A modified PCA based on the minimum error entropy

Zhenhua Guo, Hong Yue and Hong Wang

***Abstract*— Conventional principal component analysis (PCA) minimizes the total error variance, which may be inappropriate for the non-Gaussian distribution systems. In this paper the entropy is proposed as a more general index for PCA model, and then a modified PCA with the optimization for the minimum error entropy via a genetic algorithm (GA) is addressed.**

I. INTRODUCTION

PCA reduces the dimension of interrelated data set with the minimized error variance [1]. This implies that the error data are Gaussian and the existing methods may be inappropriate for the non-Gaussian error data. Hence, a more general index (i.e., the entropy) should be used to measure the uncertainties in the error. In this paper an adequate principal components are firstly selected via *cumulative percent variance* (CPV) [2] criteria, and then an optimal rotation on loadings is made via a *genetic algorithm* (GA) [3] for the minimal error entropy.

II. PCA AND MINIMUM ERROR ENTROPY

A. Principal Component Analysis

Consider an n-by-m data matrix $\mathbf{X}$ that consists of n observations with m variables. In PCA $\mathbf{X}$ is decomposed into the sum of the outer product of m pairs of vectors as follows.

$$\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \cdots + \mathbf{t}_m\mathbf{p}_m^T \tag{1}$$

where the vector $\mathbf{t}_i \in R^n$ is called ith *principal component* (PC), and the vector $\mathbf{p}_i \in R^m$ is called ith *loading*. Equation "(1)" is also subjected to $\text{var}(t_i) = \lambda_i$ and $\text{cov}(t_i, t_j) = 0 \ (i \neq j)$. Where $\text{var}(\cdot)$ is an operator for variance, and $\text{cov}(\cdot,\cdot)$ is an operator for covariance. If k PCs are selected via CPV criteria based on the cumulative variance contribution rate of the first several PCs, the following PCA model can be formulated.

$$\begin{aligned} \hat{\mathbf{X}} &= \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \cdots + \mathbf{t}_k\mathbf{p}_k^T = \mathbf{T}_k\mathbf{P}_k^T \\ \mathbf{E} &= \mathbf{t}_{k+1}\mathbf{p}_{k+1}^T + \mathbf{t}_{k+2}\mathbf{p}_{k+2}^T + \cdots + \mathbf{t}_m\mathbf{p}_m^T \end{aligned} \tag{2}$$

Where $\hat{\mathbf{X}}$ is an estimator of $\mathbf{X}$, and $\mathbf{E}$ is a corresponding residual error matrix.

Zhenhua Guo is with the School of Mechanical Science and Engineering, HUST, Wuhan, Hubei, P.R.China

Hong Yue and Hong Wang are with Control Systems Centre, UMIST, Manchester, UK (Affiliated with Inst. Auto., Beijing, P R China) hong.wang@umist.ac.uk

B. Least Mean Squared Error and Minimum Error Entropy

Consider a random vector $\mathbf{x} \in \mathbb{R}^m$ with zero mean. Least mean squared error criterion states that the best estimator $\hat{\mathbf{x}}$ for $\mathbf{x}$ is the one which minimizes *mean squared error* (MSE). This means that the best estimator $\hat{\mathbf{x}}$ is subject to $\min(\text{var}(\tilde{\mathbf{x}}))$. where $\tilde{\mathbf{x}} = \mathbf{x} - \hat{\mathbf{x}}$ is an estimated error vector. Under the assumption of m-dimensional Gaussian distribution for $\tilde{\mathbf{x}}$, the entropy $H(\tilde{\mathbf{x}})$ of $\tilde{\mathbf{x}}$ has the same monotony with determinant $\det(\text{var}(\tilde{\mathbf{x}}))$ [4]. That is

$$\min\left(\text{var}(\tilde{\mathbf{x}})\right) \Leftrightarrow \min\left(\det(\text{var}(\tilde{\mathbf{x}}))\right) \tag{3}$$

III. PCA BASED ON MINIMUM ERROR ENTROPY

Form "(3)" it can be seen that minimizing the mean squared error for Gaussian system is equivalent to the minimization of the error entropy. But entropy has more general meaning than that of variance for arbitrary random variables, it can be used to form a design criteria for general stochastic system subjected to arbitrary distribution [5]. Based on this idea a modified PCA technique with the optimization of error entropy is proposed.

To be more specific, consider an n-by-m data matrix $\mathbf{X}$ from a stochastic system that consists of n observations with m variables. From the PCA model in "(2)" the error matrix can be expressed as follows.

$$\mathbf{E} = [E_1, \cdots, E_m] = \begin{bmatrix} e_{11} & e_{12} & \cdots & e_{1m} \\ e_{21} & e_{22} & \cdots & e_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ e_{n1} & e_{n2} & \cdots & e_{nm} \end{bmatrix} \tag{4}$$

In "(2)" we only consider the total variances contributed by PCs, but don't consider the ensemble minimum error entropy of the model. Under the assumption of the independence among the errors, the probability distribution matrix of error matrix $\mathbf{E}$ and the corresponding error entropy can be obtained by statistical histogram principle.

$$Prob(\mathbf{E}) = \begin{bmatrix} \eta_{11} & \eta_{12} & \cdots & \eta_{1m} \\ \eta_{21} & \eta_{22} & \cdots & \eta_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \eta_{N1} & \eta_{N2} & \cdots & \eta_{Nm} \end{bmatrix} \tag{5}$$

$$H(\mathbf{E}) = \sum_{i=1}^{m} H(E_i) = \sum_{i=1}^{m}\sum_{j=1}^{N} \eta_{ij}\log\left(\eta_{ij}\right) \tag{6}$$

where N is a uniformly quantizing division number of error. η_{ij} is the probability of error vector E_j in its ith division. That is, for each error vector $E_i = [e_{1i},\ e_{2i},\ \cdots,\ e_{ni}]^T$ in "(4)", its elements are binned in $[\min(E_i),\ \max(E_i)]$ by N divisions. The error entropy calculated in "(6)" is only based on the conventional PCA model without any

optimization. In order to formulate the best model with minimum error entropy, the reserved loading eigenvectors in $\mathbf{P}_k \in \mathbb{R}^{m\times 1}$ in "(2)" should be optimized via a genetic algorithm [3]. For the convenience of chromosome coding in the genetic algorithm, the reserved loading eigenvectors in $\mathbf{P}_k$ are rearranged and expanded as a value vector $\mathbf{g}$ for the optimization.

$$\begin{bmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \vdots \\ \mathbf{p}_k \end{bmatrix} \Longleftrightarrow \mathbf{g} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{m\times k} \end{bmatrix} \tag{7}$$

where there is a mapping of elements between $\mathbf{p}_i$ and $\mathbf{G}$ as follows.

$$\mathbf{p}_i = \begin{bmatrix} p_{1i} \\ p_{2i} \\ \vdots \\ p_{mi} \end{bmatrix} \Longleftrightarrow \begin{bmatrix} g_{(i-1)\times m+1} \\ g_{(i-1)\times m+2} \\ \cdots \\ g_{i\times m} \end{bmatrix} \tag{8}$$

Since $\mathbf{g}$ is a deterministic value vector, it cannot be optimized directly. In order to further optimize the error entropy, an optimization model with a random vector $\mathbf{u} = [u_1, u_2, u_{m\times k}]$ pertinent to the value vector $\mathbf{g}$ in "(7)" can be built. The difference between $\mathbf{u}$ and $\mathbf{g}$ is that the elements in $\mathbf{g}$ are deterministic values, while the elements in $\mathbf{u}$ are variables. The random vector $\mathbf{u}$ can also be uniquely converted to a random loading eigenvector matrix $\mathbf{V} = [\mathbf{v}_1 \mathbf{v}_2 \cdots \mathbf{v}_k]$. Each column vector $\mathbf{v}_i$ in $\mathbf{V}$ forms a random vector akin the loading eigenvector $\mathbf{p}_i$. Consequently, the following optimization model for optimization can be formulated.

$$\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{m\times k} \end{bmatrix} \Longleftrightarrow \mathbf{V} = \begin{bmatrix} v_{11} & \cdots & v_{1k} \\ v_{21} & \cdots & v_{2k} \\ \cdots & \cdots & \cdots \\ v_{m1} & \cdots & v_{mk} \end{bmatrix} \tag{9}$$

The mapping of elements in the vector $\mathbf{u}$ to those in the matrix $\mathbf{V}$ is similar to the relationship in "(8)".

In order to optimize the random variables in $\mathbf{u}$ via the genetic algorithm, the domains of the random variables shall be decided. For the reason that the random variables are used to determinate the direction of loading eigenvectors, their values are proportional to each other. Form "(7)" the domain for the random vector $\mathbf{u}$ in "(9)" can be set as $[0.5\mathbf{g}, 1.5\mathbf{g}]$, i.e. $u_i \in [0.5g_i, 1.5g_i]$. In the meantime the loading eigenvectors in $\mathbf{V}$ shall also be subjected to

$$\mathbf{v}_i^T \mathbf{v}_i = 1 \quad \text{and} \quad \mathbf{v}_i^T \mathbf{v}_j = 0 \ (i \neq j;\ 1 \leqslant i, j \leqslant k) \tag{10}$$

Based on the results discussed above, the following calculating procedure for the proposed method can be got.

(i) Let $\mathbf{X} \in R^{n\times m}$ be a data matrix with n observations and m variables. Perform a conventional PCA on $\mathbf{X}$ and obtain the PCA model as "(2)";

(ii) Rearrange the loading matrix according to (7) and form the value vector $\mathbf{g}$, which can be used to define the domain of the random vector $\mathbf{u}$ in (iii);

(iii) Establish an optimizing model with a random vector $\mathbf{u}$ and define its domain via the value vector $\mathbf{g}$ in (ii);

(iv) Optimize the random vector via the genetic algorithm with the performance of the minimum error entropy, which can be calculated by "(4)", "(5)" and "(6)".

IV. A SIMULATION EXAMPLE

Let's consider a pseudo stochastic system with a random state vector $\mathbf{x} = [x_1,\ x_2,\ x_3]$, where $x_1 = \sin(\frac{\pi}{2} \times rand(1))$; $x_2 = x_1 + 0.5 \times rand(1)$; $x_3 = x_2 - 0.5 \times x_1$; $rand(1)$ chooses number uniformly distributed on the interval $(0.0,\ 1.0)$.

Given that we have 500 observations of $\mathbf{x}$. After simulation with the proposed method the "Fig. 1" gives the results of the decrease of error entropy and undulation of mean square error of the PCA model before and after the optimization via the genetic algorithm.

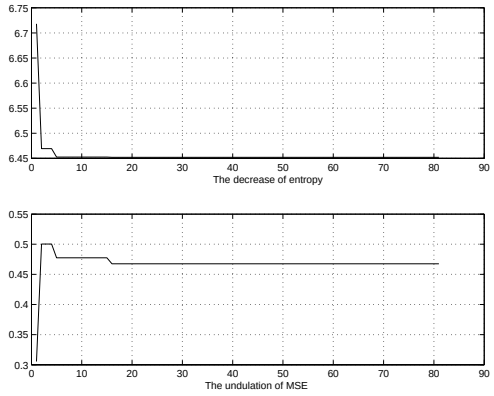

Fig. 1. The decrease of error entropy and the undulation of MSE

V. ACKNOWLEDGMENTS

The authors acknowledge the support through grants from the UK Leverhulme Trust (F/00038/D) and the NSFC of P R China (No. 60128303/60274020).

REFERENCES

[1] I. T. Jolliffe, *Principal Component Analysis*, Springer, New York, 2nd edition, 2002.

[2] S. Valle, W. Li, and S. J. Qin, "Selection of the Number of Principal Components: The Variance of the Reconstruction Error Criterion with a Comparison to Other Methods", *Ind. Eng. Chem. Res.*, 38, 1999, pp. 4389-4401.

[3] L. D. Davis, *Handbook of Genetic Algorithm*, Van Nostrand Reinhold, 1991.

[4] A. Hyvärinen , J. Karhunen and E. Oja, *Independent Component Analysis*, John Wiley & Sons, Inc., New York, 2001.

[5] H. Wang , "Minimum Entropy Control of Non-Gaussian Dynamic Stochastic System", *IEEE Transactions on Automatic Control*, Vol.47, Freb. 2002, pp. 398-403.