Study Guide: The Singular Value Decomposition (SVD)

Dan Kalman

1 Introduction to the SVD

The **Singular Value Decomposition (SVD)** of a matrix is a fundamental result in linear algebra that has both practical and theoretical significance. For an $m \times n$ matrix A, the SVD expresses A as:

$$A = U\Sigma V^T$$

where:

- U is an $m \times m$ orthogonal matrix (left singular vectors),
- V is an $n \times n$ orthogonal matrix (right singular vectors),
- Σ is an $m \times n$ diagonal matrix with non-negative real numbers σ_i, called **singular values**.

1.1 Connection to Eigenvalue Decomposition (EVD)

The SVD generalizes the eigenvalue decomposition (EVD) of a symmetric matrix. Recall that for a symmetric matrix A, the EVD is:

$$A = V\Lambda V^T$$

where V is orthogonal and Λ is diagonal. Similarly, for the SVD, the matrix $A^T A$ provides eigenvectors V that serve as the right singular vectors in the SVD of A.

2 Geometrical Interpretation

The SVD provides insight into the geometric structure of linear transformations. Specifically, the transformation $A : \mathbb{R}^n \to \mathbb{R}^m$ can be understood as follows:

- The matrix V provides an orthonormal basis for the domain (input space),
- The matrix U provides an orthonormal basis for the range (output space),
- The matrix Σ scales vectors along certain directions, with the singular values representing the magnitudes of the scaling.

For an input vector x, A stretches or compresses x along the directions of the singular vectors, transforming the unit sphere in $\mathbb{R}^n$ into an ellipsoid in $\mathbb{R}^m$.

3 Properties of the SVD

- **Singular Values**: The diagonal entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$ of Σ are the singular values of A, where r is the rank of A.
- **Orthogonality**: The matrices U and V are orthogonal, meaning their columns form orthonormal sets of vectors.
- **Existence and Uniqueness**: Every matrix A has an SVD. If the singular values are distinct, the SVD is unique up to sign.

4 Applications of the SVD

The SVD has a wide range of applications, including:

- **Least Squares Problems**: The SVD provides a stable method for solving linear least squares problems. Given the system $Ax = b$, the SVD allows us to rewrite it as $U\Sigma V^T x = b$ and solve for x with improved numerical stability.
- **Data Compression and Reduced Rank Approximation**: The SVD can be used to approximate a matrix A with a matrix of lower rank. This is useful for data compression, where the matrix A is approximated as:

$$A_r = U_r \Sigma_r V_r^T$$

  where U_r and V_r are truncated versions of U and V, respectively, and Σ_r contains the largest r singular values.
- **Principal Component Analysis (PCA)**: The SVD is a key component of PCA, which is used to reduce the dimensionality of data while preserving as much variance as possible.
- **Generalized Inverses**: The Moore-Penrose pseudoinverse A^+ can be computed using the SVD:

$$A^+ = V\Sigma^+ U^T$$

  where Σ^+ is the pseudoinverse of Σ, constructed by taking the reciprocal of the nonzero singular values and transposing the matrix.

5 Numerical Stability and Condition Numbers

The SVD is known for its numerical stability, particularly in solving ill-conditioned problems where other methods, such as the normal equations, may fail. The **condition number** of a matrix A is defined as the ratio of its largest singular value to its smallest nonzero singular value:

$$\mathrm{cond}(A) = \frac{\sigma_1}{\sigma_r}$$

A high condition number indicates that A is close to being singular, and the SVD can help manage numerical errors in such cases.

6 Outer Product Expansion of the SVD

The SVD can be written as an **outer product expansion**:

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T$$

This expresses A as a sum of rank-1 matrices, each scaled by a singular value. This form is useful for understanding the structure of A and for reduced-rank approximations.

7 Conclusion

The SVD is a powerful tool in both theoretical and applied linear algebra. It provides valuable insights into the structure of matrices and is widely used in applications such as least squares problems, data compression, and principal component analysis.

PCA, SVD, Right Principle Components, and Left Principle Components

Study Notes

1 Relationship Between Data, Principal Components, and Singular Vectors

The relationship between data, its principal components, right singular vectors, and left singular vectors is fundamentally tied to Singular Value Decomposition (SVD), a mathematical technique used to factorize matrices, and Principal Component Analysis (PCA), a statistical method that identifies the directions of maximum variance in the data.

1.1 1. Data Matrix

Consider a dataset represented as a matrix X, where:

$$X \in \mathbb{R}^{m \times n},$$

with:

- m: number of data points (samples),
- n: number of features (dimensions) for each data point.

Each row of X is a data point, and each column represents a feature. For example, for a dataset with 100 observations and 5 features, X would be a 100×5 matrix.

1.2 2. Principal Component Analysis (PCA)

PCA is a technique used to reduce the dimensionality of data while retaining as much variability as possible. It identifies new orthogonal axes (called principal components) that capture the maximum variance in the data.

- **Principal Components** are directions (or axes) in the feature space along which the variance of the data is maximized.
- The first principal component captures the largest variance in the data, the second captures the largest remaining variance orthogonal to the first, and so on.

PCA is closely related to the singular value decomposition (SVD) of the data matrix, which decomposes the data into its key components.

1.3 3. Singular Value Decomposition (SVD)

The SVD of the data matrix X is:

$$X = U\Sigma V^\top,$$

where:

- $U \in \mathbb{R}^{m \times m}$ contains the left singular vectors (orthonormal vectors related to the data samples),

- $\Sigma \in \mathbb{R}^{m \times n}$ is a diagonal matrix containing the singular values (square roots of the eigenvalues of $X^\top X$),
- $V \in \mathbb{R}^{n \times n}$ contains the right singular vectors (orthonormal vectors related to the features of the data).

Each part of the SVD has a specific relationship to the data's principal components and variance structure.

1.4 4. Right Singular Vectors and Principal Components

The right singular vectors $V = [v_1, v_2, \ldots, v_n]$ are unit vectors that define the directions in the feature space along which the variance of the data is maximized. These directions are the principal components in PCA.

- The first right singular vector v_1 corresponds to the direction of the first principal component, explaining the maximum variance in the data.
- The second right singular vector v_2 corresponds to the second principal component, and so on.

The matrix V defines the axes of the new coordinate system (the principal component directions) in the original feature space.

1.5 5. Singular Values and Variance Explained

The singular values in Σ correspond to the amount of variance explained by each principal component. If the data matrix X is centered (i.e., has zero mean), the square of the singular values (i.e., σ_i^2) represents the eigenvalues of the covariance matrix of the data. These eigenvalues quantify the variance explained by each principal component.

1.6 6. Left Singular Vectors and Principal Component Scores

The left singular vectors $U = [u_1, u_2, \ldots, u_m]$ represent the data points projected onto the principal component axes. These projections are called *principal component scores*.

- The columns of U describe how much each data point contributes to each principal component.
- The first left singular vector u_1 corresponds to the projection of the data points onto the first principal component, the second left singular vector u_2 corresponds to the projection onto the second principal component, and so on.

Thus, the left singular vectors give us the coordinates of the data points in the space defined by the principal components.

1.7 7. Link Between PCA and SVD

- In PCA, the principal components are the eigenvectors of the covariance matrix $X^\top X$, which correspond to the right singular vectors V in the SVD.
- The principal component scores (projections of data onto the principal components) are captured by the product $U\Sigma$, where U gives the directions in the new space and Σ scales these directions based on how much variance is explained by each principal component.

1.8 8. Summary of Relationships

- **Right Singular Vectors** (V) = *Principal Components* (directions of maximum variance).
- **Singular Values** (Σ) = *Measure of the variance* explained by each principal component.
- **Left Singular Vectors** (U) = *Projections of data points* onto the principal component directions (i.e., the coordinates of the data in the principal component space).
- **Principal Component Scores** = $U\Sigma$, representing the transformed data in the lower-dimensional space of principal components.

1.9 Example

Suppose you have a dataset X with 100 data points (samples) and 5 features:

- The right singular vectors V describe the 5 new axes (principal components) in the 5-dimensional feature space that capture the maximum variance.
- The singular values in Σ tell you how much variance is captured by each of these axes.
- The left singular vectors U describe how the 100 data points are projected onto these new axes.
- If you reduce the dimensionality by keeping only the first 2 principal components, you would use the first 2 columns of V (the first 2 principal components), the corresponding singular values from Σ, and the first 2 columns of U to describe your data in this new 2-dimensional space.

1.10 Conclusion

In summary, the principal components are derived from the right singular vectors of the data matrix, with singular values representing the amount of variance each component explains. The left singular vectors represent how the data points map to these new axes. Together, SVD provides a powerful tool for understanding the structure of the data, with applications in PCA and dimensionality reduction.