Application of Pattern Recognition with Principal Component Analysis for Travelling Wave Protection

E. Vázquez
Universidad Autónoma de Nuevo León
P.O. Box 89F, Monterrey, NL, 66450, México
evm@ieee.org

Abstract - **This paper is concerned with a new approach for travelling wave distance protection, based on pattern recognition with principal component analysis (PCA), to be used for transmission line ultra-high speed protection. The proposed approach explores the possibility to characterize the wave front shape for internal and external faults of the protected transmission line. In this case, a PCA with neural networks is proposed as feature extractor to implement the pattern recognition process. The approach was proven with current and voltage samples from a three-phase 230 kV power system, which was simulated using the Electro-Magnetic Transients in DC program (EMTDC). The results show the feasibility to implement an algorithm for transmission line ultra-high speed protection.**

Keywords: Pattern recognition, Principal component analysis, Power system protection, Travelling waves.

I. INTRODUCTION

Power system protection has traditionally relied on the measurement of power frequency components for the detection of faults. In conventional protection schemes, the signals of high frequency introduced by a fault are considered as interference and are filtered out [1]. However, these high frequency components contain extensive information about the fault type, location, direction and sustain time. In fact, the high frequency transient signals generated by a fault contain more information about the fault than power frequency signals [2,3].

In ultra-high-voltage systems, in order to improve the transient stability, high-speed fault clearance is always desired. The post fault voltage and current are initially dominated by electromagnetic travelling waves. Based on the analysis of these transient state signals, the travelling wave-based protective relays can detect and locate the fault within several milliseconds after the fault. Due the development of new generation fibre-optic voltage and current measuring systems, the travelling wave protection will have much wider application in the near future [1,4].

The basic principle of the travelling wave distance protection is to measure the time interval between the arrival of an incident wave toward the fault point and that of the corresponding wave reflected from it. Most of the present schemes use the correlation function method to recognize the wave front returning from the fault [5,6]. This is the radar principle. Most recently, some different techniques have been used to improve the results obtain by the correlation function, as wavelets [7], neural networks [8] and pattern recognition methods [9]. However, these new approaches use the same concept to recognize the second wave front returning from the fault. The common characteristic of these algorithms is that they have at least an operation time of 3τ, where τ is the travel time between the relay and the fault point.

This paper describes a new approach for travelling wave distance protection; the proposed approach explores the possibility to characterize the wave front behavior for internal and external faults of the protected transmission line using the first wave from the fault. A PCA algorithm with neural networks extracts the features from the relaying signals (S_1 or S_2) in order to implement a pattern recognition process in a 2D space called feature space. All information is normalized to have zero mean and unity standard deviation. The representation of the original relaying signals in the feature space show a linearly separable structure, and it could be solved by any classification technique. It allows to discriminate between internal and external faults with an operation time of τ. Finally, an algorithm is proposed to implement this function in real time for transmission line protection.

II. THEORY OF TRAVELLING WAVE PROTECTION

When a fault occurs in the transmission line, by virtue of the superposition theorem, the fault injected components v_f and i_f can be acquired by subtraction of the steady-state components from the post fault signals (incremental signals) [6]. For distributed parameter model representation of a transmission line, the fault injected components v_f and i_f can be expressed in terms of a forward and backward travelling wave:

$$V(x,t) = F_1(x-ct) + F_2(x+ct) \qquad (1)$$

$$I(x,t) = F_1(x-ct) - F_2(x+ct)/Z_0 \qquad (2)$$

where c and Z_0 are the surge velocity and line characteristic impedance, respectively, and x is the distance that a travelling wave travels from the fault point.

The forward and backward travelling waves S_1 and S_2 used in the travelling wave protection are defined as follow:

$$2F_2(t) = V(t) - Z_0 I(t) = S_1(t) \qquad (3)$$

$$2F_1(t) = V(t) + Z_0 I(t) = S_2(t) \qquad (4)$$

$S_1(t)$ and $S_2(t)$ travel along the transmission line in opposite directions. When they hit a discontinuity, part of it will be reflected, and a part will pass to other sections of the system.

The principle of the travelling wave protection can be illustrated by the power system shown in Fig. 1. When a fault

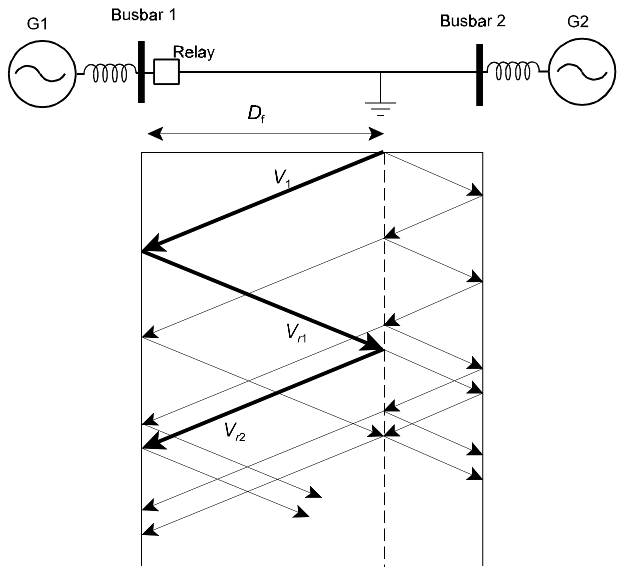


Fig. 1. Principle of the travelling wave protection.

occurs at a position that is D_f km away from the relay, travelling waves would be generated and propagate along the line. When the backward travelling wave V_1 arrives at the source G1 behind the relay, reflection would occur. The reflected wave V_{r1} would return along the line toward the fault point. At that point, part of it would be reflected, and part of it would be transmitted, if the fault resistence is not zero. The reflected wave V_{r2} would return to busbar 1 after some time.

If we can get the time interval t_0, between the arrival of V_{r1} and that of the backward wave V_{r2}, then D_f can be acquired from t_0 by:

$$D_f = v t_0 / 2 \tag{5}$$

Hence the identification of the signal V_{r2} becomes the key problem of travelling wave protection. The correlation function technique is always used to fulfil it [5,6]. For three-phase transmission lines, the mutual elements of the surge impedance matrix make travelling wave couple across the phases. To simplify the calculations, the modal analysis method is always adopted to de-couple the phase signals into three independent modal components, including one earth mode and two aerial modes. These modes have different velocity and attenuation and hence lead to dispersion effects on wave fronts describe by phase components. For fully transposed system, the two aerial modes have the same characteristic impedance and velocity. The modal transformation can be expressed by [6]:

$$\Delta v_m(t) = S^{-1}\,\Delta v(t) \tag{6}$$

$$\Delta i_m(t) = Q^{-1}\,\Delta i(t) \tag{7}$$

where $\Delta v(t)$ and $\Delta i(t)$ are the incremental phase voltages and $\Delta v_m(t)$ and $\Delta i_m(t)$ are the corresponding modal voltage and current. S^{-1} and Q^{-1} are the transformation matrices. Three of the constant modal transformation matrices for perfectly transposed lines are:

Clark transformation:

$$Q = S = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1/2 & \sqrt{3}/2 \\ 1 & -1/2 & -\sqrt{3}/2 \end{bmatrix} \tag{8}$$

Wedepohl transformation:

$$Q = S = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -2 \\ 1 & -1 & 1 \end{bmatrix} \tag{9}$$

Karrenbauer transformation:

$$Q = S = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \tag{10}$$

However, the performance of the correlation function method depends on the fault resistence, the system configuration and the mode type [9]. Thus, it is necessary to modify this method for more reliable application.

The proposed approach characterizes the wave front behavior for internal and external faults of the protected transmission line using the first wave from the fault (V_1 in Fig. 1). A PCA algorithm with neural networks extracts the features from the relaying signal S_1 in order to implement a pattern recognition process.

III. PRINCIPAL COMPONENT ANALYSIS

The PCA is a statistical technique falling under the general title of factor analysis [10]. The purpose of PCA is to identify the dependence structure behind a multivariable stochastic observation in order to obtain a compact description of it. When there is nonzero correlation between the observed variables the dimensionality, n, of the data space (number of observed variables) does not represent the number of independent variables, m, that are really needed to describe the data. We may liken m to the number representing the degrees of freedom of a physical system. In the statistical context the number n is called the superficial dimensionality of the data, whereas m is called the intrinsic dimensionality of the data. The stronger the correlation between the observed variables, the smaller the number of independent variables that can adequately describe them.

The n observed variables are thus represented as functions of m latent variables called factors, where $m<n$ and often $m<<n$. The simpler the mathematical form of the representation functions the more economical is the description of the dependence structure between variables. Traditional PCA is associated with linear transformations, which are the simplest and most mathematically tractable function forms for representation. The factor variables are also called *features* of the multivariate random signal, and the vector they form is a member of the *features space*.

The usual objective of the analysis is to see if the first few components account for most of the variation in the original data. If they do, then it is argued that the effective dimensionality of the problem is less than n. In other words, if some of the original variables are highly correlated, they are effectively the same thing and there may be near-linear constraints on the variables. In this case it is hoped that the first few components will be intuitively meaningful, will help us understand the data better, and will be useful in subsequent analysis where we can operate with a smaller number of variables. PCA transforms a set of correlated variables to a new set of uncorrelated variables.

Suppose $\boldsymbol{X}^{\mathrm{T}}$=[x_1, ..., x_n] is a n-dimensional random variable with mean μ and covariance matrix Σ. A new set of variables y_1, y_2, ..., y_n, which are uncorrelated, can be represented as a linear combination of the x_i, so that:

$$y_j = a_{1j}\, x_1 + a_{2j}\, x_2 + \ldots\ldots + a_{nj}\, x_n = a_j^{\mathrm{T}} \boldsymbol{X} \qquad (11)$$

where $a_j^{\mathrm{T}} = [\ a_{1j}\ \ a_{2j}\ \ldots\ldots\ a_{nj}\ ]$ is the vector of principal components. We can prove that the *jth* principal component is the eigenvector associated with the *jth* largest eigenvalue of Σ [11]. It is common to calculate the principal components of a set of variables after they have been standardized to have a unit variance. This mean that one is effectively finding the principal components from the correlation matrix $\boldsymbol{R}$ rather than from the covariance matrix. The mathematical derivation is the same, and the principal components are the eigenvectors of $\boldsymbol{R}$. However, it is important to realize that the eigenvalues and eigenvectors of $\boldsymbol{R}$ will generally not be the same as those of Σ.

In our case, a n-dimensional vector $\boldsymbol{X}$=[x_1 x_2 .... x_n]$^{\mathrm{T}}$ is formed with samples of the travelling wave $S_{1.}$

Fig. 2 shows a geometrical interpretation of the principal component subspace; based on the variance criterion the principal component should be the one where the signal has more energy; the least principal direction is the one with the least energy. If the signal is, zero mean that the maximum energy direction is also the direction of maximum spread or, in information theory, the direction that contains the most information of the signal (assuming it is Gaussian).

IV. PCA NEURAL NETWORK

There are two techniques to calculate the principal components: the batch PCA methods and the neural PCA models. Batch methods are used to process finite sets of data. Because of storage consideration batch methods are preferred when relatively few data are to be processed relatively few times. Adaptive methods (neural networks), on the other hand, are preferred with arbitrarily long or infinite sets of data to be processed. Such methods require less memory for data storage, since intermediate matrices are not explicitly formed. In addition, adaptive methods with constant step-size parameters that do not tend to 0 as $k\to\infty$, can track gradual changes in the optimal solution rather inexpensively compared to the batch models. In general, the interest for adaptive techniques arises when $\boldsymbol{R}$ is not known.

There are different models of PCA neural networks, namely Oja's model, Földiák's model, the GHA model, the APEX model, Rubner's model, etc. For our problem, we will use the GHA model [11].

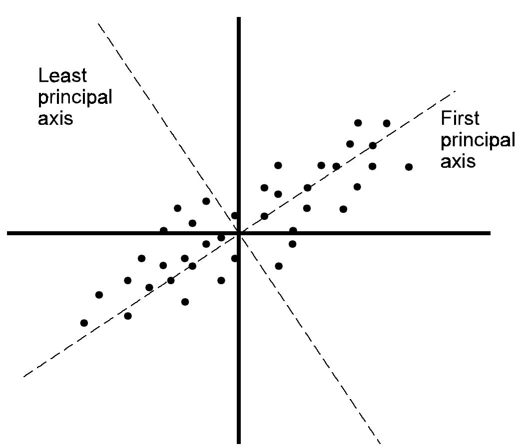


Fig. 2. Geometrical interpretation of the principal component subspaces.

V. THE GENERALIZED HEBBIAN ALGORITHM

This algorithm proposed by Sanger [11] is capable to extract all the principal components from a data set. The model has m output neurons y_1,.....,y_m and n inputs x_1,.....,x_n. There are only feedforward connections between input and output and the output is a linear function of the input:

$$y_i = w_i^{\mathrm{T}}\, x$$

The updating equations for neuron i (i=1,....., m) are:

$$\Delta w_{ij,k} = \beta_k \left(y_{ik} \left[x_{jk} - \sum_{l<i} y_{lk} w_{lj,k} \right] - y_{ik}^2 w_{ij,k} \right) \qquad (12)$$

where W_i = [w_{i1} w_{i2} ..... w_{in}]$^{\mathrm{T}}$. The model extracts the first m principal normalized eigenvectors of $\boldsymbol{R}$ under the following assumptions:

A. The input sequence {x_i}, is at least wide-sense stationary with autocorrelation matrix $\boldsymbol{R}$, whose eigenvalues are positive, arranged in descending order, and where the m largest eigenvalues are distinct: $\lambda_1 > \ldots > \lambda_m \geq \lambda_{m+1} \geq \ldots \geq \lambda_n > 0$.
B. The step-size parameter sequence β_k is such that

$$\beta_k \to 0 \quad \text{as } k \to \infty\ , \quad \text{and} \quad \sum_{k=0}^{\infty} \beta_k = \infty$$

This assumption will be useful for showing the asymptotic convergence of the algorithm. We can prove that $\lim_{(t\to\infty)} w_1=\pm e_1$, $\lim_{(t\to\infty)} w_2=\pm e_2$, ..., $\lim_{(t\to\infty)} w_m=\pm e_m$, where e_1, e_2,...., e_m are the eigenvectors of $\boldsymbol{R}$. Using a more rigorous approach, Hornik and Kuan [11] have shown that the only asymptotically stable equilibria of GHA are the points W=[w_1 ,..,w_m]$^{\mathrm{T}}$= [$\pm e_1$,.., $\pm e_m$]$^{\mathrm{T}}$ while all other equilibria are unstable. The network implementation of this local GHA rule is shown in Fig. 3.

Once the process has finished, the matrix w_i = [w_{i1} w_{i2} ..... w_{in}]$^{\mathrm{T}}$ represent the n principal components of the data storage in the matrix $\boldsymbol{X}$. Making

$$\boldsymbol{Y} = \boldsymbol{W}^{\mathrm{T}} \boldsymbol{X} \qquad (13)$$

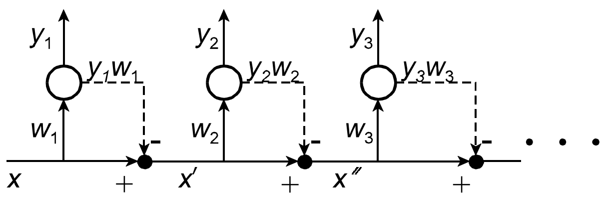


Fig. 3. The network model for the local GHA rule.

we obtain the projection of the original data on the subspace of the principal components. Suppose we have a (50x10) matrix $\boldsymbol{X}$ (50 samples of 10 variables in column format); it can be represented as a group of 10 points (or vectors) on a 50-dimension space. If we apply the GHA rule describe above for n=2 (to calculate the first two principal components from $\boldsymbol{X}$), we could represent the same 10 points in a two-dimension space ($\boldsymbol{W}$ is (50x2) and $\boldsymbol{Y}$ is (2x10)).

VI. PCA APPLICATION TO THE TRAVELLING WAVE DISTANCE PROTECTION

The principal idea to use the PCA is carried out a pattern recognition process to discriminate between internal and external faults using the basic features extracted from the first wave that arrive to the relay. We used the EMTDC program [12] to simulate the three-phase 230 kV power system shown in Fig. 4 and characterize the first wave from the fault. Details of the transmission lines used in this study are contained in the appendix. A horizontal line configuration was chosen.

For the present study, the Wedepohl transformation matrix was selected with mode 2 (aerial mode), and the n-dimensional vector $X=[x_1 \;\; x_2 \; .... \; x_n]^T$ is formed with samples of the travelling wave S_1 using a simulation time step of 1μsec. In order to consider the first wave front from the fault, we decide to use 31 samples of S_1 for distinct fault points and inception time with respect to the 60 Hz voltage signal, with 5 samples before and 26 after the fault inception.

In this case, 31 samples represent 31 μsec, which is smaller time that the travel time along the transmission lines (0.338 ms for L1 and 0.341 ms for L2). The purpose is to avoid the effect of the continuos reflection process present in the line ends. The fault conditions simulated were:

Internal faults: 10 to 90 and 95 % for 4 to 19 ms.
External faults: 5 and 10 to 90 % for 4 to 19 ms.

All 320 faults were solid three-phase to ground (R_f = 0). So, we have an input matrix $\boldsymbol{X}$ with 31 rows (samples of S_1) and 320 columns (fault conditions).

It is important to realize that the principal components of a set of variables depend critically upon the scales used to measure the variables. The practical outcome of the above result is that principal components are generally changed by scaling and that they are therefore not a unique characteristic of the data. An option is use the standard deviation for each variable to scale it. This ensures that all variables are scaled to have a unit variance and so in some sense have equal importance. This scaling procedure is still arbitrary to some extent, is data dependent and avoids rather than solves the scaling problem. One of this preprocessing methods [11] normalizes the input so that they have zero mean and unity standard deviation, as:

$$x_{i,\ new} = \frac{x_i - m}{\sigma} \qquad (14)$$

This process generates two new vectors which contain the mean and standard deviation of the original inputs. Once the network has been trained, these vectors should be used to transform any future inputs.

Using the GHA algorithm, we calculate the two first principal components to obtain a two-dimensional representation of the original data using (13). The results of this process are shown in Fig. 5 in two dimensions, where **+** are the faults in L1 and ***** are the faults in L2. We can see that principal components convert the original 31-dimension vectors in 2-dimension vectors, but there is not a specific feature that allows to implement a pattern recognition process.

The result shows in the Fig. 5 indicate a strong effect of the fault condition on the principal components. However, an analysis of this result shows that exist a specific behavior with respect to the fault inception time. So, we repeat the process for all faults with the same inception time independently its position; in other words, we obtained 16 pairs of principal component for each inception time from 4 to 19 ms. These results are shown in Fig. 6.

We can see that neural network extract the principal features from the travelling wave S_1, let it know a special data structure, which could be solved by any classification technique, as statistical, neural network, fuzzy logic, etc. However, it is necessary to test if this behavior is independent

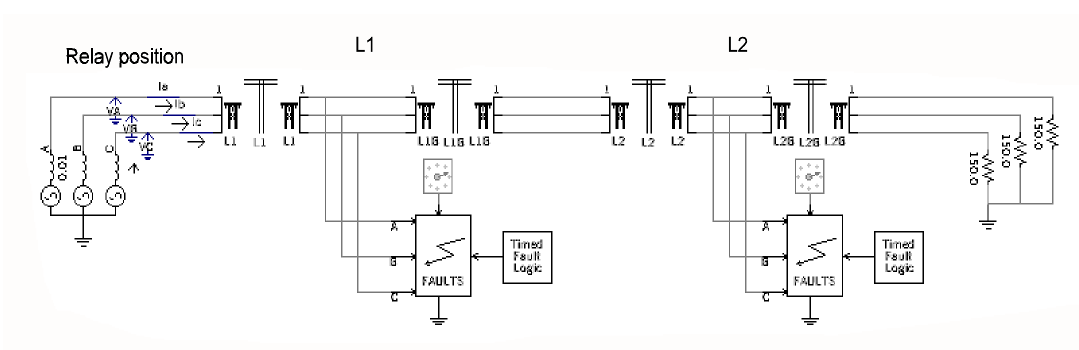


Fig. 4. Three-phase 230 kV power system.

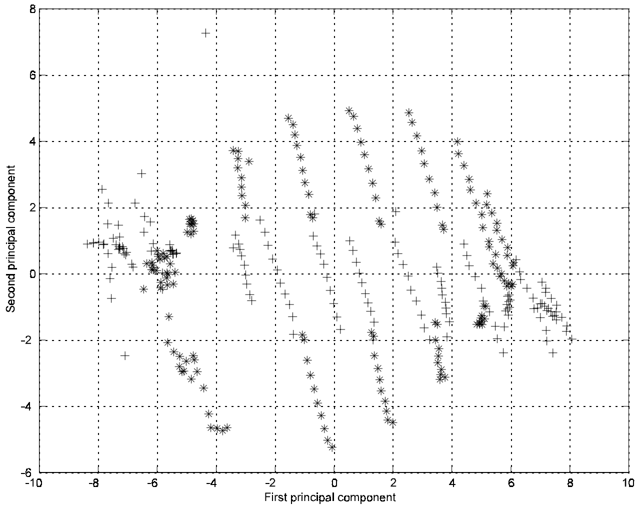


Fig. 5. Fault conditions in the PCA subspace.

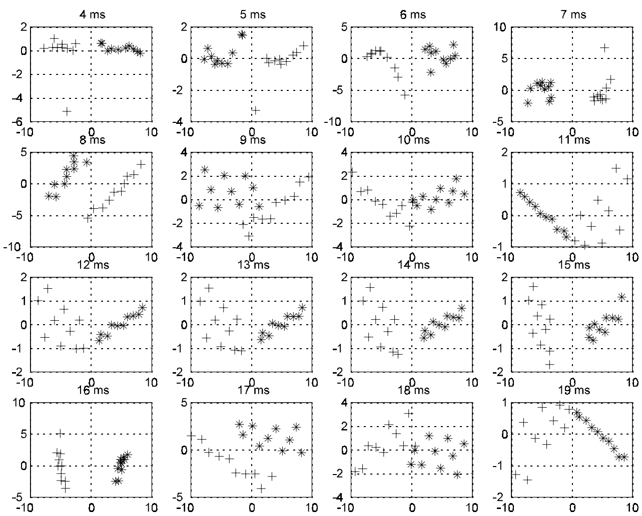


Fig. 6. Fault conditions for the same inception time represent in the PCA subspace.

of the fault position, either internal and external faults. With this purpose, the next six faults were simulated:

For an inception time of 8 ms:

- Fault in L1 at 82 and 91 km from relay (□).
- Fault in L2 at 8 and 15 km from transmission lines union (◇).

For an inception time of 8.4 ms:

- Fault in L1 at 85 km from relay (△).

For an inception time of 7.8 ms:

- Fault in L2 at 20 km from transmission lines union (▽).

These six faults were not being considerate in the principal components extract process and include two faults did not occur at 8 ms, but nearly it. Fig. 7 describes how the PCA differentiates between internal and external faults using these six faults as an example. The first two-graphics show the 31-samples of the travelling wave S_1 for each fault, before and after the scaling process. Once these signals are normalized, it is possible to observe some differences between the waveforms for either internal and external faults. The principal features of these differences were captured in the PCA vectors during the neural network training. After

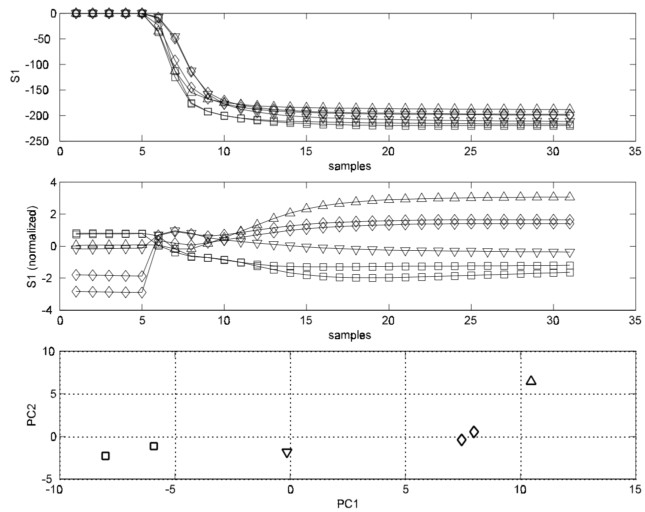


Fig. 7. Discrimination between internal and external faults using the PCA.

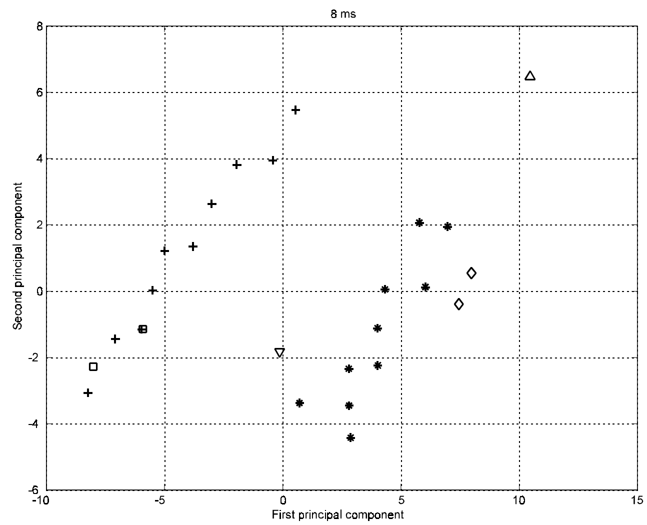


Fig. 8. Fault conditions (26) in the PCA subspace.

that, the normalized signals are represented in the principal component subspace, where those differences are amplified, showing that the final structure of the fault conditions in 2D is linearly separable (third graphic); it allows to use any classification technique. Fig. 8 show all 26 fault conditions in the PCA subspace, where the marks **+** and ***** correspond to the original 20 patterns shown in Fig. 6 for faults in L1 and L2 respectively, with an inception time of 8 ms.

VII. ALGORITHM

The results describe above indicate that is possible to implement an algorithm to protect all the transmission line using 16 pairs of principal components for distinct fault inception times on a 60 Hz cycle. The diagram blocks in Fig. 9 suggest the following logic: using a 31-data window, the S_1 signal is formed continually apply a modal transformation; when a fault occurs, the inception time could be determinate approximately using the transition from S_1=0 to $S_1 \neq 0$ (fault detector algorithm used in overcurrent digital relays [1]). Once the inception time is known, the algorithm selects the corresponding transformation vectors (previously stored) and

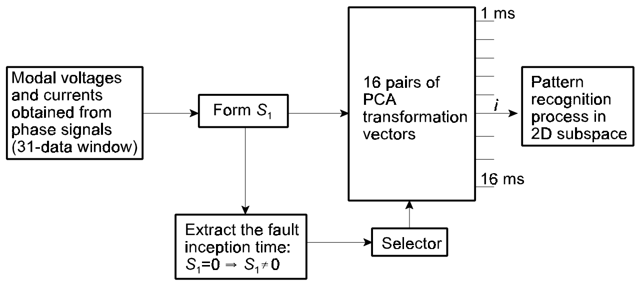


Fig. 9. Algorithm.

converts the original data to the principal component subspace to carry out a pattern recognition process and decide if the fault is internal or external of the protected transmission line.

Although the results obtained are good, it is necessary to show the feasibility of this kind of algorithm, and to study the effect of the power system configuration and the transmission line impedances in the principal components extraction process as the next steps of this approach.

VIII. CONCLUSIONS

The basic principle of the travelling wave distance protection is to measure the time interval between the arrival of an incident wave toward the fault point and that of the corresponding wave reflected from it. The correlation function, wavelets, neural networks and pattern recognition methods have been used to solve this problem.

The proposed approach characterizes the wave front behavior for internal and external faults of the protected transmission line using the first wave from the fault. A PCA algorithm with neural networks extracts the features from the relaying signal S_1 in order to implement a pattern recognition process. It will allow to discriminate between internal and external faults with an operation time of τ, instead 3τ as in the previous methods.

The final structure of the fault conditions in the PCA subspace is linearly separable for all faults with the same inception time, independently of its position. So, any classification technique could be used to discriminate between internal and external faults. Finally, an algorithm is proposed to carry out this function in real time for transmission line protection.

IX. ACKNOWLEDGMENT

The author wishes to acknowledge to the Mexican Research Council (CONACyT) for providing funds for the realization of this paper through project 25909A.

X. REFERENCES

[1] P.M. Anderson, *Power System Protection*, McGraw Hill, New York, 1999.

[2] M. Vitins, "A fundamental concept for high speed relaying", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-100, no. 1, January 1981, pp. 163-173.

[3] M.M. Mansour and G.W. Swift, "A multi-microprocessor based travelling wave relay - Theory and realization", *IEEE Transactions on Power Delivery*, vol. PWRD-1, no. 1, January 1986, pp. 272-279.

[4] Z.Q. Bo and A.T. Johns, "Transient based protection - A new concept in power system protection", *Proceedings of the 1997 International Conference on Power System Transients*, pp. 401-406, June 22-26, Seattle.

[5] M. Vitins, "A correlation method for transmission line protection", *IEEE Transactions on Power Apparatus and Systems*, vol. PAS-97, no. 5, Sept/Oct 1978, pp. 1607-1617.

[6] E.H. Shehab-Eldin and P.G. McLaren, "Travelling wave distance protection - Problems areas and solutions", *IEEE Transaction on Power Delivery*, vol. 3, no. 3, July 1988, pp. 894-902.

[7] J. Liang, S. Elangovan and J.B.X. Devotta, "A wavelet multiresolution analysis approach to fault detection and classification in transmission lines", *Electrical Power & Systems*, vol. 20, no. 5, 1998, pp. 327-332.

[8] W.M. Al-hasawi, N.H. Abbasy and M.M. Mansour, "Using fault induced transients and neural network for TL ultra high speed fault detection and classification", *Proceedings of the 1997 International Conference on Power System Transients*, pp. 412-417, June 22-26, Seattle.

[9] J. Lang, S. Elangovan and J.B.X. Devotta, "Pattern recognition technique for travelling wave protection", *Electric Machines and Power Systems*, vol. 26, 1998, pp. 1007-1018.

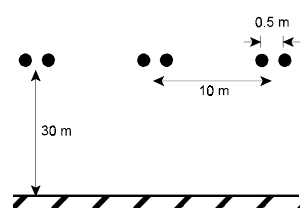
[10] V. Cherkassky and F. Mulier, *Learning from Data*, John Wiley & Sons, Inc., New York, 1998.

[11] K.I. Diamantaras and S.Y. Kung, *Principal Component Neural Networks - Theory and applications*, Wiley & Sons, Inc., New York, 1996.

[12] D. Woodford, *Introduction to PSCAD/EMTDC V3*, Manitoba HVDC Research Centre Inc., Winnipeg, 2000.

XI. APPENDIX

Line 1: two conductors per phase, r=0.01 m, R_{DC}=0.03206 ohms/km, Z_0=313.15 ohms, τ=0.338 ms.
Line 2: one conductor per phase, r=0.02034 m, R_{DC}=0.03206 ohms/km, Z_0=391.77 ohms, τ=0.341 ms.



Horizontal line configuration.

Ground resistivity: 100.0 ohm-m.
Relative ground permeability: 1.0
Shunt conductance: $1.0e^{-10}$ mhos/m.