Mathematical Geophysics

Swarandeep Sahoo

Department of Applied Geophysics

Indian Institutes of Technology (Indian School of Mines), Dhanbad

Week - 08

Lecture –40

Hello everyone, welcome to the SWAYAM NPTEL on mathematical geophysics. We continue with module number 8, data-driven analysis in geophysics. This is lecture number 5, *data-driven methods in geophysics*. In this lecture, the concepts covered are related to data-driven methods in geophysics. First, we will look into the concept of singular value decomposition.

Next, the singular value decomposition method. The third component of this lecture is principal component analysis, followed by proper orthogonal decomposition. Finally, we will look into the geophysical applications of data-driven methods. So, let us begin. Singular value decomposition is a powerful linear algebraic tool used for factorization of data.

Since we have seen that geophysical analysis and measurements use data obtained from various instrumental measurements, the representation of this dataset in a proper manner is of utmost importance for its clear interpretation. Singular value decomposition also plays a critical role in various other fields of data-driven science, such as physics and engineering. Singular value decomposition is suitable for data-driven analysis in geophysics. By data-driven analysis, what we understand is that we have a set of data, and without using any further theoretical models, we obtain the underlying structure or the representation of this data in simpler forms. This is primarily done using the data itself rather than any other theoretical equations, models, or other factors.

The data-driven analysis is primarily done using singular value decomposition, which is a matrix-based method. Now, the matrix method is a representation of a set of data in a structured format. All of us know what a matrix looks like. The singular value decomposition decomposes a matrix into three components. For example, let us consider a matrix as shown in equation 1.

We have the datasets arranged in the matrix. These are the datasets. How are these datasets obtained? Let's say we have a domain represented by the Cartesian coordinates x and y. The sample data can be obtained at the grid nodes. Now, this would give us a matrix.

An $n \times m$ matrix, where n is the number of rows and m is the number of columns. Now, the total dataset comprises $n \times m$ or nm numbers. Thus, this can be represented as datasets as given in this diagram. X_1 represents the data which is obtained from $m = 1$. This data is kept in X_1.

The second column is represented by x_2, and so on till x_m. The last column is represented by x_m. This gives us a complex matrix of $\mathbb{C}^{n\times m}$. This is one way to arrange a large dataset in a matrix form denoted by capital X. In cases where n is greater than m, this indicates a tall, skinny matrix, which means that one of the dimensions, n (the number of rows), is much larger than the number of columns. While the opposite case, n much less than m indicates a short, fat matrix. The first one is a tall, skinny matrix, while the bottom one is the short, fat matrix. We can also look into an alternative data scenario. Now, suppose instead of this spatial domain X and Y, we can have time-series data. For example, we have X along the columns and time along the rows.

At a particular time, we have all the data along X arranged in one column. Thus, the columns represent various instances of the data at various times, called snapshots. Thus, the columns are snapshots, and m is the number of snapshots. So here, we have seen how we can arrange the dataset obtained from either spatial data or spatiotemporal data into a matrix form, which will later be used for further data-driven analysis using SVD. Now, coming to the process of singular value decomposition.

The singular value decomposition decomposes any complex matrix into three matrices. This is given by:

$$X = U\Sigma V^T.$$

Here, U and V are unitary matrices with orthonormal columns. Note the dimensions of the U and V matrices. U is an $n \times n$ matrix, while V is an $m \times m$ matrix, where m and n are the rows and columns of the original dataset.

These matrices have orthonormal columns, which means that the individual columns are linearly independent from each other and have a net magnitude equal to unity. The middle matrix, that is Σ, is a real matrix. It consists of non-negative entries on the diagonal only. The off-diagonal entries are 0, and hence Σ, which is $n \times m$ in size, is a diagonal matrix. The singular value decomposition can be represented in diagrammatic form, as given in the adjacent diagram.

Here, X is the dataset matrix, decomposed into the U matrix, which is given by this matrix. The Σ and V matrices are also shown. Note the sizes of these matrices. U is $n \times n$. Now, in this particular case, it has been assumed that n is greater than m. Thus, the U matrix is larger compared to the V matrix.

Both are square matrices. However, the Σ matrix is a rectangular matrix. Now, since n is greater than m, the Σ matrix has a complete submatrix composed of zero entries. These diagonal entries are now zero. The lower submatrix is completely zero entries, while the off-diagonal entries are also zero.

Now, from here, we can proceed to represent this in matrix form as equation number 3. Note that the Σ matrix can be written as the combination of the diagonal part and the zero part, as given here.

$\hat{\Sigma}$ represents the dark-shaded region of the Σ matrix, while the zero matrix is the lighter shade represented respectively. Now, using this notation, we can write $X = U\Sigma V^T$ as U is decomposed into two submatrices: $\hat{U}$ and $\hat{U}^{\perp}$.

$sigma$ is decomposed into two submatrices: $\hat{\Sigma}$ and 0, while V^* is as it is. Now, $\hat{U}$ is the region of the U matrix shown in the dark region, while the lighter part is given by $\hat{U}^{\perp}$ in vector notations. The columns of $\hat{U}^{\perp}$ span a vector space that is complementary and orthogonal to that spanned by $\hat{U}$. This is because n is greater than m. Thus, only the part of m which is in n—that is, in the darker region—is considered. Thus, the columns of U are also called the left singular vectors of X, and the columns of V are called the right singular vectors of X. Now, since we have a $\hat{U}$ matrix or submatrix in particular, and its columns span a vector space which is orthogonal, then we have these columns of $\hat{U}$ vector as basis vectors for an entire vector space, which means we can construct any vector by linear combinations of the columns of $\hat{U}$.

These generalized vectors are nothing but the original dataset. In fact, the singular value decomposition performs segregation of the various modes in the generalized data vectors into $\hat{U}$ columns. The generalized data is also obtained from a combination of the modes. Now, these modes are the simpler forms of the data we are interested in, and they can be obtained from the $\hat{U}$ vector. Now, the diagonal elements $\hat{\Sigma}$, which are in an $n \times m$ complex matrix, are known as singular values.

These singular values are also sorted from largest to smallest in magnitude. The rank of matrix X is equal to the number of non-zero singular values, which is essentially the number of linearly independent vectors that can form the vector space spanned by the columns of the data matrix X. Now, using the singular value decomposition method, we can understand the concept of principal component analysis. Now, principal component analysis is a dimensionality reduction technique. What do we mean by dimensionality reduction?

Dimension is a form of data representation as the number of independent coordinates or the number of independent degrees of freedom in which the data can be represented. For example, let us consider the phenomenon of mantle convection. It depends on various factors such as temperature, pressure, material properties, etc. All these components contribute to mantle convection, and each can independently affect the nature of mantle convection.

Now, these form the various degrees of freedom on which the mantle convection depends. Thus, the data set—for example, fluid flow data from mantle convection—can be analyzed in various degrees of freedom. Now, here we have to reduce the dimensionality. In mathematical terms, it is useful to reduce the dimensionality for better representation of data. We will consider a simpler example for the analysis shown here.

The principal component analysis, by definition, transforms high-dimensional data into a lower-dimensional form. The principal component analysis transforms high-dimensional data into low-dimensional form. It preserves most of the significant features of the original data set, while some

negligibly important features are omitted. Now, the principal component analysis relies on the singular value decomposition to find out the most important or significant features it can retain, and to omit the not-so-essential features.

Now, coming back to the singular value decomposition, note that we have understood that the matrix $\widehat{U}$ here has the columns as the various modes. Now, each individual mode—which is essentially a column of the $\widehat{U}$ vector—represents a significant feature of the data, as significant as the singular values. Now, since the singular values are sorted from largest to smallest, the first column of the $\widehat{U}$ vector is the most significant. The second column of the $\widehat{U}$ vector is the next most significant.

In principal component analysis, it is required to retain the first few columns of $\widehat{U}$ as the most significant features and omit the rest of the data features, which are much less significant. The computational algorithm for the implementation of principal component analysis is shown here. For example, we consider the matrix X, which is the data matrix. We compute the row-wise mean of X, which is essentially taking the entire row and calculating the average.

This gives us $\bar{x}_j$, which is $\frac{1}{n}\sum_{i=1}^{n} x_{ij}$. Recall that here we have $n = 1$ to $n = n$. For each individual column, we would be averaging over this entire column. The average data obtained from this column would be represented by $\bar{x}_j$. Now, this is performed for all the columns. This gives us the mean matrix $\bar{X}$, which is equal to $1 \cdot \bar{x}$, obtained from equation 4.

Subtracting the mean from X, we obtain the mean-removed data matrix. The mean-subtracted data essentially moves the data distribution to be centered at the origin, as shown in this diagram. For example, if the raw data had been located as such, subtracting the mean from the original data would move the dataset to be centered at the origin. That is the function of mean-subtracted data.

Now we compute the covariance matrix. The covariance matrix is given by:

$$C = \frac{1}{N-1} B^T B.$$

This is the covariance matrix. Now the first principal component of this covariance matrix is u_1. This u_1 is given by equation number 8 which reads:

$$u_1 = \arg\max_{||u_1||=1} u_1^* B^* B u_1.$$

Now equation 8 can be represented in terms of matrix calculation as the eigenvector of B^*B. u_1 is the eigenvector of B^*B which has the largest magnitude. Also u_1 is the left singular value of B corresponding to the largest singular value. Essentially u_1 is the first column of the matrix $\widehat{U}$. This first column of $\widehat{U}$ is nothing but the u_1 which is the most important component of this data and hence is known as the principal component. For example, if the original data set is shown as

here, the principal component would mean this direction because this data set is depending upon two axes.

Let us say these are x and y axes, and the distribution of this data set is such that one of the axis is more important than other axis. Now that axis is neither x axis or y axis. It is the axis which is passing through the entire data set. That is this axis.

The perpendicular to this axis is this axis. So instead of choosing x and y as the independent degrees of freedom, we can choose this principal component as the axis to represent this dataset. Now, if we denote the principal axis as P_1 and the second principal axis as P_2, which are the principal components of this data, then we can represent this data in this axis. Now, one can easily distinguish the importance of these axes P_1 and P_2.

The P_1 axis is considerably more significant than P_2, as most of the dataset has alignment along P_1. P_2 is the second most significant axis. But if we look at the original data form, which depends on the X and Y axes, each of these axes is equally important, as the spread, dependence, or sensitivity of the dataset to each of these axes is equivalent. These are the ranges of the x and y axes in the original dataset, which are approximately equal. Hence, the significance of the x and y axes in the original dataset is equivalent.

However, converting this into the principal component form, we can easily see that the principal axis P_1 has much more significance than P_2, which is much less. Thus, one can now represent the entire dataset using P_1 only and neglect P_2. Thus, we have transformed the number of degrees of freedom from 2 to 1. The two degrees of freedom in the original dataset are x and y, while the single important degree of freedom in the principal component axis is P_1 only.

This is the dimensionality reduction that is reduction in the number of dimensions from 2 to 1. Now let us look into the proper orthogonal decomposition technique. The proper orthogonal decomposition technique relies on the singular value decomposition method to obtain the most significant data patterns in a complex data distribution. Inherently, POD is also related to principal component analysis and SVD. However, it is useful for time dependent and spatially dependent data.

This is used in the context of dynamical systems. POD uses singular value decomposition method to represent the data set capital X. Thus, we have the data set X decomposed into $U_k \Sigma_k V_k^*$. The k subscript denotes the restriction of the decomposition to the k most energetic modes. These most energetic modes are nothing but the economic version of the singular value decomposition. The economy version of singular value decomposition only takes in the gray shaded area from the general decomposition.

You can see that the capital X economized can be represented by only the product of $\hat{U}$, $\hat{\Sigma}$ and V^*. Here the U^{**} and zero matrices are omitted. The number of columns in these are k. This is the economic SVD. The economic SVD is used in proper orthogonal decomposition to find the most

dominant k modes which are energetically the maximum. Note that in principal component analysis, the energy maximum is not considered.

The principal component analysis is a general technique where the dimensionality is reduced, while POD is useful to obtain the most energetic modes in the dataset. Essentially, POD represents the dominant structures or patterns in the data. It can also indicate how each of these patterns evolves over time. It also indicates the relative importance of each mode. Now, let us consider these diagrams, which are the proper orthogonal decomposition of real geomagnetic field datasets.

Now consider these diagrams, where the latitude and longitude of the global distribution of the radial magnetic field are shown. The radial magnetic field is the r component of the magnetic field vector B. We have the longitudes from left to right and the latitudes from top to bottom. The central line indicates the equator. This diagram is essentially the Hammer projection of the globe.

Now, this projection is useful to represent the entire globe's data on a flat surface. Having understood the mapping, let us focus on the pattern of the magnetic field. One can see that the magnetic field, or the radial magnetic field, is positive in the northern hemisphere while negative in the southern hemisphere. Now, this is a very complicated distribution. Upon application of the proper orthogonal decomposition, we obtain two of the most energetic modes, as shown here. The most energetic mode indicates that the most significant pattern in the data set is given by the mode 1 which is essentially a dipolar structure, indicated by the spherical harmonic Y_1^0. As we know, the Earth's magnetic field is dominantly dipolar in nature. That is also indicated by the mathematical analysis of proper orthogonal decomposition, which is truly remarkable. The second mode is indicated by an oscillatory, spatially oscillatory mode which is given by POD mode 2.

This feature shows that the oscillatory patterns or spatially periodic patterns near the equator form the next most energetic dominant mode of the radial magnetic field. These modes arise from the patterns which are distributed close to the equator. Thus, from such type of analysis, we can understand that the radial magnetic field has two major patterns. One is the north-south dipolar distribution and the equatorially spatially periodic distribution. This diagram represents a typical decay of the singular values in the POD analysis.

The most energetic and the significant data structure is shown as mode number one. The significance of the individual modes gradually decay and become lesser and lesser as the mode number m increases. In geophysics, various applications of the data-driven methods are prevalent. First, let us see the seismic data processing. In seismic data processing, SVD is used to separate coherent signals from random noise.

The seismic data matrices are decomposed to obtain the dominant singular values. These dominant singular values correspond to particular and significant seismic events. The smaller singular values denote noise. In geophysics, various ill-posed inverse problems can also be solved using the

technique of singular value decomposition. The singular value decomposition regularizes these problems by decomposing the model space and selectively retaining the significant singular values.

This reduces numerical instability and makes the ill-posed problem a well-posed problem. In today's scenario, where machine learning and its applications are inherently becoming more important and significant, geophysical exploration also makes use of such applications, which are based on singular value decomposition and principal component analysis. Features such as dimensionality reduction and extraction of essential features improve the models and performance of interpretation. This helps in mineral prospective mapping and subsurface classification. Coherent seismic reflections are also separated using principal component analysis in seismic studies.

The reconstruction of the dataset using only the most significant modes enhances the signal-to-noise ratio. Large-scale geophysical inversion problems, such as those involved in gravity inversion, magnetic inversion, and electrical data inversion, utilize proper orthogonal decomposition to reduce the number of model parameters. This is essentially similar to dimensionality reduction of the data. This results in a smaller set of dominant modes, which represent the subsurface structures more accurately and clearly. Thus, we come to the conclusion of the present lecture.

First, data decomposition. SVD is effective in separating signals from noise by decomposing the data into orthogonal components. Second, noise reduction. The filtering action of singular value decomposition enhances the quality and the signal-to-noise ratio. This clarifies the geophysical signals.

Third, dimensionality reduction. The technique of principal component analysis simplifies large geophysical datasets and reduces the degrees of freedom. This helps in identifying the most important contributing factors for the geophysical phenomena and their control. Next, we have noise attenuation. Noise attenuation is achieved effectively by principal component analysis.

This enhances the data quality. Finally, proper orthogonal decomposition helps identify the most energetic modes. It helps detect significant data patterns and further enhances the interpretation and geometrical representation of these datasets, which are useful in geophysical applications. Thus, we come to the conclusion of this module and this lecture series. One can go through the following references and the various references we have discussed throughout this lecture series for obtaining a greater understanding of the subject of mathematical geophysics.

I thank you for your patience and attention throughout this course, and I hope this will be very useful and relevant for your further understanding of mathematics as well as geophysics, and the symbiotic relationship that exists between the two, which enhances our understanding of the Earth, planetary systems, and nature in general. Thank you very much.