of this dimensionality problem, regularization techniques such as SVD are almost always needed to perform the covariance matrix inversion. Because it appears to be a fundamental property of hyperspectral data, however, this dimensionality issue warrants further investigation, as it seems to indicate that data representation is highly inefficient and overly sensitive to noise.

12.2 Dimensionality Reduction

From the geometric representation in Fig. 12.2, it is apparent from the linear spectral mixing concept that signal content within hyperspectral data is likely restricted to reside within a lower-dimensional subspace of the K-dimensional data space, where subspace dimensionality is dictated by the number of spectrally distinct materials in the scene. Depending on scene complexity, this dimensionality could be small or quite large. In real sensor data, however, spectral measurements are corrupted by noise, which is completely random and not restricted in the same manner. The reduced dimensionality of the information content within hyperspectral imagery can be recognized by the high degree of correlation that typically exists between spectral bands.

For example, Fig. 12.4 illustrates three bands across the full spectral range of a Hyperion satellite image over the intersection of two major highways in Beavercreek, Ohio. These bands capture some of the same spatial features of the scene but also indicate significant spectral differences. The band-to-band correlation is modest, and each band carries a substantial amount of new information not contained within the others. On the other hand, there are other band combinations, such as those depicted in Fig. 12.5, for which correlation is very high and each band carries little additional information relative to the others. To a large degree, the band combinations are redundant. This high degree of correlation, or reduced inherent dimensionality, implies that the data can be represented in a more compact manner. Transforming the data into a reduced-dimensionality representation has multiple benefits. First, it limits the amount of data needed for processing and analysis, an obvious advantage where computer memory, network bandwidth, and computational resources are concerned. Additionally, representing the data according to the primary signal components as opposed to the sensor spectral bands accentuates underlying material content, aiding visualization and analysis. Finally, transforming to a lower-dimensional subspace should provide noise reduction, as this this process filters out the noise power in the subspace that is removed.

12.2.1 Principal-component analysis

The principal-component transformation, commonly known as either principal-component analysis (PCA) (Schowengerdt, 1997) or the

Figure 12.4 Three example bands from a Hyperion VNIR/SWIR hyperspectral image over Beavercreek, Ohio, exhibiting modest spectral correlation: (a) 610 nm, (b) 1040 nm, and (c) 1660 nm.

Figure 12.5 Three example bands from a Hyperion VNIR/SWIR hyperspectral image over Beavercreek, Ohio, exhibiting high spectral correlation: (a) 509 nm, (b) 610 nm, and (c) 2184 nm.

Karhunen–Loeve transformation (Karhunen, 1947; Loeve, 1963), addresses the issue of spectral correlation and provides one basis for dealing with data dimensionality. Assume for the moment that the inherent data dimensionality is actually K, such that the covariance matrix is full rank and therefore invertible. Spectral correlation manifests by nonzero, off-diagonal elements of the covariance matrix. Suppose that there was another set of orthogonal coordinate axes in the multidimensional space for which the covariance matrix was actually diagonal. If the data were transformed into this new coordinate system, the spectral correlation between bands would be removed. This is what the PCA transform attempts to perform; it is shown graphically for a simple 2D case in Fig. 12.6, where normally distributed data are presented in the form of a scatter plot. Relative to the sensor spectral bands, the data exhibit a high degree of spectral correlation. However, correlation would be removed in this case if the bands were redefined to correspond to the principal axes of the elliptically shaped scatter distribution, denoted as $\mathbf{v}_1$ and $\mathbf{v}_2$ using dotted vectors.

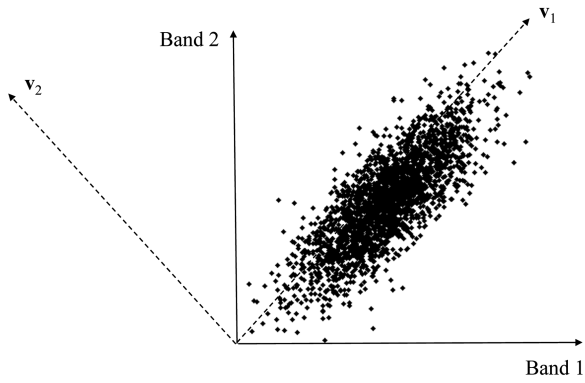

Figure 12.6 2D scatter plot of normally distributed data, illustrating the concept of redefining the bands to correspond to the principal axes in order to remove spectral correlation.

Principal-component transformation is a matter of diagonalizing a sample covariance matrix, a technique that is mathematically performed by determining its eigenvalues and eigenvectors (Strang, 1980). This begins by solving the characteristic equation

$$\det(\mathbf{C} - \sigma^2\mathbf{I}) = 0 \tag{12.16}$$

for a set of solutions $\{\sigma_j^2, j = 1, 2, \ldots, K\}$, where $\det(\mathbf{A})$ represents the determinant of matrix $\mathbf{A}$, and $\mathbf{I}$ is the $K \times K$ identity matrix. Under the full-rank assumption, K nonzero solutions exist, where each *eigenvalue* σ_j^2 represents the variance of data for a particular eigenvector direction. *Eigenvectors* $\mathbf{v}_j$ correspond to the principal directions for which the spectral correlation is removed and are computed by solving the linear system of equations

$$\mathbf{C}\mathbf{v}_j = \sigma_j^2\mathbf{v}_j \tag{12.17}$$

for their corresponding eigenvalues. Since eigenvectors can be arbitrarily scaled and still satisfy Eq. (12.17), a unitary basis is chosen, such that

$$\mathbf{v}_j^T\mathbf{v}_j = 1 \tag{12.18}$$

for all j.

Suppose that diagonal matrix $\mathbf{D}$ is formed by placing the eigenvalues along the diagonal in decreasing order, that is, from highest to lowest variance. The eigenvectors are placed in corresponding order as columns

of unitary eigenvector matrix $\mathbf{V}$. It then follows from Eq. (12.17) that

$$\mathbf{CV} = \mathbf{V}\,\mathbf{D}. \tag{12.19}$$

Since the inverse of a unitary matrix is its transpose, it follows that

$$\mathbf{C} = \mathbf{V}\,\mathbf{D}\mathbf{V}^T, \tag{12.20}$$

which indicates that the linear transformation represented by eigenvector matrix $\mathbf{V}$ diagonalizes the covariance matrix. Therefore, the principal-component transformation,

$$\mathbf{Z} = \mathbf{V}^T\mathbf{X}, \tag{12.21}$$

represents a coordinate rotation to principal-component data matrix $\mathbf{Z}$ into an orthogonal basis, such that the new principal-component bands are both uncorrelated and ordered in terms of decreasing variance.

As an example of PCA, consider again the Hyperion Beavercreek image shown in Figs. 12.4 and 12.5. Ranked eigenvalues of the sample covariance matrix are displayed on a logarithmic scale in Fig. 12.7, where it is apparent that variance in the data is predominately captured by a small set of leading principal-component directions. Figure 12.8 provides the first three eigenvectors, while the principal-component band images, corresponding to the Hyperion data, are illustrated in Fig. 12.9. These images capture primary features of the original hyperspectral image with no spectral correlation. Generally, the first principal component corresponds to the broadband intensity variation, while the next few capture the primary global spectral differences across the image. By comparing a three-band false-color composite from these three principal components with an RGB composite from the original 450-, 550-, and 650-nm bands (as illustrated in Fig. 12.10), the ability of the PCA to accentuate scene spectral differences is apparent. Statistically rare spectral features along with sensor noise dominate the low-variance, trailing principal components, three of which are illustrated in Fig. 12.11.

12.2.2 Centering and whitening

Several variations of PCA can be found in the literature and therefore warrant some discussion. The first concerns removal of the sample mean vector $\mathbf{m}$ and scaling of the diagonalized covariance matrix $\mathbf{D}$. Especially when dealing with various target detection algorithms (described in the next chapter), it is sometimes desirable to transform data into an orthogonal coordinate system centered within the data scatter as opposed

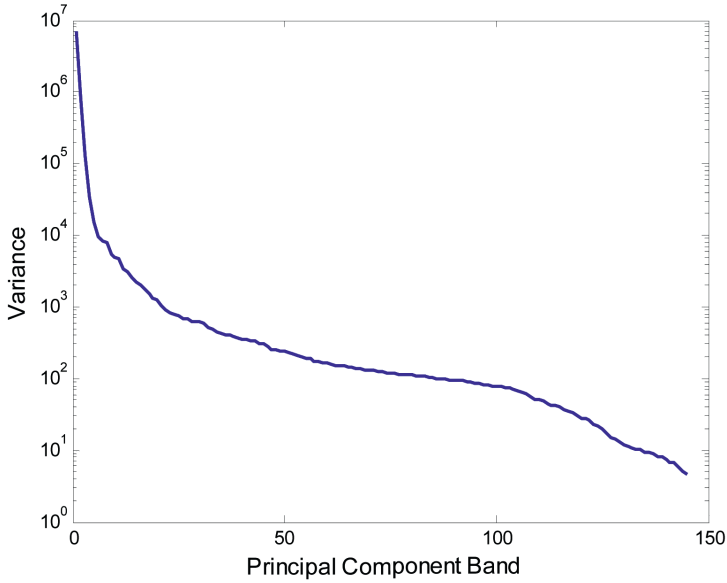

Figure 12.7 Magnitude of the ranked eigenvalues of the Hyperion Beavercreek image.

to the origin defined by the sensor data. This can be performed using the affine transformation,

$$\mathbf{Z} = \mathbf{V}^T(\mathbf{X} - \mathbf{mu}^T) \tag{12.22}$$

in place of the standard principal-component transformation given in Eq. (12.21), or equivalently, by removing the sample mean vector $\mathbf{m}$ from all original spectral vectors prior to computing and performing the principal-component transformation. This is called centering, demeaning, or mean removal. Another variation is to use the modified transformation to scale the principal-component images such that they all exhibit unit variance:

$$\mathbf{Z} = \mathbf{D}^{-1/2}\mathbf{V}^T(\mathbf{X} - \mathbf{mu}^T). \tag{12.23}$$

This is referred to as whitening the data, and $\mathbf{D}^{-1/2}$ simply refers to a diagonal matrix for which the diagonal elements are all the inverse square root of the corresponding eigenvalues in $\mathbf{D}$. To illustrate the difference between these variants of the principal-component transformation, Fig. 12.12 compares the transformed scatter plot from Fig. 12.6 in the principal component and whitened coordinate system. Data centering is performed in each case.