**MODULE – 5 LECTURE NOTES – 5**

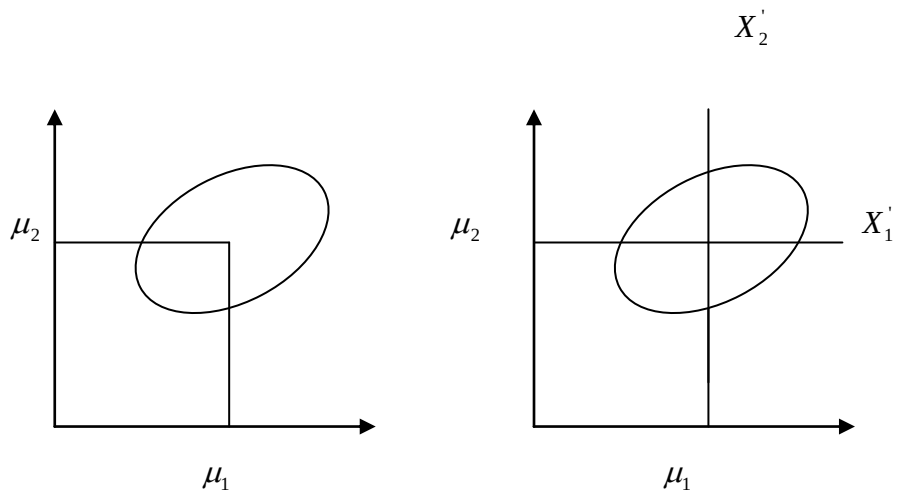
**PRINCIPAL COMPONENT ANALYSIS**

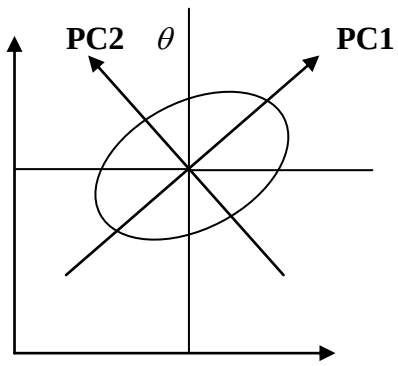
## 1. Principal Component Analysis (PCA)

Principal component analysis (PCA), also known as Karhunen-Loeve analysis, transforms the information inherent in multispectral remotely sensed data into new principal component images that are more interpretable than the original data. It compresses the information content of a number of bands into a few principal component images. This enables dimensionality reduction of hyperspectral data. Generally within a multispectral imagery, the adjacent bands will depict mutual correlation. For example, if a sensor captures information using visible/near infrared wavelengths, the vegetated areas obtained using both the bands will be negatively correlated in nature. Imagine a multispectral or a hyperspectral imagery with more than 2 bands which are inter-correlated. The inter correlations between bands depicts repetition of information between the adjacent bands.

Consider two variables x and y that are mutually correlated and which are plotted using a scatter diagram. The relationship between x and y can be very well represented using a straight line sloping upwards towards right (assuming that x and y are positively correlated). Now suppose that x and y are not perfectly correlated and that there exists a variability along some other axis. Then the dominant direction of variability can be chosen as the major axis while another second minor axes can be drawn at right angles to it. A plot with both these major and minor axis may be a better representation of the x-y structure than the original horizontal and vertical axes. Using this background information, assume that the pixel values of two bands of Thematic Mapper are drawn using a scatter plot. Let $X_1$ and $X_2$ denote the respective bands and let $\mu_1$ and $\mu_2$ represent their corresponding mean values. The spread of points (pixel values) indicates the correlation and hence the quality of information present in both the two bands. If the points are tightly clustered within a two dimensional space, it means that they would provide very little information to the end user. It means that the original axis of X1 and X2 might not be a very good representative of the 2D feature space in order to analyze the information content associated with these two bands. Principal component analysis can be used to rotate the location of original axes so that the original

brightness values ( pixel values) be redistributed or reprojected onto a new set of axes (principal axis). For example, the new coordinate system (with locations of $X_1^{'}, X_2^{'}$) can be obtained by a simple translation of $X_1^{'} = X_1 - \mu_1$ and $X_2^{'} = X_2 - \mu_2$. Once this translation is accomplished, the new coordinates can be rotated about the new origin in the new coordinate system by say some angle $\theta$. Now the first axis of $X_1^{'}$ will be associated with the maximum amount of variance which is now called as the first principal component (PC1). And the second principal component is orthogonal to PC1. Similarly, the third, fourth and subsequent components can be arrived at which will be arranged in the decreasing amount of variance found in the data set.



In order to arrive at the principal axes, certain transformation coefficients need to be obtained which can be applied to the original pixel values. The steps for this transformation are discussed below:

a) Compute the covariance matrix of the n dimensional remotely sensed data set.

The importance of variance to define the points represented by a scatter plot along the dominant direction has already been stressed. If variance be used to define the shape of the ellipsoid covering the points ( in an n dimensional variable space) then, the scales used for measuring each variable must be comparable with one another. If not, neither the variance of each variable be the same nor will the shape of enclosing ellipse remain same. This may also create further complications as the shape of one ellipsoid cannot be related mathematically to the shape of the second ellipsoid. In these circumstances, the correlation coefficient can be used rather than the covariance to measure standardized variables. To standardize the variables, the mean value can be subtracted from all measurements and then the result can be divided by their standard deviation which would convert the raw values to z scores or standard scores having zero mean and a variance of unity. It should be noted that usage of covariance matrix will yield unstandardized PCA and use of correlation matrix will yield in a standardized PCA.

b) Computation of eigenvalues and eigenvectors

Consider that there are n number of bands within a multispectral remotely sensed imagery. For these n bands there will be n rows and n columns. Quantities known as eigenvalues can be found for the chosen matrix. Eigenvalues are proportional to the length of principal axes of the ellipsoid whose units are measured using variance. In order that the variables be measured on comparable scales, standardized units of variance must be used, as stated in previous paragraph. Each eigenvalue will be associated with a set of coordinates which are known as eigenvectors. The eigenvalues and eigenvectors will together describe the lengths and directions of the principal axes. The eigenvalues will contain important information such as the total percent of variance explained by each of the principal components using the expression

$$TotalVariance(\%) = \frac{Eigenvalue}{\sum_{i=1}^{n} Eigenvalue} * 100$$

If $\lambda_{i,j}$ represents the eigenvalues of an nxn covariance matrix which can be represented as:

$$EV\ Cov\ EV^{T} = \begin{pmatrix} \lambda_{1,1} & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda_{2,2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda_{3,3} & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_{4,4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda_{5,5} & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda_{n,n} \end{pmatrix}$$

c) Estimation of factor loadings

The eigenvectors when scaled using the square roots of their corresponding eigenvalues can be interpreted as correlations between the principal components and the individual bands of the image. The correlation of each band with respect to each of the principal components can be computed. This gives us an idea regarding how each band 'loads' or otherwise is associated with respect to each principal component. The expression can be given as:

$$R_{kp} = \frac{a_{kp} * \sqrt{\lambda_p}}{\sqrt{Var_k}}$$

where $a_{kp}$ = Eigenvector for band k and component p

$\lambda_p$ = pth eigenvalue

$Var_k$ = Variance of band k in the covariance matrix

This results in factor loadings.

**Numerical Example**

PCA is based on four assumptions namely, linearity, sufficiency of mean and variance, orthogonality of principal components and that large variances have important dynamics.

The second assumption states are the mean and variance is used as sufficient statistics to fully define the probability distribution. For this assumption to be true, the probability distribution of the variable considered must be exponential in nature. This guarantees that the signal to noise ratio together with the covariance matrix is sufficient to fully describe the noise and redundancies. The third assumption indicates that the data has a high signal to noise ratio. And hence, the principal components with a larger variance will represent more dynamics than those with lower variances which will depict noise. PCA can be solved using linear algebra decomposition techniques.

Assume a hypothetical situation of an image at row1 and column1 for seven bands of a satellite sensor which are represented using a vector X such that,

$$X = \begin{bmatrix} BV_{1,1,1} = 20 \\ BV_{1,1,2} = 30 \\ BV_{1,1,3} = 22 \\ BV_{1,1,4} = 60 \\ BV_{1,1,5} = 70 \\ BV_{1,1,7} = 62 \\ BV_{1,1,6} = 50 \end{bmatrix}$$

We will now apply the appropriate transformation to this data such that it is projected onto the first principal component's axes. In this way we will find out what the new brightness value will be, for this component. It is computed using the formula:

$$newBV_{i,j,p} = \sum_{k=1}^{n} a_{kp} BV_{i,j,k}$$

Where $a_{kp}$ = eigenvectors, $BV_{ijk}$ =brightness value in band k for the pixel at row i, column j and n = number of bands. In our hypothetical example, this yields,

$$newBV_{1,1,1} = a_{1,1}(BV_{1,1,1}) + a_{2,1}(BV_{1,1,2}) + a_{3,1}(BV_{1,1,3}) + a_{4,1}(BV_{1,1,4}) + a_{5,1}(BV_{1,1,5}) + a_{6,1}(BV_{1,1,7}) + a_{7,1}(BV_{1,1,6})$$

$$= 0.205\ (20) + 0.127(30) + 0.204\ (22) + 0.443\ (60) + 0.742(70) + 0.376(62) + 0.106(50)$$

$$= 119.53$$

This pseudomeasurement is a linear combindation of original brightness value and factor scores (eigenvectors). The new brightness value for row 1, column 1 in principal component 1 after truncation to an integer is 119. This procedure takes place for every pixel in the original image data to produce the principal component 1 image dataset. Then, p is incremented by 1 and principal component 2 is created pixel by pixel

## 2. Noise Adjusted PCA

The presence of noise in any data set should be low, else no matter what the analysis technique, the information content extracted will be a minimal. Noise dominates signal in lower- order principal component images. Hence, the PCA needs to be adjusted for the noise variance. Once the PCA images have been adjusted for noise, the result can be used to generate PCA images that are unaffected by noise. Noise is relatively expressed with respect to the measurement in terms of signal to noise ratio given by the expression:

$$SNR = \frac{\sigma^2_{signal}}{\sigma^2_{noise}}$$

A high value of SNR indicates high precision data whereas a lower value indicates data contaminated with noise.

Principal components are linear combinations of the original variables (like image pixel values) with the coefficients being defined such that the criterion of maximum variance gets satisfied. The question which needs to be asked is whether there exists any other criterion other than that of minimum variance that can be used to estimated weights for linear combinations. In this context, a new criterion i.e., maximizing the SNR ratio can be followed. How to maximize this criterion?

How do we calculate $\sigma^2_{signal}$ and $\sigma^2_{noise}$?

A method should be devised that is capable of separating the measurements into two parts, with the first part showing the signal and the second part showing the contribution of noise. If the dataset consists of n number of bands, firstly the covariance matrix can be computed (C). Then, the horizontal and vertical pixel differences in each of these n bands can be determined. This in turn can be used to compute their covariance matrices which when combined produce the noise covariance matrix ($C_N$). The covariance matrix of the signal ($C_S$) is estimated by subtracting the covariance matrix of the noise from that of the measurement. This results in the criterion which can be written as maximizing the ratio of $C_S/C_N$. The outcome of noise adjusted PCA analysis is a linear combinations of the n spectral bands that are ranked from 1 ( having the highest signal to noise ratio) to n (having the lowest signal to noise ratio). The coefficients are applied to the data in exactly a similar manner as PCA coefficients. Hence, the principal components are estimated using the least signal to noise ratio instead of the minimum variance criterion.

**Bibliography**


1. Bastin, L. (1997) 'Comparison of fuzzy C-Means classification, linear mixture modelling and MLC probabilities as tools for unmixing coarse pixels', *International Journal of Remote sensing,* Vol.18, 3629-3648.
2. Binaghi, E., Brivio, P.A., Ghezzi, P., Rampini, A. (1999) 'A fuzzy set-based accuracy assessment of soft classification', *Pattern Recognition Letters,* Vol. 20, 935-948.
3. Congalton, R.G. and Green, K. (1998) Assessing the Accuracy of Remotely Sensed Data: Principles and Practices , Lewis Publishers, New York.
4. Deering, D. W., J. W. Rouse, R. H. Haas, and J. A. Schell, 1975,"Measuring Forage production of grazing units from Landsat MSS Data," Proceedings, 10th International Symposium on Remote Sensing of Environment, 2:1169-1178.
5. Dunn, J.C. (1973) 'A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters', *Journal of Cybernetics,* Vol. 3, 32-57.
6. Foody, G.M. (2002) 'Status of land cover classification accuracy assessment', *Remote Sensing of Environment,* Vol.80,185–201.
7. Foody, G.M. and Cox, D.P. (1994) 'Sub-pixel land cover composition estimation using a linear mixture model and fuzzy membership functions', *International Journal of Remote Sensing,* Vol.15, 619–631.
8. Groten, S. M., 1993," NDVI-Crop Monitoring and Early Warning Yield Assessment of Burkina Faso," International Journal of Remote Sensing, 14 (8):1495-1515.
9. Guyot, G. and Gu, X.-F., 1994, Effect of radiometric corrections on NDVI determined from SPOT-HRV and Landsat – TM data. Remote Sensing of Environment, 49, 169-180.
10. Huete, A., 1989, Soil influences in remotely sensed vegetation-canopy spectra. In: Asrar, G. (ed.) (1929), 107-141.
11. John R. Jensen, 1996, Introductory Digital Image Processing, Prentice Hall
12. Kauth, R. J. and G. S. Thomas, 1976,"The Tasseled Cap-A Graphic description of the spectral-temporal development of agricultural crops as seen by Landsat," Proceedings, Symposium on Machine Processing of remotely sensed data. West Lafayette, IN:Laboratory for applications of remote sensing, pp. 41-51.

13. Kauth, R. J., P. F. Lambeck, W. Richardson, G. S. Thomas, and A. P. Petland, 1979, " Feature extraction applied to agricultural crops as seen by Landsat," Proceedings, of the Technical session, LACIE Symposium. Houston: National Aeronautics and space administration, 705-721.
14. Lillesand T. M. & Kiefer R. W., 2000. *Remote Sensing and Image Interpretation, 4th ed.* Wiley & Sons.
15. Lunetta, R. S. and Elvridge, C. D. (eds.), 1998, Remote Sensing Change Detection: Environmental Monitoring, Methods and Applications. Chelsea, MI: Ann Arbor Press.
16. Paul. MK. Mather, 2004, Computer Processing of Remotely- Sensed Images, Wiley & Sons.
17. Perry, C. R., and L. F. Lautenschlager, 1984, " Functional Equivalence of spectral vegetation indices," Remote Sensing of Environment, 14:169-182.
18. Richardson, A. J. and C. L. Wiegand, 1977, " Distinguishing vegetation from soil background information," Remote sensing of environment, 8:307-312.
19. Rouse, J. W., R. H. Haas, J. A. Schell, and D. W. Deering, 1973, "Monitoring vegetation systems in the great plains with ERTS, Proceedings, 3rd ERTS Symposium, Vol. 1, pp. 48-62.
20. Sellers, P., 1989, Vegetation-canopy reflectance and biophysical properties. In: Asrar, G. (ed.) (1989), 297-335.
21. Steven, M. D., 1998, The sensitivity of the OSAVI vegetation index to observational parameters. Remote Sensing of Environment, 63, 49-60.
22. Thompson, D. R., and O. A. Wehmanen, 1980, " Using Landsat Digital Data to detect moisture stress in corn-soybean growing regions," Photogrammetric Engineering & Remote Sensing, 46:1082-1089.
23. Wang, F. (1990) 'Improving Remote Sensing Image Analysis through Fuzzy Information Representation', *Photogrammetric Engineering And Remote Sensing*, Vol. 56, 1163-1169.
24. Wu, K.L. and Yang, M.S. ( 2002) 'Alternative c-means clustering algorithms', *Pattern Recognition*, Vol.35, 2267–2278.
25. Yang, M.S., Hwang, P.Y. and Chem, D.H. (2003) 'Fuzzy clustering algorithms for mixed feature variables', *Fuzzy Sets and Systems*, Vol.141, 301–317.

26. Zadeh, L.A. (1973) 'Outline of a new approach to the analysis of complex systems and decision processes', *IEEE Transactions on systems, Man And Cybernetics*, Vol.SMC-3, No.1,28-44.
27. Zhang, J., Foody, G.M. (1998) 'A fuzzy classification of sub-urban land cover from remotely sensed imagery', *International Journal of Remote Sensing*, Vol.19, No 14, 2721-2738.