Study Guide on Projections and Norms (Conceptual Understanding)

1 Projections

A projection is a way of "mapping" a vector onto a subspace, essentially reducing its dimensionality while preserving as much information as possible in the new subspace. Geometrically, projections can be thought of as "shadows" or the closest point on a subspace to the original vector.

1.1 Conceptual Explanation of Projection

- **Purpose:** The idea of a projection is to find the component of a vector that lies within a particular subspace. For example, given a vector in space, its projection onto a line is the closest point on the line to that vector.
- **Orthogonal Projection:** The most common type of projection is the orthogonal projection, where we project a vector onto a subspace, and the resulting vector is the closest point to the original vector on that subspace. The error between the original vector and its projection is minimized, and the difference (error) is orthogonal (i.e., at a right angle) to the subspace.

1.2 Projections in Terms of Matrices

In the context of matrices, a projection matrix P is a square matrix that, when multiplied by a vector v, gives the projection of v onto a subspace:

$$Pv = \text{proj}_W v$$

where W is the subspace onto which we are projecting.

- **Idempotence:** A projection matrix P satisfies $P^2 = P$. This means that projecting a vector twice has the same effect as projecting it once.
- **Symmetry:** In the case of orthogonal projections, the matrix is symmetric, i.e., $P = P^T$.

1.3 Geometric Intuition of Projections

Imagine a 3D space where you have a vector v and a plane. The projection of v onto the plane is the point on the plane that is "closest" to v. The vector connecting v to its projection is orthogonal (perpendicular) to the plane.

2 Norms

A norm is a measure of the "size" or "length" of a vector. Norms generalize the concept of distance in vector spaces and provide a way to quantify how large a vector is.

2.1 Conceptual Explanation of Norms

- **Purpose:** Norms are used to measure the magnitude or length of vectors. For example, in 2D or 3D space, the Euclidean norm (or ℓ_2-norm) of a vector gives its straight-line distance from the origin.

2.2 Common Norms

- **ℓ_2-Norm (Euclidean Norm):** The most familiar norm, which measures the straight-line distance between a vector and the origin.

  $$\|v\|_2 = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$$

  This is what we typically think of as "length" in geometry.

- **ℓ_1-Norm (Manhattan Norm):** This norm sums the absolute values of a vector's components, measuring the distance you would travel in a grid-like path.

$$\|v\|_1 = |v_1| + |v_2| + \cdots + |v_n|$$

- **ℓ_∞-Norm (Maximum Norm):** This norm measures the size of the largest component in the vector.

$$\|v\|_\infty = \max(|v_1|, |v_2|, \ldots, |v_n|)$$

2.3 Norms in Terms of Matrices

Norms can also apply to matrices, where they measure the "size" of the matrix in various ways. One common norm is the Frobenius norm, which is similar to the Euclidean norm for matrices, summing the squares of all elements:

$$\|A\|_F = \sqrt{\sum_{i,j} |a_{ij}|^2}$$

2.4 Geometric Intuition of Norms

- **Vector Norm:** Imagine a vector as an arrow pointing from the origin to some point in space. The norm measures the length of this arrow, which gives an intuitive sense of the "size" or "magnitude" of the vector.
- **Matrix Norm:** The norm of a matrix can be thought of as measuring the "stretching" effect the matrix has when applied to a vector. For instance, applying a matrix to a vector changes its length, and the norm of the matrix captures the maximum amount by which the matrix can stretch a vector.

3 Relationship Between Projections and Norms

- **Projections Minimize Norms:** One way projections and norms are connected is through the idea that a projection minimizes the distance (as measured by a norm) between the original vector and the subspace. For an orthogonal projection of a vector v onto a subspace, the error (difference between the original vector and the projection) has the smallest possible norm (or length).
- **Norm of a Projection:** When you project a vector onto a subspace, the norm of the projected vector gives you the length of the vector in the subspace. The norm of the difference between the original vector and its projection measures how far the original vector is from the subspace.

4 How Projections and Norms Relate to Vectors and Matrices

- **Vectors:** Projections are used to decompose vectors into components along a subspace, such as finding the part of a vector that lies along a specific direction. Norms are used to measure the length or size of these vector components.
- **Matrices:** Projections are represented by matrices that map vectors onto subspaces. Norms of matrices help us understand how much a matrix can stretch a vector. For example, in optimization, we often want to project a vector onto a feasible region, and matrix norms help evaluate how certain transformations affect the vector's size.

5 Summary

- Projections map a vector onto a subspace, and orthogonal projections find the closest point in the subspace to the vector. Mathematically, this is often represented by a projection matrix.
- Norms measure the size or length of vectors (or matrices). The ℓ_2-norm is the most common and corresponds to the Euclidean length.
- Projections and norms are connected because projections minimize the norm of the difference between the original vector and its projection, meaning they find the closest point in the subspace.

By understanding these concepts, you can better grasp the geometric meaning of vector operations in linear algebra and how these operations relate to optimization, machine learning, and other mathematical fields.

Spans, Orthogonality, and Subspaces

Study Notes

1 Spans

The span of a set of vectors is the collection of all possible linear combinations of those vectors.

1.1 Definition

The span of vectors $v_1, v_2, \ldots, v_k$ is the set of all vectors that can be written as:

$$\text{Span}(v_1, v_2, \ldots, v_k) = \{c_1 v_1 + c_2 v_2 + \cdots + c_k v_k \mid c_1, c_2, \ldots, c_k \in \mathbb{R}\}$$

1.2 Geometric Interpretation

The span represents the space "covered" by all combinations of the given vectors. For example:

- The span of one vector is a line through the origin.
- The span of two vectors (if linearly independent) is a plane.

2 Orthogonality

Two vectors are orthogonal if they are at a 90-degree angle to each other (perpendicular).

2.1 Definition

Vectors u and v are orthogonal if their dot product is zero:

$$u \cdot v = 0$$

2.2 Geometric Interpretation

Orthogonal vectors point in completely independent directions. For instance, in 2D, the x-axis and y-axis vectors are orthogonal.

2.3 Orthogonal Projections

When projecting a vector onto a subspace, the difference between the vector and its projection is orthogonal to the subspace.

3 Subspaces

A subspace is a set of vectors that forms a smaller vector space within a larger one. It must satisfy three conditions: containing the zero vector, being closed under vector addition, and being closed under scalar multiplication.

3.1 Definition

A subspace W of a vector space V is any subset of V that is itself a vector space under the same operations.

3.2 Examples

- A line through the origin is a subspace of $\mathbb{R}^2$.
- A plane through the origin is a subspace of $\mathbb{R}^3$.

3.3 Relation to Span

The span of a set of vectors forms a subspace, since it contains all possible linear combinations of those vectors.

4 Summary

- **Span:** The set of all linear combinations of given vectors, describing the space they "cover."
- **Orthogonality:** Two vectors are orthogonal if their dot product is zero (perpendicular).
- **Subspaces:** A subset of a vector space that is itself a vector space, closed under addition and scalar multiplication.

These concepts are fundamental to understanding vector spaces, linear independence, and vector geometry.

Transpose of a Matrix

Study Notes

1 What is the Transpose of a Matrix?

The transpose of a matrix is an operation that flips a matrix over its diagonal. This means that the rows of the original matrix become the columns in the transposed matrix, and the columns become rows.

1.1 Notation

If A is a matrix, its transpose is denoted by A^T.

2 How the Transpose Works

For a matrix A with elements A_{ij} (where i represents the row index and j represents the column index), the transpose A^T is a matrix where:

$$A_{ij}^T = A_{ji}$$

In simple terms, the element in the i-th row and j-th column of the original matrix becomes the element in the j-th row and i-th column of the transposed matrix.

3 Example

Consider the following 2×3 matrix A:

$$A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$$

- The rows of A are: $[1, 2, 3]$ and $[4, 5, 6]$.
- The columns of A are: $[1, 4]$, $[2, 5]$, and $[3, 6]$.

Now, the transpose A^T will swap rows and columns:

$$A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

4 Step-by-Step Transposition

- The first row of A, $[1, 2, 3]$, becomes the first column of A^T.
- The second row of A, $[4, 5, 6]$, becomes the second column of A^T.

Thus, the matrix A^T is a 3×2 matrix (the dimensions are flipped from 2×3 to 3×2), and it looks like this:

$$A^T = \begin{pmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{pmatrix}$$

5 Summary

- The transpose operation swaps rows and columns.
- If A is an $m \times n$ matrix, its transpose A^T will be an $n \times m$ matrix.

6 The Matrix Product $A^T A$

The matrix product $A^T A$ (the transpose of matrix A multiplied by matrix A) has several important interpretations and properties, especially in linear algebra, machine learning, and data analysis. Below are key insights into what $A^T A$ tells us:

6.1 Gram Matrix and Dot Products

$A^T A$ is often referred to as the Gram matrix of A. It contains the dot products between the columns of A.

- If A is an $m \times n$ matrix, then $A^T A$ is an $n \times n$ matrix. Each entry $(A^T A)_{ij}$ of $A^T A$ is the dot product between the i-th and j-th columns of A:

  $$(A^T A)_{ij} = a_i^T a_j$$

  where a_i and a_j are the i-th and j-th columns of A.

This tells us about the similarity between the columns of A. If the columns of A are orthogonal, $A^T A$ will be diagonal, meaning the dot products between different columns are zero.

6.2 Covariance Matrix Interpretation (for Data Matrices)

In data analysis or machine learning, if the rows of A represent data points and the columns represent features, $A^T A$ is closely related to the covariance matrix of the data.

- If A is centered (i.e., the mean of each column is zero), then $A^T A$ is proportional to the covariance matrix. The diagonal entries of $A^T A$ represent the variance of each feature, and the off-diagonal entries represent the covariance between different features.
- Specifically, the covariance matrix is given by:

  $$\frac{1}{m-1} A^T A$$

  where m is the number of rows (data points) in A.

6.3 Positive Semidefiniteness

$A^T A$ is always positive semidefinite, meaning for any vector v:

$$v^T (A^T A) v \geq 0$$

This property ensures that the eigenvalues of $A^T A$ are non-negative, which is useful in optimization and matrix factorization applications.

6.4 Rank and Invertibility

The rank of $A^T A$ is equal to the rank of A. This is because the rank of the matrix product $A^T A$ cannot exceed the rank of the original matrix A.

- If A has full column rank (i.e., its columns are linearly independent), then $A^T A$ is invertible. In this case, $A^T A$ has full rank, and the inverse $(A^T A)^{-1}$ exists.

6.5 Least Squares Solution

$A^T A$ is fundamental in the solution to the least squares problem. When solving for x in an overdetermined system $Ax = b$ (where there are more equations than unknowns), the normal equation is:

$$A^T A x = A^T b$$

This equation minimizes the error between the predicted values Ax and the observed values b, and $A^T A$ helps find the best fit.

6.6 Example

Consider a simple matrix A:

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

First, compute A^T:

$$A^T = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix}$$

Now compute $A^T A$:

$$A^T A = \begin{pmatrix} 1 & 3 \\ 2 & 4 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} = \begin{pmatrix} 1^2 + 3^2 & 1 \cdot 2 + 3 \cdot 4 \\ 2 \cdot 1 + 4 \cdot 3 & 2^2 + 4^2 \end{pmatrix} = \begin{pmatrix} 10 & 14 \\ 14 & 20 \end{pmatrix}$$

Here, $A^T A$ tells us:

- The diagonal entries (10 and 20) are the squared norms (or magnitudes) of the columns of A.
- The off-diagonal entries (14) are the dot products between the columns of A.

6.7 Summary

- $A^T A$ contains information about the similarity (dot products) between the columns of A.
- It is positive semidefinite and can be used in solving least squares problems.
- $A^T A$ is related to the covariance matrix in data analysis when A is a data matrix.
- It provides key insights into the rank and invertibility of A.