

Bearing Fault Diagnosis by EXIN CCA

G. Cirrincione, H. Henao
LTI-EESA Lab
University of Picardy Jules Verne
Amiens, France
exin@u-picardie.fr

M. Delgado, J.A. Ortega
MCIA Center
Technical University of Catalonia (UPC)
Terrassa, Catalonia, Spain
miguel.delgado@upc.edu

Abstract—EXIN CCA is an extension of the Curvilinear Component Analysis (CCA), which solves for the noninvariant CCA projection and allows representing data drawn under different operating conditions. It can be applied to data visualization, interpretation (as a kind of sensor of the underlying physical phenomenon) and classification for real time industrial applications. Here an example is given for bearing fault diagnostics in an electromechanical device.

Keywords: bearing fault, classification, curvilinear component analysis, least squares, intrinsic dimension, multilayer perceptron, principal component analysis, visualization.

I. INTRODUCTION

The curvilinear component analysis (CCA, [1] [2] [3]) is a self-organizing neural network which performs the quantization of a data training set (TS) for estimating the corresponding nonlinear projection from the data space to a lower dimensional space (latent space). In the recall phase, by means of the same training algorithm, it outputs the projection of a data input. CCA is a better version of Sammon mapping [1], because the CCA error (*stress*) function uses weights depending on distances in latent space. There are different versions of CCA according to its weighting factor. The EXIN extension of CCA considers the CCA algorithm weighted by a decreasing exponential, whose corresponding error function is the *right Bregman divergence* [4]. Indeed, this function penalizes inconsistent long distances and its asymmetry allows a better unfolding of data. As a consequence, it is well suited for classification, in which distances between clusters have to be well represented. The CCA projection is not invariant. Indeed, it changes at each estimation, because it is only constrained by the preservation of distance topology: there are two sources of randomness, namely the initial conditions in the latent space and the index sequence of the fixed points in the stochastic gradient algorithm. This variability is a problem in case of classification. In [3] constraints are added in order to let the axis of maximum variance be horizontal : however, it is only a partial solution to the invariance problem, because this principal axis depends on the TS and corresponding working conditions under which it has been drawn. It will be shown that it cannot be used in classification problems with TS's drawn under different conditions, because a unique stable framework is needed.

The next section explains the invariant EXIN CCA neural network. As an example, an application to the classification of

bearing defects in an electromechanical device is proposed, which also needs a particular hierarchical multilayer perceptron.

II. THE EXIN CCA NEURAL NETWORK

The EXIN version of CCA extends the dimensionality reduction by taking into account the varying environment: non stationarity of the unknown data distribution which may depend on time or other observable parameters (both called here *working conditions* and regrouped in a vector \mathbf{p}). For each choice of \mathbf{p} , a data TS is drawn and the corresponding CCA projection is computed: all CCA's are collected in a *pool*. If it is required to project an input data drawn under new working conditions, a new recall phase is needed, which depends on the CCA pool.

Define as $\mathbf{X}_i^l \in \mathfrak{R}^m$ the i -th data point drawn from a TS (named l) collected under working conditions represented by \mathbf{p}_l . The projection is given by $\mathbf{Y}_i^l \in \mathfrak{R}^s$, $s \leq m$. The Euclidean distances between points i and j are called D_{ij} in data space and L_{ij} in latent space. In the n th dimensional parameter space, represented in fig. 1, each CCA of the pool is represented by the corresponding parameter vector. EXIN CCA requires a weighting factor which takes into account the new working conditions \mathbf{p} w.r.t. the conditions \mathbf{p}_l of each TS of the CCA pool. This factor, inspired by the Nadaraya-Watson distance [5] for smoothening the interpolation, uses Gaussian pdf's (see fig. 1) centered at the associated working conditions and is given by:

$$w_l(\mathbf{p}) = \frac{e^{-(\mathbf{p}-\mathbf{p}_l)^T \Sigma^{-1}(\mathbf{p}-\mathbf{p}_l)}}{\sum_{j=1}^k e^{-(\mathbf{p}-\mathbf{p}_j)^T \Sigma^{-1}(\mathbf{p}-\mathbf{p}_j)}} \quad (1)$$

where k yields the problem-dependent number of neighbor CCA's whose working conditions are close to \mathbf{p} (in fig. 1 four neighbors in a square grid are considered, as in the application of the next section). The covariance matrix Σ has to take into account the resolution of the interpolation and depends on the physical phenomenon represented by \mathbf{p} . Choosing k means deciding how many projections are used in the recall phase. However, when the working operation point is too far, the associated projection is given less importance by decreasing the weight factor.

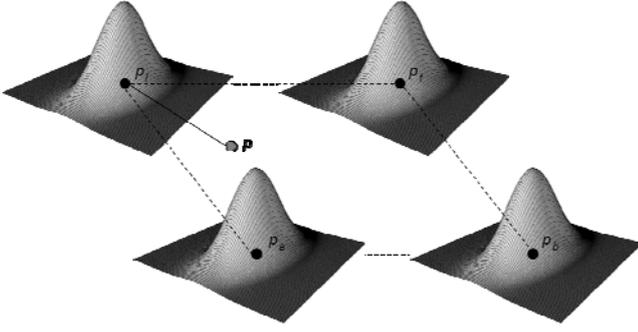


Fig. 1: Gaussians for the weight factor: case of four neighbors in a square grid.

The EXIN CCA right Bregman divergence [4] is given by:

$$E_{Bregman}(\mathbf{Y}^p) = \lambda^2 \sum_i \sum_j w_{l(i)}(p(j)) \left[e^{-\frac{D_{ij}}{\lambda}} - e^{-\frac{L_{ij}}{\lambda}} + (D_{ij} - L_{ij}) \frac{e^{-\frac{L_{ij}}{\lambda}}}{\lambda} \right] \quad (2)$$

being $l(i)$ and $p(j)$ the indexes of the working conditions under which data i and j have been drawn, respectively. The stochastic gradient algorithm for minimizing (2) is then:

$$\mathbf{Y}_j^p \leftarrow \mathbf{Y}_j^p - \alpha w_l(p) \frac{L_{ij} - D_{ij}}{L_{ij}} e^{-\frac{L_{ij}}{\lambda}} (\mathbf{Y}_j^p - \mathbf{Y}_i^l) \quad (3)$$

where $i \neq j$ and α (learning rate) and λ are scalars. This formula assumes that only the projected data from the k nearest neighbor CCA's are to be considered in training and recall. Observe that the EXIN weight acts by multiplying (modulating) the learning rate of the CCA gradient algorithm.

In order to achieve the CCA invariance, three strategies are applied (in case of classification, this is repeated for each class):

1. The initial conditions of the points in latent space are chosen from the projection points of the closest (w.r.t. the working conditions) CCA. If the former points are less than the latter points, the choice is done by using repetition (with replacement).
2. The initial learning rate α is lowered in order to have fewer variations in the projection (then, it is exponentially decreased).
3. A refinement (affine) transformation is applied to all CCA projection points in order to refer each CCA projection map to a unique one, here called *rated CCA*.

For the estimation of the affine transformation, consider that for each data point \mathbf{X}_i^p in the TS p , the rated CCA and the p th CCA output two different s -dimensional projected points \mathbf{Y}_i^{rated} and \mathbf{Y}_i^p . In the stabilization of the CCA training, the entire TS p (containing m data points) is projected and then transformed to the rated projection, by means of $\mathbf{Y}_i^{rated} = \mathbf{A}\mathbf{Y}_i^p + \mathbf{b}$ (if matrix \mathbf{A} is orthogonal, the transformation is a rotation, if not it is affine) for each i , which

is computed by solving the over determined system of linear equations given in matrix form by:

$$\begin{pmatrix} \mathbf{I}_s \otimes \mathbf{Y}_1^{pT} & \mathbf{I}_s \\ \vdots & \vdots \\ \mathbf{I}_s \otimes \mathbf{Y}_m^{pT} & \mathbf{I}_s \end{pmatrix} \boldsymbol{\theta} = \begin{pmatrix} \mathbf{Y}_1^{rated} \\ \vdots \\ \mathbf{Y}_m^{rated} \end{pmatrix} \quad (4)$$

where the data matrix contains the Kronecker product \otimes and $\boldsymbol{\theta} = \left((\text{vec } \mathbf{A})^T \quad \mathbf{b}^T \right)^T$ is the affine parameter vector. The

system can be solved by ordinary least squares (OLS), even if the noisy structured data matrix suggests the use of the constrained least squares, CTLS [6], which looks for the minimum noise subspace. Indeed, only ms values (the m \mathbf{Y}_i^p latent points) in the $ms \times s(m+1)$ data matrix in (4) are noisy. In the recall phase, instead, the parameters of the tuning affine transformation are estimated by interpolation (e.g., using a Radial Basis Function neural network, RBF [5], as in the example of the paper). This approach is possible because all CCA frameworks are coherent, and then an interpolation of the affine transformations is possible. The final projection, unlike [3], which is data dependent, is invariant, in the sense that it respects the reference axis of the rated CCA projection.

III. APPLICATION TO BEARING FAULT DIAGNOSIS

Early diagnosis of faults in induction machines is an extensively investigated field for cost and maintenance savings. In fact, induction motors are still the most important rotating electrical machines in industry.

Most electrical machines use either ball or rolling-element bearings which consist of outer and inner rings. Balls or rolling elements rotate in raceways inside the rings. Bearings are the major source of faults. Depending on the type and size of the machine, these faults vary from about 40% to about 90% from large to small machines. Bearing defects under normal operational conditions often occur because of material fatigue. They have been categorized as distributed or local [7]. Distributed defects include mainly surface roughness, waviness and misaligned races. Localized defects include cracks, pits and spalls, basically. Localized bearing faults are classified by the fault specific location in: inner race, outer race and ball faults mainly, although some studies include also cage faults. Generalized roughness, instead, is a type of fault where the condition of a bearing surface has degraded considerably over a large area and become rough, irregular, or deformed. A common example is the overall surface roughness produced by a contamination or loss of lubricant.

Although different physical magnitudes such as stator currents or acoustic emissions have been included to develop bearing faults diagnosis methodologies, the vibration analysis is the most popular in practice [8]. Most of the bearings monitoring schemes are focused on bearing localized defects, because they allow an enough reliable frequency analysis. On the contrary, the generalized roughness faults, for example, produce unpredictable broadband effects in the machine

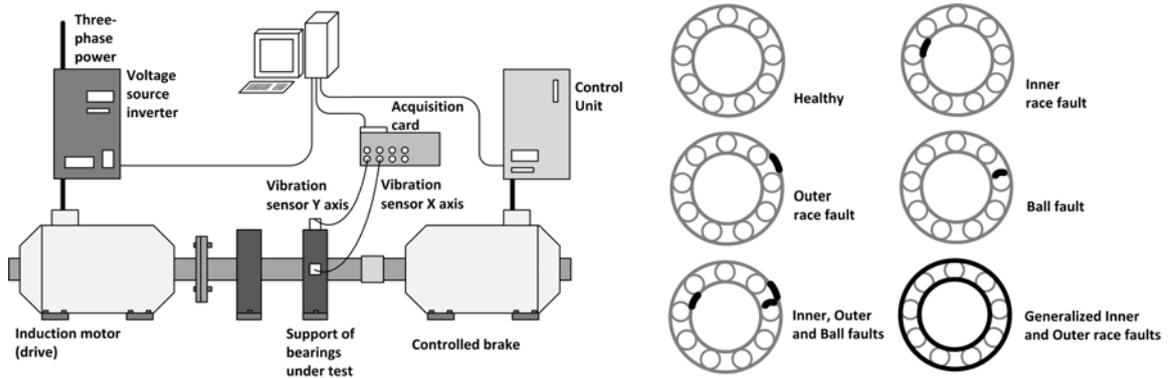


Fig. 2: a) Scheme of the experimental setup b) Six bearing defaults to be classified

vibration spectrum and the characteristic fault frequencies do not appear. These faults are common in industry, while they are often neglected in the research literature.

The bearing faults produce a variation of the vibration mode in the machine, and statistical features from time [9] [10], as RMS, crest factor, standard deviation, rectified skew and others are useful to exhibit the system vibration condition. However, the fault effects in the acquired signal will be attenuated as the vibration is acquired further from the origin of the fault. But if the fault has enough presence, this will be detected at any point of the machine.

Neural networks [11] [12] have been applied to the identification of defect locations (e.g., on the inner or outer ring of the bearing). Other techniques such as autoregressive modeling have also been used to analyze the spectra of defective bearings for classification [13]. However, in general, feedforward neural networks are generally more accurate than linear autoregressive models and RBF networks to classify healthy and defective bearings. In [14] features are selected by using the principal component analysis (PCA), using the too simple method of choosing the attributes having the largest principal component (all other information is lost). The classification is computed by means of a multilayer perceptron (MLP) or a RBF.

The experimental setup is based on one induction motor and a controlled brake. They are connected by means of an additional shaft in which two bearing supports are mounted. The driving motor is controlled by an inverter. The drive is a 0.37 kW motor at 2780 rpm rated speed.

A set of six identical bearings have been used covering the most important bearing fault scenarios : healthy (h), inner race fault (i), outer race fault (o), ball fault (b), inner-outer-ball (iob) faults at the same time and generalized degradation in inner and outer races (gdi). The experimental setup and the bearing set are shown in fig. 2. Two monoaxial (orthogonal X and Y axis) piezoelectric accelerometers are attached using screw mounting to one of the bearing support and its data were collected using an acquisition card sampling at 10 kS/s, 1 second for each measurement. In order to take into account different speed and torque combinations, 25 different stationary working conditions have been included: from 80%

to 100% of rated speed, and from 80% to 100% of rated torque acquiring every variation of 5% of torque and speed. From the acquired vibration time signals a set of 30 (15 for each axis) statistical features was extracted. Feature analysis and discriminant analysis [5] have reduced the number of features to six: for each axis, RMS, variance and shape factor.

A. The EXIN CCA approach

The intrinsic dimension of data is 2 for the rated TS. It increases for the pool of 25 TS's. Table 1 shows some intrinsic dimension estimators [15] for both cases: the geodesic minimum spanning tree (GMST [16]), the correlation dimension (CORR [17]), the PCA and the maximum likelihood (MLE [18]).

Figure 3 shows the CCA projection (right) for the TS drawn under *rated* working conditions (100% rated speed, 100% rated torque): the six classes correspond in pairs (in the sense of vicinity): h/i, o/b and iob/gdio. This is the same for each working condition. The corresponding $dy-dx$ diagram (left) shows that most points lie on the bisector, which means the six dimensional manifold is quasi linear. Indeed, the smallest distances, which represent the class interdistances, that is the distances between members of the same class, are represented very close to the bisector: it means that each pair of classes lies on a linear manifold. However, the class interdistances, that is the distances between classes, show a few well separated clusters around the bisector. Each cluster represents the distances between pairs of clusters. Hence, it demonstrates the presence of distant pairs. Resuming, it can be deduced that, in the six dimensional data space, the three linear manifolds are not coplanar: the data manifold is non-Riemannian (hence, the principal component analysis is not the better projection tool [15]).

TABLE I. INTRINSIC DIMENSIONALITY

estimator	rated CCA	CCA pool
GMST	1.8475	3.1259
CORR	0.74223	1.4441
PCA	2	2
MLE	2.0862	2.9851

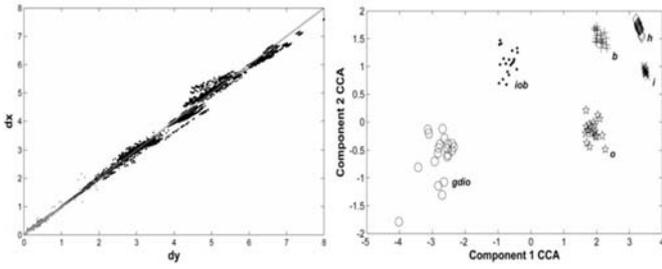


Fig. 3: Rated CCA, 25 samples per class, $\alpha=0.5$, 10 epochs

The projections corresponding to the 25 working conditions (represented by the two dimensional vector \mathbf{p}) are computed by using EXIN CCA with the OLS refinement. The initial learning rate (which decreases exponentially in time) is equal to 0.5 for the rated CCA and 0.1 otherwise. The initial conditions in the latent space are chosen by using the latent points of the nearest CCA, in the sense of working conditions. In this application, the initial values in the latent space are drawn from the latent points of the CCA whose TS has been collected with the same torque value, but with the previous value of speed (i.e. 5% more). If the TS is drawn under rated speed, the CCA corresponding to the same value of speed, but with the previous value of torque (i.e. 5% more) is considered.

The OLS refinement follows. It has been computed by using the normal equations [5]. If a RBF is trained with the 25 affine parameter vectors, which, in this application are six dimensional (the four components of matrix \mathbf{A} and the two components of vector \mathbf{b}), the values of the transformation for the interpolation case are estimated. Figure 4 shows the RBF outputs. The shape of the surfaces is a function of the CCA variability, after having considered the particular values for the learning rate and the initial conditions. The remaining variability depends on the change of class position and shape under different values of the operating conditions, which is exploited in the next step (classification).

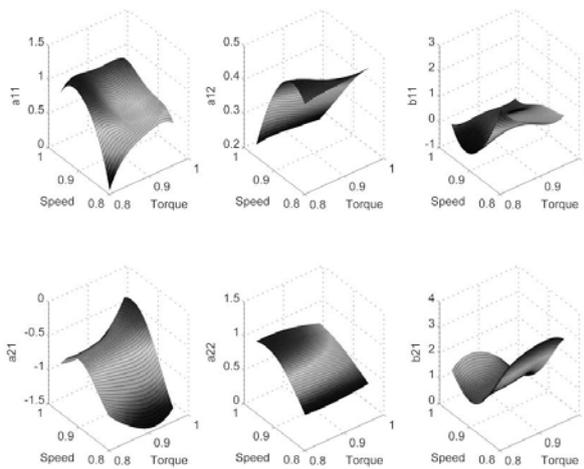


Fig. 4: RBF outputs for different operating conditions (per unit)

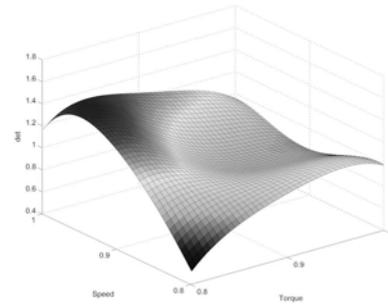


Fig. 5: Determinant of matrix \mathbf{A} for different operating conditions (p.u.)

Figure 5 shows the determinant of matrix \mathbf{A} estimated by using the RBF outputs for different operating conditions. This determinant is close to 1 (orthogonal transformation) for a certain range of speed and torque. In the borders of the domain, it decreases, always remaining positive. It means that the transformation is no more a rotation, but is affine: This is also due to the physical phenomenon which gives a distortion of the transformation. However, there is no reflection.

The *global* rated CCA projection, namely the output of EXIN CCA, is invariant to the CCA projection variability. The class centroids of the 25 CCA's in the rated projection are visualized in fig. 6. The class pairs are well separated, but there is an overlapping in h/i. This figure shows how the working conditions influence the bearing defects.

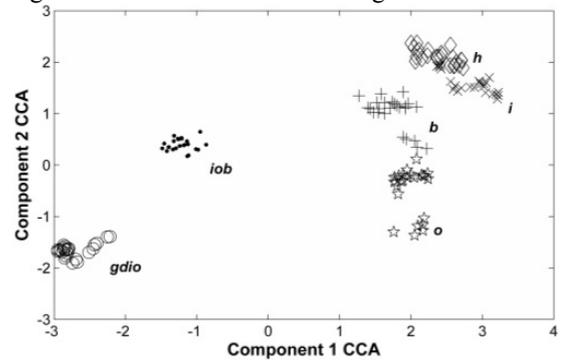


Fig. 6: EXIN CCA : class centroids of the global TS

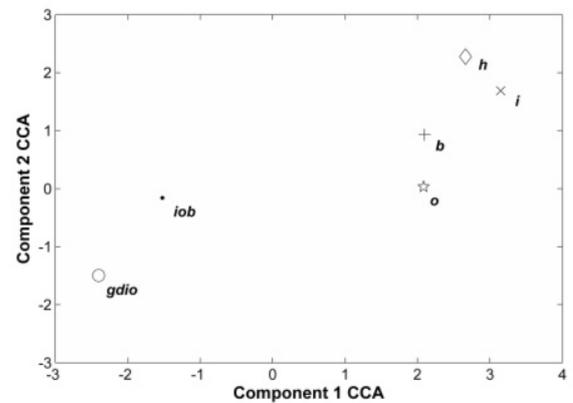


Fig. 7: EXIN CCA : class centroids of the test TS (new working conditions: 98% rated speed, 93% rated torque)

In order to assess the quality of the projection in the global rated CCA, the Local Continuity Meta-Criterion (LCMC [19]) has been used as quality measure. It measures the invariance of the neighborhood of data in the projection. It is in the range [0,1] and is higher for better projection. Using five neighbors, the following values have been found :

$$\begin{pmatrix} 0.87 & 0.78 & 0.83 & 0.82 & 0.81 \\ 0.80 & 0.83 & 0.83 & 0.81 & 0.83 \\ 0.82 & 0.83 & 0.82 & 0.81 & 0.80 \\ 0.86 & 0.88 & 0.84 & 0.83 & 0.82 \\ 0.82 & 0.89 & 0.84 & 0.78 & 0.78 \end{pmatrix}$$

where each component corresponds to one of the 25 working conditions, that is the 1,1-component yields the value of LCMC for 100% speed and 100% torque (rated case) and the 5,5-component yields the value of LCMC for 80% speed and 80%. The quality of the EXIN CCA output is high.

Resuming, EXIN CCA is important both for classification (only one classification is needed) and as a *sensor* of the physical phenomenon (the trajectories of the centroids show the influence of speed and torque on the bearing faults).

For checking the generalization properties of the proposed setup, a test set for the recall phase (interpolation) has been acquired: it is composed of 120 experimental data vectors, measured for 98% rated speed and 93% rated torque.

The corresponding tuning vector θ is computed by using the RBF neural network. After tuning, the new data are reported to the rated CCA for classification (see fig. 7). The fact that the OLS correction basically depends on the CCA remaining variability and only, as a secondary effect, on the physical phenomenon, can be deduced by comparing figs. 6 and 7: the centroids of the projections of the test set are placed in the corresponding regions, projections of the training set. Figure 8 shows the setup for the recall phase. The six dimensional feature vector is fed to the CCA neural network and the corresponding torque and speed parameters are fed to the RBF in order to find the parameters of the affine transformation (least squares tuning). A simple madaline, composed by two adalines [5], is used for transforming the CCA projection to the rated map. The whole classification setup is EXIN CCA.

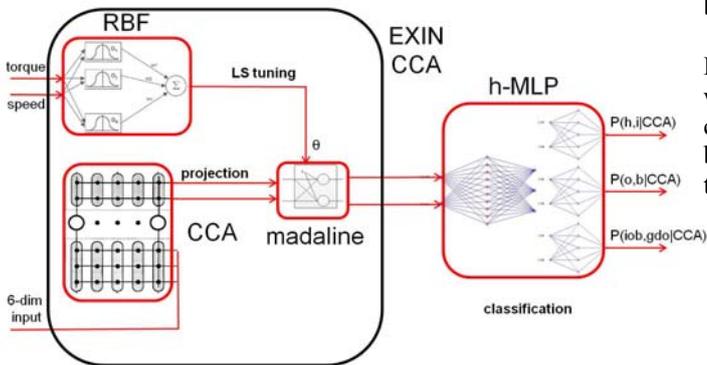


Fig. 8: Projection and classification setup

B. The hierarchical MLP for classification

The building block of the classification setup is a multilayer perceptron. It has two layers and the hidden activation function is the hyperbolic tangent. For the two-class problem, the output activation function is the logistic sigmoid. For the case of more classes, the softmax function is used. Training uses the backpropagation rule for the gradient estimation and the scaled conjugate gradient as minimization technique [5]. With this configuration, MLP outputs the a posteriori probability of each class, conditioned by the input.

A hierarchy of multilayer perceptrons (hMLP) has also been used (see fig. 8) in order to exploit the particular configuration of points in data and latent space. A first MLP classifies the three class pairs and enables three MLP's, each trained on a pair of classes. The posterior probability for each class is then given by the product of the three-class MLP output with the corresponding two-class MLP output. For instance, define as h the event *healthy*, as hi the event pair *healthy/inner race fault* and as CCA the two dimensional input from EXIN CCA. It follows:

$$P(h|CCA) = \underbrace{P(h|hi, CCA)}_{\text{output 2-class MLP}} \underbrace{P(hi|CCA)}_{\text{output 3-class MLP}} \quad (5)$$

At first, a classification is done by directly using all the six dimensional vectors (i.e. all the TS's) without projection. The results are given in tab. 2. Despite the good TS classification, the test set classification is poor. It means that a good generalization has not been achieved.

Secondly, PCA is used in order to check the validity of the linear projection. By using the hMLP, the classification ratio for the test set is 86.7% and the corresponding confusion matrix is given by:

$$\begin{pmatrix} 20 & 0 & 0 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 & 0 & 0 \\ 0 & 0 & 16 & 4 & 0 & 0 \\ 0 & 0 & 2 & 18 & 0 & 0 \\ 0 & 0 & 0 & 0 & 10 & 10 \\ 0 & 0 & 0 & 0 & 0 & 20 \end{pmatrix}$$

where the rows represent the true classes and the columns represent the predicted classes They are both ordered as h, i, o, b, iob and gdo.

Thirdly, the output of EXIN CCA is classified by a single MLP. The decision regions for the test set are shown, together with the EXIN CCA projected data, in fig. 9 and the classification ratios in tab. 2. A very good improvement has been achieved, thanks to the EXIN CCA projection and tuning.

TAB II. CLASSIFICATION RATIOS

%	6-dim input, hMLP	2-dim input MLP	2-dim input hMLP
training set	95	97.8	98.3
test set	72	95.8	95

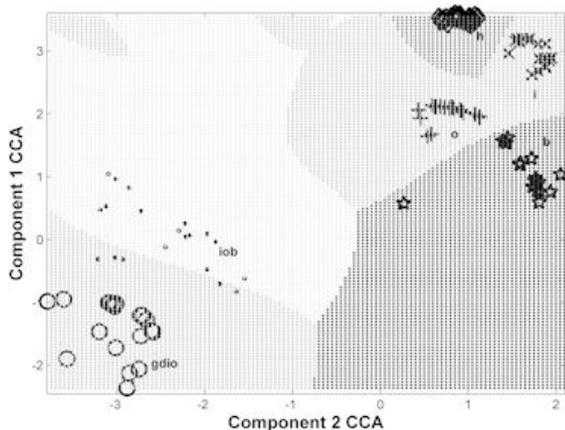


Fig. 9: Decision regions for the single MLP

The importance of EXIN CCA is confirmed by the last experiment, where the EXIN CCA projection is fed to the hMLP. Here, the training classification ratio is 98.3% (3-class: 100%, h/i: 99.3%, o/b: 98.9%, iob/gdio: 96.7). The classification ratio for the test set is 95%. The decision regions are shown in fig. 5. It can be seen that all points corresponding to healthy machine are correctly classified. Indeed, the confusion matrix for the latter case is given by:

$$\begin{pmatrix} 20 & 0 & 0 & 0 & 0 & 0 \\ 0 & 20 & 0 & 0 & 0 & 0 \\ 0 & 0 & 19 & 1 & 0 & 0 \\ 0 & 0 & 0 & 20 & 0 & 0 \\ 0 & 0 & 0 & 0 & 15 & 5 \\ 0 & 0 & 0 & 0 & 0 & 20 \end{pmatrix}$$

Table 2 shows that the obtained results are of the same order as in the case of a unique MLP (only a slight improvement in the training set classification). Hence, it can be deduced that the most important tool for the neural approach is EXIN CCA.

IV. CONCLUSIONS AND FUTURE WORK

This paper explains the EXIN CCA neural network, a variant of CCA. It is a *stabilizer* of the CCA projection, that is it yields an invariant nonlinear projection. It achieves this result by operating on the initial conditions and the learning rate. It refines the result by applying an affine transformation to a unique invariant map. This transformation is computed by means of the ordinary least squares technique, which is, however, suboptimal for the data matrix in (4) is noisy. Future work will deal with the use of the constrained total least squares which takes into account the structure of the data matrix. EXIN CCA also uses a weighting factor multiplying the learning rate in order to take into account the neighbor working conditions for the recall/interpolation phase. An open problem remains the analysis of extrapolation, i.e. working conditions outside the range.

From the bearing fault diagnosis point of view, the use of temporal information has given very high classification ratios by using a neural approach. Future work will deal with time frequency integration and with the use of stator currents instead of vibration signals.

The research proposed in this paper will be also used in other classification problems, mainly the classification of faults in electrical machines, where the goal is to improve the results obtained by CCA in [20].

Thanks to the potential of the EXIN CCA + hMLP setup, there is also the possibility of using this technique as a sensor of the physical phenomenon, in the sense that the accurate invariant projections, the posterior probabilities and the corresponding decision regions yield a deep insight of the problem at hand. This possibility allows, for instance, to improve the feature selection, as a feedback from the classification to the system to be analyzed.

REFERENCES

- [1] P. Demartines and J. Héroult., "Curvilinear Component Analysis: a Self-Organising Neural Network for Non-Linear Mapping of Data Sets", *IEEE Trans. on Neural Networks*, vol 8, n.1, pp.148-154, 1997.
- [2] J. Héroult, C. Jausions-Picaud and A. Guérin-Dugué, "Curvilinear component analysis for high-dimensional data representation: I. Theoretical aspects and practical use in the presence of noise", *Proceedings of In J. Mira and J. Sánchez-Andrés, editors, proceedings of IWANN 1999, Lecture Notes in Computer Science 1607*, pp. 625-634, Springer-Verlag, 1999.
- [3] A. Guérin-Dugué, P. Teissier, G. Delso-Gafaro. and J. Héroult, "Curvilinear Component Analysis for High-dimensional Data Representation: II. Examples of introducing additional mapping constraints for specific applications", *Proceedings of In J. Mira and J. Sánchez-Andrés, editors, proceedings of IWANN 1999, Lecture Notes in Computer Science 1607*, pp. 635-644, Springer-Verlag, 1999.
- [4] J. Sun, M. Crowe and C. Fyfe, "Extending metric multidimensional scaling with Bregman divergences". *Pattern Recognition*, vol. 44, n.5, pp. 1137-1154, 2011.
- [5] C. Bishop, *Pattern Recognition and Machine Learning*, Springer; 2007.
- [6] G.Cirrincone and M. Cirrincone, *Neural Based Orthogonal Data Fitting: The EXIN Neural Networks*, *Wiley and Sons* (New York, USA), series : *Adaptive and Learning Systems for Signal Processing, Communications and Control*, November 2010.
- [7] J.R. Stack, T.G. Habetler, and R.G. Harley, *Fault classification and fault signature production for rolling element bearings in electric machines*, *IEEE Tr. On Industry Applications*, vol.40, n°3, pp. 735- 739, 2004.
- [8] Immovilli, F.; Bellini, A.; Rubini, R.; Tassoni, C., "Diagnosis of Bearing Faults in Induction Machines by Vibration or Current Signals: A Critical Comparison," *Industry Applications, IEEE Transactions on*, vol.46, no.4, pp.1350-1359, July-Aug. 2010.
- [9] L. B. Jack, A. K. Nandi, and A. C. McCormick, "Diagnosis of rolling element bearing faults using radial basis function networks," *Appl. Signal Process.*, vol. 6, pp. 25–32, 1999.
- [10] R. B. W. Heng and M. J. M. Nor, "Statistical analysis of sound and vibration signals for monitoring rolling-element bearing condition," *Appl. Acoust.*, vol. 53, pp. 211–226, 1998.
- [11] T. Liu and J. Mengel, "Intelligent monitoring of ball bearings," *Mech. Syst. Signal Process.*, vol. 6, no. 5, pp. 419–431, 1992.
- [12] I. Alguindigue, A. Loskiewicz-Buczak, and R. Uhrig, "Monitoring and diagnosis of rolling element bearings using artificial neural networks," *IEEE Trans. Ind. Electron.*, vol. 40, no. 2, pp. 209–216, 1993.

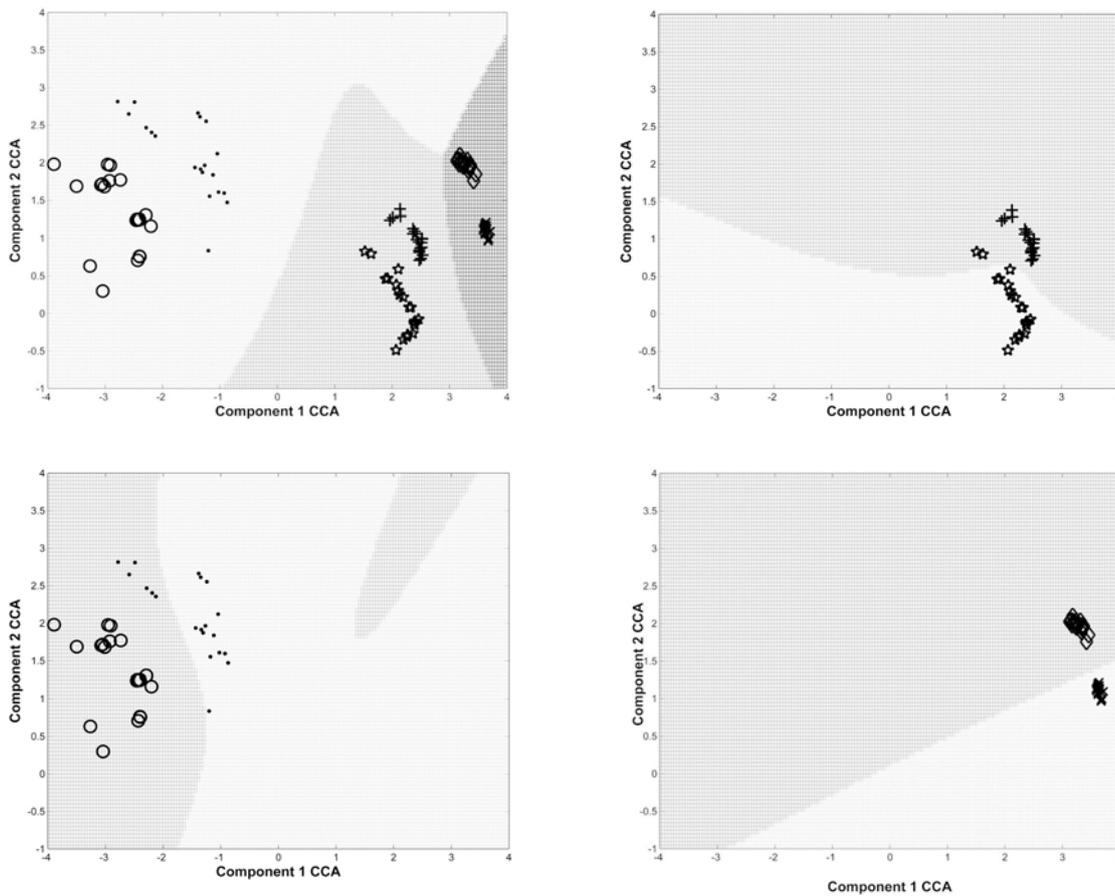


Fig. 10: Decision regions for the four MLP's of the hierarchical MLP

- [13] D. C. Baillie and J. Mathew, "A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings," *Mech. Syst. Signal Process.*, vol. 10, pp. 1-17, 1996.
- [14] A. Malhi and R.X Gao, "PCA-based feature selection scheme for machine defect classification", *IEEE Transactions on Instrumentation and Measurement*, vol.53, n.6, pp. 1517- 1525, 2004.
- [15] L.J.P. van der Maaten, E.O. Postma and H.J. van den Herik, "Dimensionality reduction: a comparative review", *Tilburg University Technical Report, TiCC-TR 2009-005*, 2009-
- [16] J.A. Costa and A. O. Hero, "Geodesic entropic graphs for dimension and entropy estimation in manifold learning", *IEEE Trans. on Signal Processing*, , pp. 2210-2221, vol.52, n.8, 2004.
- [17] D. Mo and S. H. Huang, "Fractal-based intrinsic dimension estimation and its application in dimensionality reduction", *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, n.1, pp. 59-71, 2012.
- [18] E. Levina and P. Bickel, "Maximum likelihood estimation of intrinsic dimension". In: *17th Annual Conf. on Neural Information Processing Systems*, 2005.
- [19] L. Chen and A. Buja, "Local multidimensional scaling for nonlinear dimensionality reduction, graph layout, and proximity analysis", *PhD thesis, University of Pennsylvania*, July 2006.
- [20] G. Cirrincione, M. Cirrincione and G. Vitale, " Diagnosis of three-phase converters using the VQP neural network", *2nd IFAC Workshop on Computer Software Structures, integrating AI/KBS System in Process Control*, Lund, Sweden, 11/13 August 1994.