

Matrices, the singular value decomposition, and energy-selective imaging

The mathematical tools that I will use to study attenuation coefficients are based on matrix theory because the dimensionality of the space is closely related to the rank of a matrix.

To understand this, consider the experiment shown in Fig. 1. We measure the transmitted x-ray photons with a line-spectrum source with an energy-resolving detector. The source has 5 energies E_1, E_2, \dots, E_5 and the sample consists of varying amounts of 5 different materials with attenuation coefficients μ_1, \dots, μ_5 . The detector counts the number of transmitted photons in 5 energy bins that correspond to the energies of the source. To facilitate the processing, we divide each measurement by the mean number of photons with no object in the beam and then take minus the logarithm

$$L_1 = -\log\left(\frac{n_1}{n_{1,0}}\right) \quad (1)$$

From my discussion in this post (see Eq. 4)

$$n_1 = n_{1,0}e^{-T(a_1\mu_1(E_1)+\dots+a_5\mu_5(E_1))}$$

where T is the thickness of the sample. Substituting in (1),

$$L_1 = T(a_1\mu_1(E_1) + \dots a_5\mu_5(E_1)).$$

Letting $A_i = a_iT$, we can group $\mathbf{L} = [L_1, \dots, L_5]'$ and $\mathbf{A} = [A_1, \dots, A_5]'$ in vectors to form the matrix equation (the ' means transpose since I like to work with column vectors)

$$\begin{bmatrix} \mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} & \mu_{15} \\ \mu_{21} & & & & \\ & & \ddots & & \\ & & & & \mu_{55} \end{bmatrix} \begin{bmatrix} A_1 \\ \vdots \\ A_5 \end{bmatrix} = \begin{bmatrix} L_1 \\ \vdots \\ L_5 \end{bmatrix} \quad (2)$$

where $\mu_{i,j} = \mu_i(E_j)$.

Eq. 2 expresses the fundamental operation of an energy-selective x-ray imaging system. Since the μ_{ij} are assumed known beforehand, all the information about the object is carried by the A-vector. The fundamental operation of the imaging system is to calculate the A-vector from the measurement L-vector.

With a line-spectrum source, Eq. 2 is linear so we can invert the matrix of the μ 's, which I will call \mathbf{M} , multiply times \mathbf{L} and then compute \mathbf{A} from the measurements

$$\mathbf{MA} = \mathbf{L} \Rightarrow \mathbf{A} = \mathbf{M}^{-1}\mathbf{L}. \quad (3)$$

Does this mean that we can make images of as many materials as we want simply by adding more energies? In principle this is possible if the matrix \mathbf{M} has full rank. If the entries of \mathbf{M} and our measurements \mathbf{L} have infinite precision, almost all matrices will have full rank (that is, the rank is equal to the smaller of the number of rows or columns) and we can invert them. However, suppose another matrix with less than full rank is "close" to the original matrix. If the "distance" between the matrices is less than the accuracy of the

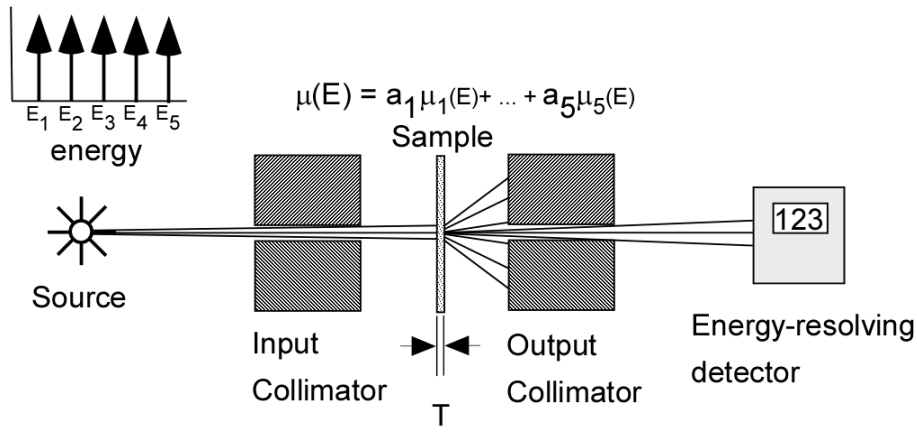


Figure 1: Measurements with line source and energy-resolving detector.

measurements, then small errors in \mathbf{L} will cause much larger errors in \mathbf{A} . In this case, the matrix \mathbf{M} is effectively not of full rank and Eq. 3 is not numerically stable.

We can study this using the concept of numerical or effective rank (Stewart 1973[1]). The theory starts by defining a matrix norm in order to measure “closeness”. The norm is a scalar function of the elements of the matrix which is a measure of its size. For effective rank, the norm is used to measure the difference between the full rank matrix and its lower rank approximating matrix.

The norm can be any function that satisfies the necessary properties of definiteness, homogeneity, and the triangle inequality[1]. A convenient norm can be defined by analogy with a vector norm. The 2-norm for a vector \mathbf{V} of length N with components v_i is

$$\|\mathbf{V}\| = \sqrt{\sum_{i=1}^N v_i^2}. \quad (4)$$

The analogous norm for a matrix is called the Frobenius norm. For a matrix \mathbf{B} with dimensions M by N , the norm is

$$\|\mathbf{B}\| = \sqrt{\sum_{i=1}^M \sum_{j=1}^N b_{ij}^2} \quad (5)$$

A matrix norm defined in this way has all the desirable properties outlined above (Stewart 1973).

The difference between infinite precision and numerical rank can be understood by considering a diagonal matrix, \mathbf{D} . The infinite precision rank for this matrix is equal to the number of non-zero entries. The effective rank can also be determined from these values. Suppose we replace the small diagonal entries with zeros. The rank of the modified matrix will be equal to the number of rows or columns minus the number of entries dropped. If the entries are in numerical order and those in columns $r + 1$ to N are set equal to zero, the difference between the limited accuracy \mathbf{D}_a and full precision matrix \mathbf{D} , is

$$\|\mathbf{D} - \mathbf{D}_a\| = \sqrt{\sum_{i=r+1}^N d_i^2} \quad (6)$$

If this distance divided by the number of entries is small compared to the error in each of the members, then the effective rank of the original matrix is equal to the rank of the reduced matrix.

Suppose the matrix of interest \mathbf{B} is not diagonal. In this case, the singular value decomposition theorem may be applied to transform this case to the diagonal matrix case just described. According to this theorem, there exist unitary matrices \mathbf{U} and \mathbf{V} such that

$$\mathbf{B} = \mathbf{U}\mathbf{D}\mathbf{V}^H, \quad (7)$$

where \mathbf{D} is a diagonal matrix and \mathbf{V}^H is the complex conjugate of the transpose of \mathbf{V} .

Since \mathbf{U} and \mathbf{V} are unitary matrices, multiplying by them does not change the value of the norm. Thus the norms of the original and the diagonal matrices are the same. Suppose the columns of the matrices are rearranged so that the diagonal elements of \mathbf{D} are in descending order. Let \mathbf{D}_a be the matrix with diagonal entries $r + 1$ to N equal to zero. The matrix

$$\mathbf{B}_a = \mathbf{U}\mathbf{D}_a\mathbf{V}^H \quad (8)$$

will have the property that no other matrix of rank r will be closer to \mathbf{B} . That is,

$$\|\mathbf{B} - \mathbf{B}_a\| = \sqrt{\sum_{i=r+1}^N d_i^2} \quad (9)$$

is minimum for all matrices of rank r . Thus, depending on the diagonal elements of the matrix \mathbf{D} , a matrix of reduced rank may be found which is closer to the original matrix than the error in the terms (with closeness measured in terms of the Frobenius norm). From an experimental point of view, the effective rank of the original matrix is equal to the reduced rank.

The singular values can be calculated by using the singular value decomposition theorem. There are extremely stable numerical methods to compute the SVD so I will use it in the next post to explore the dimensionality of attenuation coefficients.

The SVD is closely related to another tool used to study lower dimension approximations, principal components analysis (PCA). The SVD is more widely known in numerical analysis while PCA is more known in statistics. Statisticians use PCA to “reduce the dimensionality of a data set in which there are a large number of inter-related variables, while retaining as much as possible of the variation present in the data set.”[2]

Suppose the observations are arranged as rows of a matrix \mathbf{X} . PCA is implemented by computing the eigenvectors and eigenvalues of the covariance matrix of the data. These have interesting and useful properties (see Ch.1 of Jackson[3]). First, the sum of the eigenvalues is equal to the sum of the variances of the original data. Suppose we arrange the eigenvalues and the corresponding eigenvectors in decreasing order and “center” the data *i.e.* subtract the averages of the columns from the original data. Then the new variables, the principal components, formed by multiplying the eigenvectors times the columns of the centered data are uncorrelated and each represents the maximum part of the variance of the original data. That is the first principal component represents the maximum part of the variance possible with a single linear transformation of the data. The second represents the maximum part of the remaining variance and so on. Similar to the discussion above, if the distribution of eigenvalues is not uniform, as is often the case, then we can represent most of the variance of the original data by a few principal components.

We can find the relationship between the SVD and PCA as follows. Let \mathbf{X}_c be the centered data. The covariance matrix is $\mathbf{C} = \mathbf{X}_c^T \mathbf{X}_c / n$ where n is the number of observations *i.e.* rows in \mathbf{X}_c . If \mathbf{W} is the matrix of the eigenvectors of \mathbf{C} arranged along the columns and $\mathbf{\Lambda}$ is a diagonal matrix of the eigenvalues, then

$$\mathbf{C} = \mathbf{W}\mathbf{\Lambda}\mathbf{W}^T. \quad (10)$$

Using the SVD of the centered data, $\mathbf{X}_c = \mathbf{U}\mathbf{D}\mathbf{V}^T$. I have assumed real data so the Hermitian of a matrix is its transpose, $\mathbf{V}^H = \mathbf{V}^T$. Substituting in the definition of the covariance

$$\mathbf{C} = \mathbf{X}_c^T \mathbf{X}_c / n = \mathbf{V}\mathbf{D}\mathbf{U}^T \mathbf{U}\mathbf{D}\mathbf{V}^T / n \quad (11)$$

Since \mathbf{U} is unitary, $\mathbf{U}^T\mathbf{U} = \mathbf{I}$, where \mathbf{I} is the identity matrix. Therefore

$$\mathbf{C} = \mathbf{V}\mathbf{D}^2\mathbf{V}^T/n. \quad (12)$$

Comparing Eqs. 10 and 12, we can see that \mathbf{V} is the eigenvectors of the covariance and the square of the singular values divided by the number of observations is equal to the eigenvalues of the covariance.

©2011 by Aprend Technology and Robert E. Alvarez

Linking is allowed but reposting or mirroring is expressly forbidden.

Edited to add material about principal components.

References

- [1] G. W. Stewart, *Introduction to Matrix Computations*, Academic Press, 1973.
- [2] I. T. Jolliffe, *Principal Component Analysis*, New York: Springer-Verlag, 2002.
- [3] J. E. Jackson, *A User's Guide to Principal Components*, New York: John Wiley and Sons, 1991.