POWER QUALITY DATA COMPRESSION USING PRINCIPAL COMPONENT ANALYSIS

Huaying ZHANG, Zhengguo ZHU, Senjing YAO
Shenzhen Power Supply Co. Ltd., China Southern Power Grid – P. R. China
E-mail address: zhyszpower@163.com

Bingbing ZHAO, Junwei CAO
Tsinghua University – P. R. China
E-mail address: jcao@tsinghua.edu.cn

ABSTRACT

With the increasing of non-linear, burst or un-balanced load, power quality issues in the grid is becoming important. With more power quality monitors installed with higher sampling rates, an expanded size of power quality data brings difficulty to storage, transmission and analysis. In this paper, principal component analysis (PCA), which is a popular feature extraction algorithm in pattern recognition, is applied to power quality event data compression. In a power grid, different nodes and phases normally have high correlations, and PCA projects original data to a lower dimensional space to reduce redundancy. The compression ratio is determined by the number of principal components. With more principal components, the error of data recovery is reduced. We also compare the performance of two derivative algorithms, probabilistic PCA (PPCA) and kernel PCA (KPCA). Experimental results show smaller errors with higher complexity, comparing PPCA and KPCA with PCA.

INTRODUCTION

Ideal signals of a power grid should be perfect sine waves with constant frequency. In three phases AC, the voltage and current of each phase are expected to be symmetric. However, with the increasing of non-linear, burst or un-balanced load, power quality issues in the grid is becoming important. Nowadays, power quality problems can have a threat to safety and stability of the power grid, resulting in huge financial loss [1].

Since some power quality issues are transient, e.g. voltage sags, data sampling for power quality monitoring requires high frequency, resulting in large amount of data. For example, the city Shenzhen of P. R. China has deployed over 600 power quality monitoring nodes for several years [2]. Management and analysis of such big data is becoming a challenging issue. The large volume of power quality monitoring data brings difficulty to storing, transmitting, querying and data mining. Applications of such big data for advanced analysis are indeed required, though currently with very high overhead.

More data may not show more information, since correlations may bring redundancy. In this work, features of power quality monitoring data are further investigated. We believe there are high correlations among different channels and nodes of power quality monitoring data, which make data compression or feature extraction possible. For example, power quality events, happening simultaneously among the three phases, are highly correlated. A highly efficient data compression method should be able to reduce overhead for data storage and analysis.

Data compression is not a new research topic, and many approaches have been investigated to reduce data sizes by improve coding scheme such as Huffman coding. In recent years, many methods have been proposed specially to compress power quality data using wavelet and wavelet packet transforms [3-6]. In [3] and [4], compression is achieved by thresholding wavelet transform coefficients and reconstructing the signal using significant coefficients. Using the same transform thresholding techniques, variations of the wavelets, such as slantlets, are used in [5]. In [6], minimum description length criterion is used for compressing with wavelet packets.

In this work, principal component analysis (PCA) [7] is adopted for power quality event data compression. PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, so-called principal components.

In the second section, we will give a brief introduction to PCA, and its derivative algorithms, PPCA (Probabilistic Principal Component Analysis) [8] and KPCA (Kernel Principal Component Analysis) [9]. PPCA can better keep features of sample data, instead of simply removing non-principal components, which improve performance on succeed recognition and classification. KPCA map the data to Hilbert space by using kernel function, which makes it easier to extract principal components. In the third section, we will compress real PQ data using these three algorithms with reconstruction. Performance metrics include *compression ratio* and *recovery error*. The paper is concluded in the fourth section with an introduction to future research directions.

PCA, KPCA AND PPCA

PCA

Suppose that we have N samples of n -dimension vector x , and each row is a sample, column is $x_1, x_2, \cdots x_m$. We wish to reduce the dimension from n to m . Principal component analysis completes this by finding linear combinations, $a_1x_1, a_2x_2, \cdots a_mx_m$, called principal components, which have maximum variance, and subject to being uncorrelated with previous principal components. The PCA tries to reduce dimensions of data considerably while still retaining much of the information in it. **Figure 2.1** shows that how PCA works. The principal component orientation (also the signal) has maximum variance compared to non-principal component (also the noise).

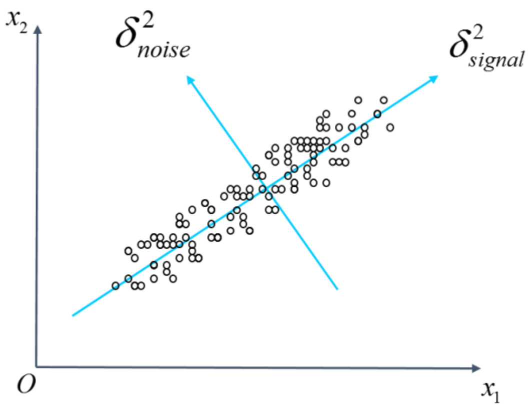


Figure 2.1. An illustration of PCA

Specific steps of PCA are derived as follows:

Normalize the sample data by:

$$x_i(t) = \frac{x_i(t) - \mu(x_i)}{\sqrt{\delta^2(x_i)}}$$

where:

$\mu(x_i)$ is the mean of x_i

$\delta(x_i)$ is the standard deviation of x_i

Compute the covariance matrix of sample data after normalization:

$$XX^T$$

Compute the eigenvalues and eigenvectors of XX^T and sort all eigenvalues. Select the corresponding eigenvectors the biggest m eigenvalues as principal component orientation.

$$\alpha^1, \alpha^2, \cdots \alpha^m$$

Then we compute the projection of sample data on principal component (also the compressed data):

$$y(t) = [\alpha^1, \alpha^2, \cdots \alpha^m]^T X$$

PPCA

Compared with traditional PCA, PPCA get more information from non-principal components instead of simply discarding it. PPCA suppose the non-principal components as noise which is subject to Gaussian distribution. By using maximum likelihood estimation (MLE), we can get the parameters of the distribution.

Specific steps of PPCA is derived as follows:

Each sample data is a n-dimension vector, and there is a m-dimension($m < n$) vector t which satisfies

$$t = Wx + \mu + \varepsilon$$

where:

W is a $n \times m$ matrix

μ is the sample mean

ε is noise

Hidden variable x is subject to Gaussian distribution $x \sim N(0, I)$

$\mu = \frac{1}{N}\sum_i^N x_i$, $\varepsilon \sim N(0, \psi)$ is the diagonal covariance matrix.

By further derivation, the MLE of PPCA is

$$\sigma^2 = \frac{1}{d-q}\sum_{j=q+1}^{d} \lambda_j$$

$$W = U_q(\Lambda_q - \sigma^2 I)^{1/2} I$$

where:

$\lambda_i (i = m+1, m+2, \cdots n)$ is the $d-q$ smallest eigenvalues of sample covariance matrix

$S = \frac{1}{N}\sum_{i=1}^{N}(x_i - u)(x_i - u)^T$

$\Lambda_q = diag(\lambda_{q+1}, \lambda_{q+2}, \cdots \lambda_d)$

U_q is a matrix composed of eigenvectors corresponding $\lambda_i (i = m+1, m+2, \cdots n)$

I is a ($d-q$)-dimension identity matrix

The data after compression x is

$$t = W^T(x - \mu)$$

KPCA

There are many algorithms using the Kernel method in pattern recognition such as support vector machine (SVM). Because the relationship may be not clear in lower dimensional space, the kernel converts it into higher dimensional space which is called feature mapping. In the new space, extract principal components or classification is possible.

Kernel methods owe their names to the use of kernel functions, which enable them to operate in a high-dimensional implicit feature space without ever computing the coordinates of the data in that space. Most popular kernel functions are linear kernel, Gaussian kernel, sigmoid kernel, etc.

In the high-dimensional space, we define a mapping $\phi: X_d \to F$ which convert a d-dimension vector to Hilbert space. In this space, the covariance matrix is:

$$K = \frac{1}{N}\phi(X)\phi(X)^T$$

Then compute the eigenvalues and eigenvectors of K which are:

$$\lambda^1, \lambda^2, \cdots \lambda^m, \alpha^1, \alpha^2, \cdots \alpha^m$$

and the largest p eigenvalues are:

$$\lambda^1, \lambda^2, \cdots \lambda^p$$

Then normalize corresponding eigenvalues which makes:

$$\| \alpha^i \|^2 = \frac{1}{\lambda^i}$$

Then we can compute the projection on principal components:

$$\phi^t(X)V_j = \sum_{i=1}^{m} \alpha_i^{\,j} \phi^t(X)\phi(X_i)$$

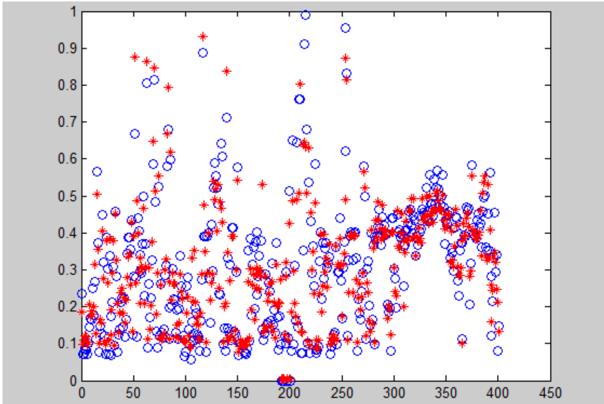
EXPERIMENTAL RESULTS

Experiment data of this paper is the power quality events record of Shenzhen city from 2010-2012. **Table 3.1** is an example. It shows the high correlation between three phases in voltage sag depth data.

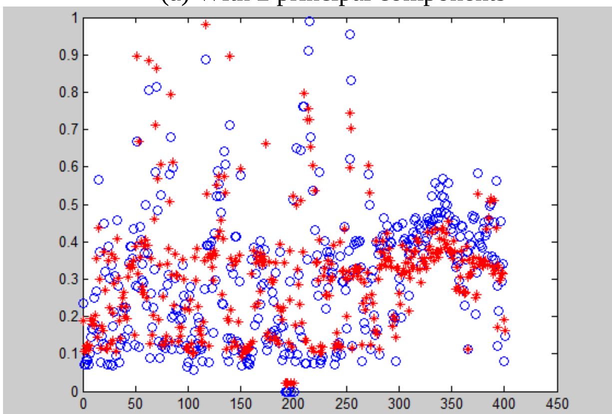
Phase	A	B	C
Fuyong Station	0.305	0.313	0.328
Qingshui Station	0.813	0.816	0.803
Shuitian Station	0.782	0.756	0.778
Yuxin Station	0.424	0.43	0.421
Tangwei Station	0.25	0.235	0.227

Table 3.1. Co-relation between three phases in voltage sag depth data

Experimental results are included in **Figure 3.1**.The x-coordinate shows that a total of 400 samples are utilized in this work. The y-coordinate is the voltage sag depth of the samples. Original data are presented in blue circles and recovered data after data compression are presented in red stars.



(a) With 2 principal components



(b) With 1 principal component

Figure 3.1. Comparison of original and recovered data on voltage sag depth

Regarding data compression, two most essential performance metrics are *compression ratio* and *recovery error*. From **Figure 3.1**, it is illustrated that original data compressed with 2 principal components can be recovered with less errors, compared that only 1 principal component is utilized. These are also investigated quantitatively.

Firstly consider the compression ratio. Suppose that the sample data are composed of N n-dimensional vectors, and we extract m principal components. The following data need to be stored:

- the projections of N vectors on m principal components, $m \times N$ in all
- m principal component vectors, $m \times n$ in all
- sample means, n in all

Therefore, the compression ratio (CR) is:

$$CR = \frac{m \times N + m \times n + n}{n \times N}$$

Then consider the recovery error, which measures the difference of original data and recovery data.

$$error = \frac{\sum_{t=1}^{N} \left(\tilde{x}(t) - \hat{x}(t)\right)^T \left(\tilde{x}(t) - \hat{x}(t)\right)}{\sum_{t=1}^{N} \tilde{x}^T(t)\tilde{x}(t)}$$

Table 3.2 compares the compression ratio and recovery error with 1 or 2 principal components. The compression rate with only 1 principal component can reach 33.8% with the given dataset, but the recovery error is as high as 8.15%. With 2 principal components applied, the recovery error is reduced to 2.87%, though the compression rate is increased to 67.4%. In actual applications, there is always tradeoff to be made between these two performance metrics.

Number of principal components	1	2
Compression ratio	0.338	0.674
Recovery error	8.15%	2.87%

Table 3.2. Compression ratio and recovery error with 1 or 2 principal components

Table 3.3 compares the performance of PCA, PPCA, and KPCA. Since compression ratios of these three methods are all the same, therefore, just recovery errors are listed. As we can see, the recovery errors of PPCA and KPCA are only 5.61% and 5.82% with 1 component, which are significantly smaller than that of PCA. This proves that PPCA and KPCA can indeed extract more information with same data size. PPCA achieves this by estimation of noise and KPCA by using kernel functions.

Number of principal components	1	2
PCA	8.15%	2.87%
PPCA	5.61%	1.06%
KPCA	5.82%	1.22%

Table 3.3. Recovery errors of PCA, PPCA and KPCA with 1 or 2 principal components

In this paper, we just implement PCA on data between different phases. In fact, many power quality events are caused by other events spread in grid, leading to higher correlation. At the same time, different physical quantity, e.g. voltage, current, active power, reactive power, are also correlated, making data compression by PCA is more feasible.

CONCLUSIONS

In this paper, PCA is applied for compression of data pertaining to the Power Quality events. Unlike compression algorithms using wavelet, PCA is simple and easy to put into practice. The compression performance is evaluated by compression ratios and errors between original data and recovery data. We also use PPCA and KPCA, which are derivative algorithms of PCA, to implement data compression. Performance of these methods is compared, and we find that PPCA and KPCA both have smaller errors than PCA but with higher complexity.

Future research directions include the application of the methods presented in this paper to more data, e.g. power experience [10], coupling of data compression with data cleaning [11], power quality data optimization [12], and advanced data analysis on power quality events [13]. All these work will be carried out using real power quality data collected in the city Shenzhen of P. R. China.

Acknowledgements

This work is supported in part by National Natural Science Foundation of China (grants No. 61472200 and No. 61233016), Ministry of Science and Technology of China under National 973 Basic Research Program (grant No. 2013CB228206), and China Southern Power Grid Co., Ltd (K-SZ2012-026).

REFERENCES

[1] T. Gao, J. Cao, Y. Xu, H. Zhang, P. Yu, and S. Yao, 2013, "From Power Quality to Power Experience", *Proceedings of 4th International Conference on Networking and Distributed Computing*, Hongkong, China, 116-120

[2] Z. Hu, Y. Han, S. Yao, Z. Huang, J. Cao, H. Zhang, P. Yu, S. Shi, and M. Wang, 2014, "Comprehensive Control of Power Quality in a City Power Grid", *High Voltage Engineering*, vol. 40, supplement 2, 9-19

[3] S. Santoso, E. J. Powers, & W. M. Grady, 1997, "Power Quality Disturbance Data Compression using Wavelet Transform Methods", *IEEE Transactions on Power Delivery*, vol. 12, no. 3, 1250-1257.

[4] T. B. Littler, D. J. Morrow, 1999, "Wavelets for the Analysis and Compression of Power System Disturbances", *IEEE Transactions on Power Delivery*, vol. 14, no. 2, 358-364

[5] G. Panda, P. K. Dash, A. K. Pradhan, S. K. Meher, 2002, "Data Compression of Power Quality Events

using the Slantlet Transform", *IEEE Transactions on Power Delivery*, vol. 17, no. 2, 662-667

[6] E. Y. Hamid, Z. I. Kawasaki, 2002, "Wavelet-based Data Compression of Power System Disturbances using the Minimum Description Length Criterion", *IEEE Transactions on Power Delivery*, vol. 17, no. 2, 460-466

[7] S. Wold, 1987, "Principal Component Analysis", *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1-3, 37-52

[8] M. E. Tipping, C. M. Bishop, 1999, "Probabilistic Principal Component Analysis", *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, no. 3, 611-622

[9] B. Schölkopf, A. Smola, K. R. Müller, 1997, "Kernel Principal Component Analysis", *Proceedings of International Conference on Artificial Neural Networks*, 583-588

[10] T. Gao, J. Cao, Y. Xu, H. Zhang, P. Yu, and S. Yao, 2013, "From Power Quality to Power Experience", *Proceedings of 4th International Conference on Networking and Distributed Computing*, Hongkong, China, 116-120

[11] Z. Yang, J. Cao, Y. Xu, H. Zhang, P. Yu, and S. Yao, 2013, "Data Cleaning for Power Quality Monitoring", *Proceedings of 4th International Conference on Networking and Distributed Computing*, Hongkong, China, 111-115

[12] Y. Wan, J. Cao, H. Zhang, Z. Zhu, and S. Yao, 2014, "Optimization of the Power Quality Monitor Number in Smart Grid", *Proceedings of 5th IEEE International Conference on Smart Grid Communications*, Venice, Italy, 236-241

[13] Z. Huang, T. Gao, H. Zhang, X. Han, J. Cao, Z. Hu, S. Yao, and Z. Zhu, 2014, "Transient Power Quality Assessment based on Big Data Analysis", *Proceedings of 2014 China International Conference on Electricity Distribution*, Shenzhen, China, 1308-1312, 2014.