. Kawaguchi, "MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection," *Proceedings of the Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019)*, pp. 209–213, 2019.

[20] Y. Koizumi, S. Saito, H. Uematsu, N. Harada, and K. Imoto, "ToyADMOS: A dataset of miniature-machine operating sounds for anomalous sound detection," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 313–317, 2019.

[21] R. Bhatia, *Positive Definite Matrices*, vol. 24, Princeton University press, 2009.

[22] O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans Pattern Anal Mach Intell*, vol. 30, no. 10, pp. 1713–1727, 2008.

[23] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[24] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing one-class support vector machines for unsupervised anomaly detection," *Proceedings of the ACM SIGKDD Workshop on Outlier Detection and Description*, pp. 8–15, 2013.

[25] M. A. A. Cox and T. F. Cox, "Multidimensional scaling," in *Handbook of Data Visualization*, pp. 315–347. Springer, 2008.

[26] M. T. Harandi, M. Salzmann, and R. Hartley, "From manifold to manifold: Geometry-aware dimensionality reduction for SPD matrices," *Comput Vis ECCV*, pp. 17–32, 2014.

[27] Y. Koizumi, S. Saito, M. Yamaguchi, S. Murata, and N. Harada, "Batch Uniformization for Minimizing Maximum Anomaly Score of DNN-based Anomaly Detection in Sounds," *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 6–10, 2019.

[28] Y. Kawaguchi, "Anomaly detection based on feature reconstruction from subsampled audio signals," *Proc Eur Signal Process Conf EUSIPCO*, pp. 2524–2528, 2018.