3.5 Empirical Orthogonal Functions

Consider an arbitrary $M \times N$ matrix $\mathbf{M}$. Suppose the matrix were representable, accurately, as the product of two vectors,

$$\mathbf{M} \approx \mathbf{a}\mathbf{b}^T,$$

where $\mathbf{a}$ was $M \times 1$, and $\mathbf{b}$ was $N \times 1$. Approximation is intended in the sense that

$$\left\|\mathbf{M} - \mathbf{a}\mathbf{b}^T\right\| < \varepsilon,$$

for some acceptably small ε. Then one could conclude that the MN elements of $\mathbf{A}$ contain only $M + N$ pieces of information contained in $\mathbf{a}, \mathbf{b}$. Such an inference has many uses, including the ability to recreate the matrix accurately from only $M + N$ numbers, to physical interpretations of the meaning of $\mathbf{a}, \mathbf{b}$. More generally, if one pair of vectors is inadequate, some small number might suffice:

$$\mathbf{M} \approx \mathbf{a}_1\mathbf{b}_1^T + \mathbf{a}_2\mathbf{b}_2 + ... + \mathbf{a}_K\mathbf{b}_K^T. \tag{3.53}$$

{eckart1}

A general mathematical approach to finding such a representation is through the SVD in a form sometimes known as the "Eckart-Young-Mirsky theorem."[84] This theorem states that the most efficient representation of a matrix in the form,

$$\mathbf{M} \approx \sum_i^K \mathbf{u}_i \lambda_i \mathbf{v}_i^T \tag{3.54}$$

{eq:53003}

where the $\mathbf{u}_i$, $\mathbf{v}_i$ are orthonormal is achieved by choosing the vectors to be the singular vectors, with λ_i providing the amplitude information.

The connection to the subject of regression analysis is readily made by noticing that the sets of singular vectors are the eigenvectors of the two matrices $\mathbf{MM}^T$, $\mathbf{M}^T\mathbf{M}$ (Eqs. 2.250, 2.251). If each row of $\mathbf{M}$ is regarded as a set of observations at a fixed coordinate, then $\mathbf{MM}^T$ is just proportional to the sample second-moment matrix of all the observations, and its eigenvectors, $\mathbf{u}_i$, are the EOFs. Alternatively, if each column is regarded as the observation set for a fixed coordinate, then $\mathbf{M}^T\mathbf{M}$ is the corresponding sample second-moment matrix, and the $\mathbf{v}_i$ are the EOFs.

A large literature provides various statistical rules for use of EOFs. For example, the rank determination in the SVD becomes a test of the statistical significance of the contribution of singular vectors to the structure of $\mathbf{M}$.[85] In the wider context, however, one is dealing with the problem of efficient relationships amongst variables known or suspected to carry mutual correlations. Because of its widespread use, this subject is plagued by multiple discovery and thus multiple jargon. In different contexts and details (e.g., how the matrix is weighted), the problem is known as that of "principal components"[86], "empirical orthogonal functions" (EOFs), the Karhunen-Loève expansion (in mathematics and electrical engineering)[87], "proper orthogonal decomposition"[88], etc. Examples of the use of EOFs will be provided in Chapter 6.