

# CLASSIFICATION OF BIOACOUSTIC SIGNALS WITH TANGENT SINGULAR SPECTRUM ANALYSIS

Lincon S. Souza<sup>\*</sup>, Bernardo B. Gatto<sup>†</sup>, Kazuhiro Fukui<sup>\*†</sup>

lincons@cvlab.cs.tsukuba.ac.jp, bernard.gatto@gmail.com, kfukui@cs.tsukuba.ac.jp

<sup>\*</sup>University of Tsukuba, Japan

<sup>†</sup> Center for Artificial Intelligence Research (C-AIR), Japan

## ABSTRACT

Automatic classification of bioacoustic signals is an essential tool in biology for laborious tasks such as environmental monitoring in areas of difficult access. A working system applied in the field must be able to run on small scale machines and make reasonable predictions from a small sample of data. Recently, a method called Grassmann singular spectrum analysis (GSSA) was introduced as the latest development in a line of research where bioacoustic signals are represented by subspaces. While this paradigm is compact and introduces a straightforward discriminant analysis for classification, it is based on a Grassmann kernel, which approximates the Grassmann manifold by a reproducing Hilbert kernel space, thus depending on a choice of a dictionary and not being able to capture the signals complexity from a small class sample. In this paper, we propose a method named tangent singular spectrum analysis (TSSA), which continues to exploit the advantages of subspace representation but does not rely on approximating the Grassmann manifold by a low-dimensional kernel. We formulate a discriminant analysis on a tangent space to the data sample mean, using the extrinsic coordinates of the manifold. The validity of TSSA is demonstrated through experiments on the Amazon rainforest Anuran dataset.

**Index Terms**— Bioacoustic signals, tangent singular spectrum analysis, Grassmann manifold.

## 1. INTRODUCTION

Bioacoustic signal classification has become an indispensable tool in biology, by providing experts with the means to efficiently collect information for monitoring of ecological communities, and searching clues for understanding the cladistics of animals from the perspective of similitude of their bioacoustic mechanisms. Monitoring species remotely and in real-time allows the gathering of critical information not only to predict changes in the populations, but observing correlated phenomena, e.g. pollution of water, air, soil and climate changes.

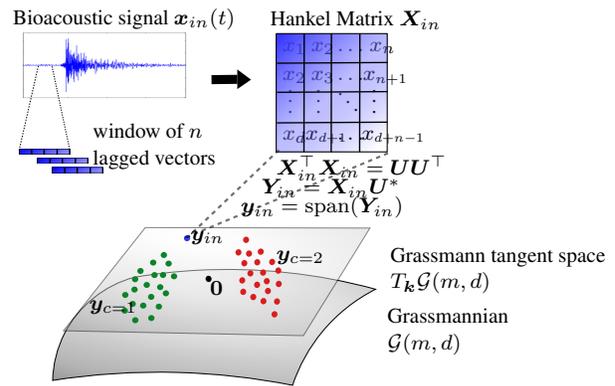


Fig. 1. Conceptual diagram of the proposed TSSA.

A classification algorithm for bioacoustics needs to have low computational cost during its acquisition, preprocessing and classification stages. This constraint is necessary to perform monitoring through small-scale devices, deployable on the environment. Likewise, such an algorithm needs to produce reliable predictions from a small dictionary of data, mainly for two reasons: first, gathering and labelling data for an initial dictionary can be expensive and time-consuming for biologists. Second, some species are extremely rare or under risk of extinction, and only a few samples will be available in a practical timeframe.

Various of the recently proposed bioacoustic signal classification frameworks [1, 2] are comprised of multiples elements, including noise [3] and dimensionality reduction [4], feature selection [5], model training [6] and classification [7].

Motivated by the constraint on computational cost, mutual singular spectrum analysis (MSSA) [8] was introduced as a novel bioacoustic signal representation based on subspaces. Its model is compact and requires no cost-intensive preprocessing techniques such as segmentation, noise reduction or syllable extraction. The input signals are represented within the singular spectrum analysis (SSA) paradigm, by a set of leftmost eigenvectors extracted based on the signals' accumulated energy. This set forms a basis for a subspace in the vector space, since the vectors are orthonormal [9, 10].

Then, canonical angles [11, 12] between the different bioacoustic subspaces are used to measure their similarity, and 1-NN classification is performed.

Later, Grassmann singular spectrum analysis (GSSA) [13] has been proposed as an extension of MSSA, addressing the issue that it had no discriminant mechanism to separate classes (e.g. species of animals). They introduce a formulation based on the fact the set of subspaces form a Riemannian manifold called Grassmannian. Inspired by the popularity of discriminant analysis on a Grassmannian (GDA) in the field of image set classification [14, 15, 16, 17], the paper introduces this paradigm for the classification of bioacoustic signals. Each subspace, and thereby, each bioacoustic signal is represented by a point on the Grassmann manifold.

However, a caveat of GSSA is that it conducts kernel discriminant analysis through the kernel trick with a Grassmann kernel [18]. It does not perform discriminant analysis on extrinsic matrix variables, but instead it creates a low-dimensional parameterisation of the Grassmannian as a reproducing Hilbert kernel space. The problem with this premise in the setting of bioacoustic signal classification is that we have a constraint on available data as discussed above, and the Grassmann kernel's parameterisation is intrinsically dependent on a choice of dictionary. Using such a kernel with a small dictionary may impair the classification ability, as the kernel may not be able to capture the signals' complexity.

To tackle the problem of bioacoustic signal classification with a constraint on the number of available samples, we formulate a new algorithm of discriminant analysis for subspaces, temporarily called tangent singular spectrum analysis (TSSA), which is depicted as a conceptual illustration in the Figure 1. Instead of relying on a kernel parameterisation of the Grassmannian, we utilise the extrinsic matrix structure of the manifold; the subspaces are mapped through the logarithmic map onto a tangent space at the sample mean of the data, called Karcher mean; then discriminant analysis in matrix space is performed.

We summarise our contributions as follows: (1) We propose a discriminant bioacoustics classification algorithm called TSSA which does not approximate the Grassmann manifold using a kernel parameterisation. (2) We continue to exploit subspaces advantages to bioacoustic systems, such as a low storage requirements; integrated filtering and simple selection of compactness ratio; robustness to arbitrary signal lengths; and self-containment, not requiring preprocessing techniques, such as segmentation, or syllable extraction.

## 2. PROPOSED METHOD

### 2.1. Review of MSSA

Consider  $N_c$  reference bioacoustic signals  $\mathbf{x}_c^i(t)$  for each  $c$ -th class ( $c = 1, \dots, C$ ), where  $i = 1, \dots, N_c$ , and  $N_c$  can be different for each class. The index  $t = 1, \dots, L_c^i$  indicates

the ordering of the signal in time up to its length  $L_c^i$ .

We consider only the variant MSSA-II from [13]. It assumes that each signal  $\mathbf{x}_c^i(t)$  can be represented by a linear mapping in terms of its autocorrelation, different from the original MSSA that extended this assumption to a whole class of signals. The reason is that the original assumption was too strong and did not provide the necessary flexibility for a class with non-gaussian noise, also not scaling well with an increasing number of signals per class. There is no cost penalty to consider the second variant, which we refer to as simply MSSA.

Each signal in MSSA is represented by a subspace  $\mathbf{y}_c^i$ ; thus a class of signals is effectively represented by a set of subspaces  $\{\mathbf{y}_c\}$ . The dimension of subspaces  $m$ , corresponding to the signal's compactness ratio, is selected empirically during the training. For a given unknown input bioacoustic signal  $\mathbf{x}_{in}(t)$  of length  $L_{in}$ , the task is to compute a subspace  $\mathbf{y}_{in}$  and predict its corresponding bioacoustic class (e.g. species) based on the nearest subspace.

Now we review the representation by a SSA subspace; for that, consider a single signal  $\mathbf{x}(t)$ . First, we apply a sliding window over  $\mathbf{x}(t)$  to turn the 1-dimensional signal of length  $L$  in a sequence of  $n$  lagged vectors of length  $d$  arranged in a Hankel matrix  $\mathbf{H} \in \mathbb{R}^{d \times n}$ , i.e. the vectors form blocks of a matrix  $\mathbf{H}_{i,j} = \mathbf{H}_{i+j-2}$ , where the number of columns  $n$  is the number of desired oscillatory components (or principal components), which has a direct correspondence with the maximum delay of the autocorrelation. Therefore, the total number of columns in matrix  $\mathbf{H}$  is given by relationship  $n = L - d + 1$ . Note that the anti-diagonal elements in  $\mathbf{H}$  are equal.

The singular value decomposition (SVD) of the autocorrelation matrix  $\mathbf{H}^\top \mathbf{H}$  is performed to find the directions of maximum variance of the Hankel matrix  $\mathbf{H}$ ; that is, for this case:  $\mathbf{U} \mathbf{\Lambda} \mathbf{U}^\top = \mathbf{H}^\top \mathbf{H}$ , where the columns of  $\mathbf{U} \in \mathbb{R}^{n \times n}$  corresponds to the eigenvectors of  $\mathbf{H}^\top \mathbf{H}$  and the diagonal of  $\mathbf{\Lambda}$  corresponds to the singular values  $\lambda_j$  ( $j = 1, \dots, n$ ), i.e.  $\text{diag}(\mathbf{\Lambda}) = \lambda_1, \lambda_2, \dots, \lambda_n$  in decreasing order. Given the compactness ratio  $m$ , the  $m$ -leftmost matrix  $\mathbf{U}^* \in \mathbb{R}^{n \times m}$  from  $\mathbf{U}$  is selected, containing the first  $m$  eigenvectors corresponding to the highest eigenvalues. The Hankel matrix  $\mathbf{H}$  is projected onto the subspace spanned by  $\mathbf{U}^*$  to obtain  $\mathbf{X} = \mathbf{H} \mathbf{U}^* \in \mathbb{R}^{d \times m}$ , which is the best  $m$ -rank approximation of  $\mathbf{H}$  and spans the subspace  $\mathbf{y}$  that characterizes our signal. In our classification framework, we utilize the previous procedure to compute basis matrices  $\mathbf{X}_c^i \in \mathbb{R}^{d \times m}$  for a SSA subspace  $\mathbf{y}_c^i$  corresponding to each reference signal  $\mathbf{x}_c^i(t)$ .

### 2.2. Grassmann Manifold

Now, we introduce here the concept of Grassmann manifold and operations we can perform on it. First, the Grassmann manifold  $\mathcal{G}(m, d)$  is defined as the set of  $m$ -dimensional linear subspaces of  $\mathbb{R}^d$ . It is an  $m(d - m)$ -dimensional compact

manifold and can be written as a quotient space of orthogonal groups  $\mathcal{G}(m, d) = \mathcal{O}(d)/\mathcal{O}(m) \times \mathcal{O}(d - m)$ , where  $\mathcal{O}(m)$  is the group of  $m \times m$  orthonormal matrices. A subspace can be represented an orthogonal basis matrix, and the set of tall orthogonal matrices comprises the compact Stiefel manifold  $\mathcal{S}(m, d)$ .

A tangent space  $T_{\mathbf{y}}\mathcal{G}$  at  $\mathbf{y} \in \mathcal{G}(m, d)$  can be seen intuitively as a subspace of  $\mathbb{R}^{d \times m}$ , since  $T_{\mathbf{y}}\mathcal{G}$  can be derived as a horizontal space. Given a point  $\mathbf{W} \in \mathcal{S}(m, d)$ , such that  $\text{span}(\mathbf{W}) = \mathbf{y}$ , a horizontal space  $H$  is a subspace of the tangent space  $T_{\mathbf{W}}\mathcal{S}$  (a copy of  $\mathbb{R}^{d \times m}$ ) which contains the directions of infinitesimal variation of  $\mathbf{W}$  that modify its span.

The Grassmann manifold is a Riemannian manifold, when endowed with a Riemannian metric  $g$ . As a quotient manifold of an embedded submanifold of  $\mathbb{R}^{d \times m}$ , the Euclidean metric induces the Grassmannian canonical metric  $g : \mathbf{U}_1 \times \mathbf{U}_2 \rightarrow \text{tr} \mathbf{U}_1^T \mathbf{U}_2$ ,  $\mathbf{U}_1, \mathbf{U}_2 \in T_{\mathbf{y}}\mathcal{G}$ . The metric  $g$  can be intuitively seen as the restriction of the Euclidean metric to a Grassmann tangent space (a horizontal space  $H$ ).

### 2.3. Exponential map

The exponential map  $\text{Exp} : \mathcal{G} \times T\mathcal{G} \times \mathbb{R} \rightarrow \mathcal{G}$  can be used to calculate an specific point on a geodesic  $\gamma(t)$ , given a point  $\gamma(0)$ , a direction  $\dot{\gamma}(t)$  and a length  $t$ . It is denoted by  $\gamma(t) = \text{Exp}_{\mathbf{y}} \mathbf{u}$ , meaning the point  $\gamma(t)$  in the geodesic emanating from  $\mathbf{y} = \gamma(0)$  in the direction of  $\mathbf{u} = t \frac{\dot{\gamma}(0)}{\|\dot{\gamma}(0)\|}$ . In this paper, we utilise the extrinsic function [19]  $S(t) : t \in [-\epsilon, \epsilon] \rightarrow \mathcal{S}(m, d)$  for a geodesic  $\gamma(t)$  on the Grassmann manifold, where  $\gamma(t) = \text{span}(S(t))$ . Consider  $S(0) = \mathbf{S}$  and  $\dot{S}(0) = \mathbf{U}$ , then the equation follows:  $S(t) = \text{orth}(\mathbf{S}\mathbf{Q}(\cos \Sigma t)\mathbf{Q}^T + \mathbf{J}(\sin \Sigma t)\mathbf{Q}^T)$ , where  $\mathbf{J}\Sigma\mathbf{Q}^T = \mathbf{U}$  is the compact singular value decomposition of  $\mathbf{U}$ . Here,  $\mathbf{J}, \mathbf{S}, S(t) \in \mathcal{S}(m, d)$  and  $\mathbf{Q} \in \mathcal{O}(m)$  are orthonormal matrices, and  $\Sigma \in \text{diag}(m)$  is a diagonal matrix.

### 2.4. Logarithmic map

The inverse of the exponential map is the logarithmic map (or log map)  $\text{Log} : \mathcal{G} \times \mathcal{G} \rightarrow T\mathcal{G}$ , denoted by  $\mathbf{u} = \text{Log}_{\mathbf{x}} \mathbf{y}$ . Given two points on the manifold  $\mathbf{x}$  and  $\mathbf{y}$ , one wants to find the tangent vector  $\mathbf{u}$  at  $\mathbf{x}$  pointing towards  $\mathbf{y}$ . Note that  $\text{Log}_{\mathbf{x}} \mathbf{y} \neq \text{Log}_{\mathbf{y}} \mathbf{x}$ . In this work, following [19], given two basis matrices  $\mathbf{X}$  and  $\mathbf{Y}$  for  $\mathbf{x}$  and  $\mathbf{y}$ , we calculate the log map by solving the SVD  $((\mathbf{Y}^T \mathbf{X})^{-1}(\mathbf{Y}^T - \mathbf{Y}^T \mathbf{X} \mathbf{X}^T))^T = \mathbf{W}\Theta\mathbf{Z}^T$ , and then reconstructing the tangent vector by  $\mathbf{u} = \mathbf{W}^* \arctan(\Theta^*)\mathbf{Z}^{*T}$ , where  $\mathbf{W}^*, \Theta^*$  and  $\mathbf{Z}^*$  are the  $m$ -leftmost matrices of  $\mathbf{W}, \Theta$  and  $\mathbf{Z}$  respectively.

### 2.5. Sample mean and variance in the Grassmann manifold

Since a manifold is not necessarily a vector space, additive mean may not be valid. Instead, inspired by the approach to

statistics on manifolds of previous works [20, 21], the Karcher mean [22] is utilised, which can be computed through algorithm 1. Likewise, sample variance of the data is defined as the expected value of the squared Riemannian distance from the mean, i.e.  $E[\|\text{Log}_{\mathbf{k}} \mathbf{x}\|^2]$ , where  $\mathbf{k}$  is the Karcher mean and  $\mathbf{x}$  is a random point;

### 2.6. Orthogonal projection in matrix space

We use the notation  $P_{\mathbf{A}_j}(\mathbf{X}) : \mathbf{X} \rightarrow \sum_{j=1}^n \mathbf{A}_j \text{tr} \mathbf{A}_j^T \mathbf{X}$  to denote the projection of  $\mathbf{X}$  onto a  $n$ -dimensional subspace spanned by the matrix space basis  $\{\mathbf{A}_j\}, (j = 1, \dots, n)$ . Note that this does not refer to subspaces of the Grassmannian of data, but refers to projection onto subspaces of the tangent space.

---

#### Algorithm 1: Computation of the Karcher mean

---

**input** : subspaces  $\{\mathbf{x}_i\} \in \mathcal{G}$ , where  $i = 1, \dots, N$ ; a step  $\tau$ ; and a patience  $\epsilon$ .

$\mathbf{k}_0 = \mathbf{x}_1$  // 1: initialise the mean with a sample

**do**

$\Delta \mathbf{k} = \frac{\tau}{N} \sum_{i=1}^N \text{Log}_{\mathbf{k}_j} \mathbf{x}_i$  // 2: log-map points to  $T_{\mathbf{k}_j} \mathcal{M}$

$\mathbf{k}_{j+1} = \text{Exp}_{\mathbf{k}_j} \Delta \mathbf{k}$  // 3: exp-map velocities to the manifold

**while**  $\|\mathbf{k}\| > \epsilon$  // 4: stop when the candidate become small

**output**:  $\mathbf{k} \in \mathcal{G}$ , the Karcher mean

---

### 2.7. Tangent Discriminant Analysis

The algorithm of discriminant analysis (LDA) [23] can be realised for matrix space as follows. Let  $\mathbf{X}_c^i \in \mathbb{R}^{d \times m}$  be the basis matrix for a SSA subspace  $\mathbf{y}_c^i$  corresponding to each reference signal  $\mathbf{x}_c^i(t)$ . Each class  $c$  has  $N_c$  number of samples. We calculate the Karcher mean  $\mathbf{k}$  of all subspaces through algorithm 1. Let  $\bar{\mathbf{X}}_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \text{Log}_{\mathbf{k}} \mathbf{X}_c^i$  be the sample mean of class  $c$ . Note that the overall mean is  $\bar{\mathbf{X}} = \frac{1}{N} \sum_{c=1}^C \sum_{i=1}^{N_c} \text{Log}_{\mathbf{k}} \mathbf{X}_c^i = \text{Log}_{\mathbf{k}} \mathbf{k} = 0$ . In the vector discriminant analysis one seeks the discriminant direction  $\mathbf{W}$  which maximizes the Rayleigh quotient  $R(\mathbf{w}) = \mathbf{w}^T \mathbf{B} \mathbf{w} / \mathbf{w}^T \mathbf{S} \mathbf{w}$  where  $\mathbf{B}$  and  $\mathbf{S}$  are the between-class and within-class covariance matrices respectively. This vector-valued function can be rewritten for matrices as:

$$R(\mathbf{W}) = \frac{\text{tr} \mathbf{W}^T \mathbf{P}_{\mathbf{B}}(\mathbf{W})}{\text{tr} \mathbf{W}^T \mathbf{P}_{\mathbf{S}}(\mathbf{W})}, \quad (1)$$

where  $\mathbf{B}, \mathbf{S} \in \mathbb{R}^{dm \times dm}$  are the between-class and within-class covariance matrices. The class covariance matrices are defined as:  $\mathbf{B}_c = N_c(\text{vec} \bar{\mathbf{X}}_c \otimes \text{vec} \bar{\mathbf{X}}_c^T)$ , and  $\mathbf{S}_c = \sum_{i=1}^{N_c} \text{vec}(\text{Log}_{\mathbf{k}} \mathbf{X}_c^i - \bar{\mathbf{X}}_c) \otimes \text{vec}(\text{Log}_{\mathbf{k}} \mathbf{X}_c^i - \bar{\mathbf{X}}_c)^T$ .

Then, the regularised class covariance matrices are given by:  $\hat{\mathbf{B}}_c(\delta) = (1 - \delta)\mathbf{B}_c + \delta \sum_{c=1}^C \mathbf{B}_c$ ,  $\hat{\mathbf{S}}_c(\delta) = (1 - \delta)\mathbf{S}_c + \delta \sum_{c=1}^C \mathbf{S}_c$  and  $\tilde{\mathbf{S}}_c(\delta, \gamma) = (1 - \gamma)\hat{\mathbf{S}}_c(\delta) + \gamma \text{tr}(\hat{\mathbf{S}}_c(\delta))\mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix, the parameter  $\delta$  controls the degree of shrinkage of the individual class covariance matrix estimates towards the pooled estimate, and  $\gamma$  controls shrinkage towards a multiple of the identity matrix, in order to alleviate the problem of matrix degeneration when taking the inverse. Both parameters can be set in a range between 0 and 1. Finally, we have that  $\mathbf{B} = \frac{1}{N} \sum_{c=1}^C \hat{\mathbf{B}}_c(\delta)$  and  $\mathbf{S} = \frac{1}{N} \sum_{c=1}^C \tilde{\mathbf{S}}_c(\delta, \gamma)$ .

The optimal  $\mathbf{W}$  is obtained from reshaping the largest eigenvectors of  $\mathbf{S}^{-1}\mathbf{B}$ . Since  $\mathbf{S}^{-1}\mathbf{B}$  has rank  $C - 1$ , there are  $C - 1$  optima  $\mathbf{W}_1, \dots, \mathbf{W}_{C-1}$ . Data can be projected onto the discriminant space spanned by  $\{\mathbf{W}_j\}$  by  $P_{\mathbf{W}_j}(\text{Log}_k \mathbf{X}_c^i)$ . This corresponds to feature extraction of data onto the most discriminant subspace.

We apply this projection operation to the reference subspace matrices  $\mathbf{X}_c^i$  to generate reference vectors  $\mathbf{V}_c^i = P_{\mathbf{W}_j}(\text{Log}_k \mathbf{X}_c^i)$ . When given an unknown bioacoustic signal  $\mathbf{x}_{in}(t)$ , we compute its SSA subspace basis  $\mathbf{X}_{in}$  and map it onto the discriminant tangent space to generate a vector  $\mathbf{V}_{in}$ ; then we predict its corresponding bioacoustic class (e.g. species) based on the nearest reference vector (1-NN) using the canonical metric  $\text{tr} \mathbf{V}_{in}^T \mathbf{V}_c^i$ .

### 3. EXPERIMENTAL RESULTS

In this section, we evaluate the validity of the proposed TSSA through an experiment using the Anuran records dataset, that was also used in [13]. It consists of 60 bioacoustic signals with different duration recorded in Amazon rainforest containing anuran's croaks and ribbits and various real background noises from the surrounding nature. Anura is the name of an order of animals in the class Amphibia that includes frogs and toads, so the diversity of signal is complex. The classes of this dataset consists of 10 species of anurans.

We evaluate TSSA against its conventional counterparts, MSSA and GSSA. We performed a 10-fold cross-validation by dividing the signals randomly, with 30 for training and 30 for test, and always ensuring that at least one signal of each anuran species is present in both groups. The parameters were varied in the following manner: the number of lagged vectors  $n$  was varied from 10 to 50 and the dimension of SSA subspaces were varied from 3 to 9.

The experimental results can be seen in Table 1. The accuracy refers to the average among all folds, and the standard deviation is calculated from the folds assuming a Gaussian distribution. TSSA outperforms the state-of-art methods: GSSA by approximately 4.5% and MSSA by about 13.12%. This result is a compelling evidence that the discriminant mechanism of the proposed method is able to enhance the performance of bioacoustic classification in this subspace representation paradigm.

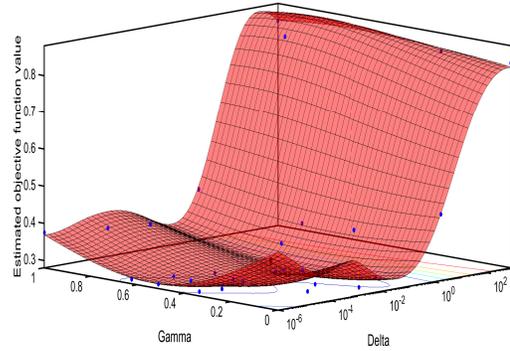


Fig. 2. Plot of error as a function of the shrinkage parameters.

We have also investigated the variation of performance due to changes on the shrinkage parameters  $\gamma$  and  $\delta$ . The subspace dimension was fixed at  $m = 4$  and the length of the lagged vectors was fixed at  $d = 34$ , both of which showed good results at the previous experiment. The plot on Fig. 2 shows the error rate variation when these parameter changed. The blue dots indicate observed results, while the surface is a graphic completion for visualisation. From the plot, we can see that in general a low value of  $\delta$  yields better results, which is closer to a quadratic discriminant model. In addition  $\gamma$  should not be too high or too low, and results around 0.4 yielded the best results.

Table 1. Results for the experiment with the Anuran data. The accuracy refers to the average among all folds.

Methods	Average (%)	Std. Dev.
MSSA	62.88	4.08
GSSA	71.67	3.60
TSSA	76.00	4.39

### 4. CONCLUSIONS

In this paper we have proposed tangent singular spectrum analysis (TSSA), an extension to MSSA, to address more effectively the classification of bioacoustic signals. The key idea of our proposed method is to introduce a discriminatory mechanism for class separation, utilising the extrinsic matrix structure of the manifold instead of relying on a kernel parameterisation of the Grassmannian. The subspaces are mapped through the logarithmic map onto a tangent space at the sample mean of the data. Then discriminant analysis in matrix space is performed. Our method also inherits various advantages of MSSA, such as low storage, consistent compactness ratio selection, signal length free formulation, and no need preprocessing techniques. The validity of TSSA was demonstrated through a classification experiment with the Anuran data where it outperformed the state-of-the-art methods MSSA and TSSA.

## 5. REFERENCES

- [1] Ana Alves, Ricardo Antunes, Anna Bird, Peter L Tyack, Patrick JO Miller, Frans-Peter A Lam, and Petter H Kvadsheim, "Vocal matching of naval sonar signals by long-finned pilot whales (*globicephala melas*)," *Marine Mammal Science*, vol. 30, no. 3, pp. 1248–1257, 2014.
- [2] Juan Gabriel Colonna, Marco Cristo, Mario Salvatierra, and Eduardo Freire Nakamura, "An incremental technique for real-time bioacoustic signal segmentation," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7367–7374, 2015.
- [3] Jesús B Alonso, Josué Cabrera, Rohit Shyamnani, Carlos M Travieso, Federico Bolaños, Adrián García, Alexander Villegas, and Mark Wainwright, "Automatic anuran identification using noise removal and audio activity detection," *Expert Systems with Applications*, vol. 72, pp. 83–92, 2017.
- [4] JF Ruiz-Muñoz, Zeyu You, Raviv Raich, and Xiaoli Z Fern, "Dictionary learning for bioacoustics monitoring with applications to species classification," *Journal of Signal Processing Systems*, pp. 1–15, 2016.
- [5] Farook Sattar, Sarika Cullis-Suzuki, and Feng Jin, "Identification of fish vocalizations from ocean acoustic data," *Applied Acoustics*, vol. 110, pp. 248–255, 2016.
- [6] Ilyas Potamitis, Stavros Ntalampiras, Olaf Jahn, and Klaus Riede, "Automatic bird sound detection in long real-field recordings: Applications and tools," *Applied Acoustics*, vol. 80, pp. 1–9, 2014.
- [7] Mohammad Pourhomayoun, Peter Dugan, Marian Popescu, and Christopher Clark, "Bioacoustic signal classification based on continuous region processing, grid masking and artificial neural network," *arXiv preprint arXiv:1305.3635*, 2013.
- [8] Bernardo B Gatto, Juan G Colonna, Eulanda M dos Santos, and Eduardo F Nakamura, "Mutual singular spectrum analysis for bioacoustics classification," in *Machine Learning for Signal Processing (MLSP), 2017 IEEE 27th International Workshop on*. IEEE, 2017, pp. 1–6.
- [9] Kazuhiro Fukui and Atsuto Maki, "Difference subspace and its generalization for subspace-based methods," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 11, pp. 2164–2177, 2015.
- [10] Bernardo Bentes Gatto, Lincon S. Souza, and Eulanda M dos Santos, "A deep network model based on subspaces: A novel approach for image classification," in *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*. IEEE, 2017, pp. 436–439.
- [11] H. Hotelling, "Relation between two sets of variables," *Biometrika*, vol. 28, pp. 322–377, 1936.
- [12] SN Afriat, "Orthogonal and oblique projectors and the characteristics of pairs of vector spaces," *Proc. Cambridge Philos. Soc.*, vol. 53, pp. 800–816, 1957.
- [13] Lincon S. Souza, Bernardo B. Gatto, and Kazuhiro Fukui, "Grassmann singular spectrum analysis for bioacoustics classification," in *2018 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. IEEE, 2018.
- [14] Jihun Hamm and Daniel D Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 376–383.
- [15] Pavan Turaga, Ashok Veeraraghavan, Anuj Srivastava, and Rama Chellappa, "Statistical computations on grassmann and stiefel manifolds for image and video-based recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2273–2286, 2011.
- [16] Lincon Souza, Hideitsu Hino, and Kazuhiro Fukui, "3d object recognition with enhanced grassmann discriminant analysis," in *ACCV 2016 Workshop (HIS 2016)*, 2016.
- [17] Lincon S. Souza, Bernardo Bentes Gatto, and Kazuhiro Fukui, "Enhancing discriminability of randomized time warping for motion recognition," in *Machine Vision Applications (MVA), 2017 Fifteenth IAPR International Conference on*. IEEE, 2017, pp. 77–80.
- [18] Jihun Hamm, "Subspace-based learning with grassmann kernels," 2008.
- [19] P-A Absil, Robert Mahony, and Rodolphe Sepulchre, "Riemannian geometry of grassmann manifolds with a view on algorithmic computation," *Acta Applicandae Mathematica*, vol. 80, no. 2, pp. 199–220, 2004.
- [20] P Thomas Fletcher, Conglin Lu, Stephen M Pizer, and Sarang Joshi, "Principal geodesic analysis for the study of nonlinear statistics of shape," *IEEE transactions on medical imaging*, vol. 23, no. 8, pp. 995–1005, 2004.
- [21] Pavan Turaga, Ashok Veeraraghavan, Anuj Srivastava, and Rama Chellappa, "Statistical computations on grassmann and stiefel manifolds for image and video-based recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2273–2286, 2011.
- [22] Hermann Karcher, "Riemannian center of mass and mollifier smoothing," *Communications on pure and applied mathematics*, vol. 30, no. 5, pp. 509–541, 1977.
- [23] Reinosuke Fukunaga, "Statistical pattern recognition," 1990.