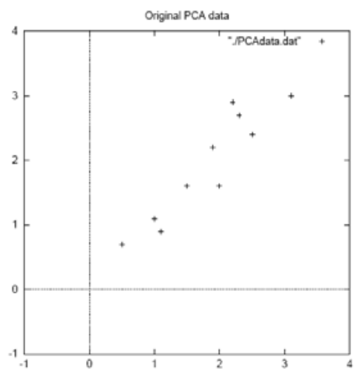
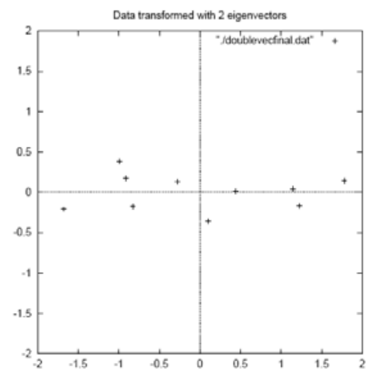
Important Terms in
Principle Component Analysis (PCA)

1. Factor analysis: The search for the “factors” (i.e. band combinations) that contain the most information.
2. Original Data: The set of brightness values for n bands and m pixels. A two-band example with m=11 data points is shown in the figure. (diagrams from LI Smith)





3. PCA: A linear method of factor analysis that uses the mathematical concepts of eigenvalues and eigenvectors. It amounts to a “rotation” of the coordinate axes to identify the Principle Components.
4. Principle Component: An optimum linear combination of band values comprising a new data layer (or image).
5. Co-variance: A measure of the redundancy of two bands (i and j), created by summing the product of the two band values over all the pixels (M).
6. Correlation: Co-variance normalized by the variances of the two bands

$$CC_{ij} = \frac{\sum_{k=1}^{M}(B_i - \overline{B}_i)(B_j - \overline{B}_j)}{\left[\sum_{k=1}^{M}(B_i - \overline{B}_i)^2 \sum_{k=1}^{M}(B_j - \overline{B}_j)^2\right]^{1/2}}$$

7. Redundant bands: Bands with a CC=1 contain the same information.
8. Correlation Matrix: a square symmetric matrix containing the correlation coefficients between every pair of bands. It contains statistical information about the data.
9. Eigenvector: The set of weights applied to band values to obtain the PC
10. Eigenvalue: A measure of the variance in a PC.
11. Axes Rotation: Multiplying the original data matrix by a matrix of eigenvectors is equivalent to rotating the data to a new coordinate system. (Note the pattern in the example above on the right)

Outline of Principle Component Analysis

1. Start with an image data set including the reflectance values from n bands with m pixels. This non-square nxm matrix will be called $\overline{\overline{D}}$. The location of the pixels does not enter this description of the data set.

2. This data set may contain "redundancies", i.e. bands whose reflectances correlate perfectly with another band. It may also contain noise. Our definition of noise is signal that does not correlate at all between bands

3. Subtract the means from each band and compute the variances and co-variances between each pair of bands. Place these values into an nxn square matrix. It is symmetric. Normalize the co-variances by the square-root of the variances to form the correlation matrix $\overline{\overline{A}}$. This is a useful matrix to study, and it forms the basis of PCA. [At this point, one could just delete bands that correlate well with other bands. This action would reduce the size of the data set. The PCA method below is more objective and systematic.]

4. Find the eigenvalues and eigenvectors of the dataset by solving

$$(\overline{\overline{A}} - \lambda)\overline{V} = 0 \qquad (1)$$

   where λ is an eigenvalue and $\overline{V}$ is an eigenvector. The word "eigen" means that these quantities are characteristics of the correlation matrix $\overline{\overline{A}}$. They reveal the hidden properties of $\overline{\overline{A}}$. Typically there will be n different solutions to (1), so there will be n paired eigenvalues and eigenvectors (i.e. λ_i .and $\overline{V}_i$). The eigenvalues will be real and positive (because $\overline{\overline{A}}$ is symmetric). We usually list these eigenvalues and eigenvectors in order of decreasing eigenvalue. That is, the first eigenvector corresponds to the largest eigenvalue.

   [In words, (1) says that for a given square matrix $\overline{\overline{A}}$, there is some vector $\overline{V}$ such that the product $\overline{\overline{A}}\,\overline{V}$ equals the product of some scalar (λ) times that same vector $\overline{V}$. We omit a discussion of how (1) is solved.]

5. The eigenvectors have a dimension equal to n, i.e. the number of original bands. The first eigenvector represents a synthetic spectrum containing the largest variance across the scene.

Ron's Rule of Thumb. "The first eigenvector usually resembles the difference in spectral signatures between the two most dominant land cover classes." For example, in a landscape composed of vegetation and bare soil, the first eigenvector will resemble the difference in the spectral signature between vegetation and soil. When this eigenvector is multiplied times the dataset matrix, it is almost like computing an NVDI layer. For a landscape composed entirely of deciduous trees and conifers, the first eigenvector will resemble the subtle difference between these two vegetation types. It would represent an "deciduous/conifer index". But these indices include all the bands, not just two.

6. The eigenvectors are orthogonal to each other, for example $\overline{V_1} \cdot \overline{V_2} = 0$. They are normally scaled so that their length is unity, that is $|\overline{V_i}| = 1$. With these two properties, the multiplying the original dataset by an eigenvector rotates the reflectance vector for a pixel.

$$\overline{\overline{D}} \cdot \overline{V_1} = \overline{P}\overline{C_1} \qquad (2)$$

where PC_1 is the first Principle Component. It is a vector with m components representing a brightness value for each pixel, i.e. it is a new single band image. Its pixel values are linear combinations of the original band values for that pixel. The weights are given by the components of the first eigenvector.

Because of our ordering of eigenvalues, this image contains the most "information" of any single image. The first eigenvalue is proportional to the brightness variance in the first PC.

To obtain the other Principle Components, we repeat (2) with the other eigenvectors so that

$$\overline{\overline{D}} \cdot \overline{V_i} = \overline{P}\overline{C_i} \qquad i = 2 \text{ to } n \qquad (3)$$

and the PC data layers can be stacked to form a new "data cube". If desired, only the first few PCs can be kept, reducing the size of the dataset. For example, if the original dataset had 200 bands (i.e. n = 200), you could keep only the ten PCs with the largest eigenvalues. This dataset is only 1/20 of the original size. The number of pixels is unchanged.

7. A remarkable property of the new data cube is that the band values are completely uncorrelated. There is no more redundancy! (Actually, the uncorrelated data in the original scene is pushed off into the high PC components.) Another property is that the bands are ordered by their "information content" (i.e. variance). Remember however that the mathematician's definition of information may not be the same as yours! You might find something important in one of the higher PCs.

A disadvantage of the PC representation is that we can no longer identify spectral signatures of objects. A "pixel profile" in the new data cube is not a spectral signature (i.e. reflectance plotted against wavelength).

References:

(Web) - A Tutorial on Principle Component Analysis by Lindsay I. Smith (Cornell University) www.cs.otago.ac.nz/cosc453/student_**tutorial**s/**principal_components**.pdf

The ENVI software manual

Introduction to Linear Algebra (Chapter on Eigenvalues)

Some textbooks on Remote Sensing have short Chapters on PCA

R.B. Smith
November, 2005