Summary Notes: Linear Algebra & Singular Value Decomposition

---

1 Core Concepts in Linear Algebra

1.1 Inner Product and Norms

Definition 1.1 (Inner Product on a Real Vector Space)**.** An **inner product** on a real vector space X is a function

$$\langle \cdot, \cdot \rangle : X \times X \to \mathbb{R}$$

that satisfies the following for all $x, y, z \in X$ and $\alpha, \beta \in \mathbb{R}$:

1. **Positivity:** $\langle x, x \rangle \geq 0$, and $\langle x, x \rangle = 0$ if and only if $x = 0$.
2. **Symmetry:** $\langle x, y \rangle = \langle y, x \rangle$.
3. **Linearity in the first argument:** $\langle \alpha x + \beta y, z \rangle = \alpha \langle x, z \rangle + \beta \langle y, z \rangle$.

Remark: For complex inner-product spaces, the inner product is conjugate-linear in one argument and linear in the other.

Definition 1.2 (Norm and Distance)**.** For a vector x in a real inner product space, the **norm** of x is:

$$||x|| = \sqrt{\langle x, x \rangle}$$

The **distance** between two vectors x and y is:

$$d(x, y) = ||x - y||$$

1.2 Orthogonality and Spectral Decomposition

Definition 1.3 (Orthogonal Matrix)**.** A real square matrix $S \in \mathbb{R}^{n \times n}$ is said to be **orthogonal** if its transpose is its inverse:

$$SS^T = S^T S = I$$

The columns (and rows) of an orthogonal matrix form an orthonormal basis for $\mathbb{R}^n$.

Theorem 1.1 (Spectral Decomposition)**.** *A real square matrix A is orthogonally diagonalizable if and only if it is* ***symmetric****. For any symmetric matrix $A \in \mathbb{R}^{n \times n}$, there exists an orthogonal matrix S and a diagonal matrix D such that:*

$$A = SDS^T = \sum_{i=1}^{n} \lambda_i u_i u_i^T$$

where:

- *λ_i are the eigenvalues of A,*
- *$\{u_1, \ldots, u_n\}$ is an orthonormal set of eigenvectors forming the columns of S,*
- *D is diagonal with entries λ_i.*

Remark: *If eigenvalues have multiplicity greater than one, an orthonormal basis can be chosen in each eigenspace.*

---

2 Singular Value Decomposition (SVD)

Definition 2.1 (Rank of a Matrix)**.** The **rank** of a matrix M, denoted $\text{rank}(M)$, is the number of linearly independent columns (or rows) of M. It is equal to the number of non-zero singular values of M.

Theorem 2.1 (Singular Value Decomposition)**.** *Any real matrix $A \in \mathbb{R}^{m\times n}$ can be decomposed as:*

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

where:

- *U is an $m \times m$ orthogonal matrix whose columns are eigenvectors of AA^T.*
- *V is an $n \times n$ orthogonal matrix whose columns are eigenvectors of $A^T A$.*
- *Σ is an $m \times n$ diagonal matrix with non-negative entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$ (the **singular values**).*

Economy SVD: *If $r = \text{rank}(A)$, we often use a reduced form:*

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

where:

$$U_r \in \mathbb{R}^{m\times r}, \quad \Sigma_r \in \mathbb{R}^{r\times r}, \quad V_r \in \mathbb{R}^{n\times r}.$$

Here, U_r and V_r have orthonormal columns, and Σ_r contains only the r positive singular values.

---

3 Applications of SVD

3.1 Low-Rank Approximation

The primary application of SVD is finding the best low-rank approximation of a matrix.

Theorem 3.1 (Eckart-Young-Mirsky Theorem)**.** *Let the SVD of M be $U\Sigma V^T$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq 0$. The best rank-$k$ approximation to M, denoted $\tilde{M}_k$, is:*

$$\tilde{M}_k = \sum_{i=1}^{k} \sigma_i u_i v_i^T$$

This minimizes the approximation error in both spectral and Frobenius norms:

1. ***Spectral Norm:***
$$||M - \tilde{M}_k||_2 = \sigma_{k+1}$$
2. ***Frobenius Norm:***
$$||M - \tilde{M}_k||_F^2 = \sum_{i=k+1}^{\min(m,n)} \sigma_i^2$$

3.2 Real-World Scenarios

- **Image Compression:** An image can be represented as a matrix of pixel intensities. A low-rank approximation using SVD significantly reduces storage while preserving visual quality.
- **Recommendation Systems:** In systems like Netflix, user ratings form a large, sparse matrix. SVD helps reveal hidden latent features (e.g., genres, actors) and predict missing ratings for better recommendations.

3.3 Gram Matrix

The **Gram matrix** provides a way to express an inner product with respect to a specific set of vectors.

- **Setup:** Consider a real inner product space V with a set of vectors $B = \{b_1, b_2, \ldots, b_n\}$.
- **Definition:** The Gram matrix G is the $n \times n$ matrix:

$$g_{ij} = \langle b_i, b_j \rangle$$

- **Inner Product Formula:** If x_β and y_β are coordinate vectors of x and y relative to B,

$$\langle x, y \rangle = x_\beta^T G y_\beta$$

- **Property:** G is always symmetric and **positive semidefinite**. If the vectors in B are linearly independent (i.e., B is a basis), then G is positive definite, meaning:

$$x^T G x > 0 \quad \text{for all } x \neq 0.$$