23MAT204
Class Notes

Adithya Nair

Chapter 1

Revision Linear Algebra

Theorem 1.0.1. Every matrix satisfies its own characteristic matrix.

What this means is

$$\lambda^3 - 2\lambda^2 + \lambda - 4 = 0$$
$$A^3 - 2A^2 + A - 4I = 0$$
$$A^2 - 2A + I - 4A^{-1} = 0$$
$$A^{-1} = \frac{1}{4}[A^2 - 2A + I]X \qquad = A^{-1}b = \frac{1}{4}[A^2 - 2A + I]b$$

There are some implications behind this. Here's a problem, generate a random 3x3 matrix, find the ranks of $(A - \lambda_1 I)$ and then $(A - \lambda_1 I)(A - \lambda_2 I$

What you will see is that the rank reduces upon multiplying roots of the characteristic equation

1.1 Ways To Calculate Whether A Matrix Is Positive Definite

1. The minors are positive then negative alternating.
2. The eigenvalues are positive.

1.2 Large Eigenvalue Computation

Large eigenvalues are computed by using a matrix multiplication method. This method involves:

1.3 Specral Decomposition

$$S = Q\Lambda Q^T = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \cdots + \lambda_n q_n q_n^T$$

This allows us to represent a matrix by the sum of n rank 1 matrices. This is used in square symmetric matrices.

This is the basis behind Principal Component Analysis.

1.4 Singular Value Decomposition

For rectangular matrices.

$$A = U\Sigma V^T = \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_r u_r v_r^T$$

This is mainly used for getting useful properties about the matrix without needing to perform significant operations on this matrix, such as the orthogonality.

Note:-

Singular values are the eigenvalues of a matrix which are **non-zero**

So a rectangular matrix $m \times n$ would have AA^T which is $m \times m$ and $A^T A$ which is $n \times n$. The number of singular values is whether m or n is lower.

Note:-

The trace of $A^T A$ equal to the sum of all a_{ij}^2 The trace is the sum of all eigenvalues of $A^T A$, and For $A_{m \times n}$ that's the sum of the eigenvalues squared.
This is known as the **Frobenius Norm**

This method can also be used to generate orthonormal bases for the 4 Fundamental Subspaces

1.5 The Geometry of SVD

Since we have an orthogonal matrix, diagonal and orthogonal matrix... It's a rotation step followed by a stretching step, finally ended by another rotation step.

1.6 Polar Decomposition

This decomposition is a special case of the Singular Value Decomposition. Where you decompose the elements into the polar form with an orthogonal matrix and a positive semi-definite matrix.

For 2D,

$$\vec{x} = r(\cos(\theta) + sin(\theta))$$

$$\begin{aligned} A &= U\Sigma V^T \\ &= U(V^T V)\Sigma V^T \\ &= (UV^T)(V\Sigma V^T) \\ &= Q \times S \end{aligned}$$

Here S is the scaling matrix, and Q is the orthogonal matrix.

Note:-

Although my assumptions might be wrong, this might be a way to generate rotation matrices for higher-dimension spaces. Might be useful in exploring the semantic spaces of LLMs.

1.7 Principal Component Analysis

Why do we use normalization while working with data? We use normalization to ensure that the units are eliminated while working with that given data. **This normalization is done by subtracting by the average and divide by the standard deviation.**

Definition 1.7.1: Magic Matrix

Matrix where the row and columns sums are equal.

1.8 Norms Of Vectors And Matrices

Norms are a way to measure the size of a vector/matrix. The norm of a vector which calculates the magnitude is known as the L^2 norm or the Euclidean norm $\|v\|_2$. This number gives us the length of the vector.

In general, the $\|u\|_p$ or the p-norm of a vector is $[|u_1|^p + |u_2|^p + \cdots + |u_n|^p]^{\frac{1}{p}}$.

Now, the L_1 norm would be the sum of the components of the vector. This is the sum of the projections to each corresponding axis.

The L_3 norm would be $[|u_1|^3 + |u_2|^3 + \cdots + |u_n|^3]^{\frac{1}{3}}$

The L_∞ norm would be $[|u_1|^\infty + |u_2|^\infty + \cdots + |u_n|^\infty]^{\frac{1}{\infty}}$. This returns the maximum vector component of the vector.

Let's take the equation, $\|v\|_1 = 1$

$$\begin{aligned} x + y &= 1 \\ x - y &= 1 \\ -x + y &= 1 \\ -x - y = 1 \end{aligned}$$

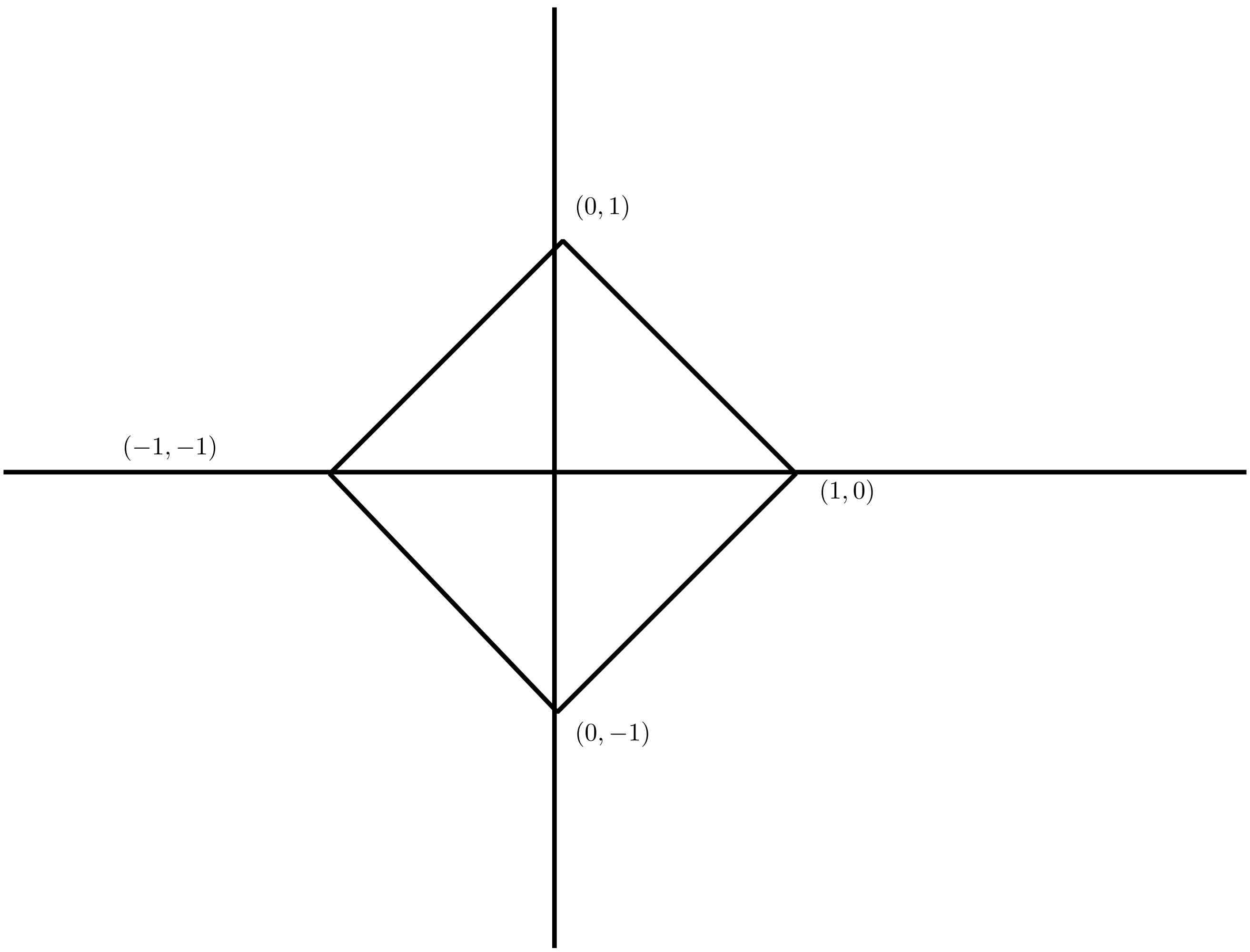

Figure 1.1: We find 4 lines to satisfy these conditions.

Definition 1.8.1: S-norm

The S-norm of a vector $\vec{x}$ is $\vec{x}^T S \vec{x}$. When S is a symmetric positive definite matrix, this S-norm is known as the energy of vector $\vec{v}$

There are three types of Matrix norms:

1. Spectral Norm
2. Frobenius Norm = $\sqrt{\sigma_1^2 + \sigma_2^2 + \cdots + \sigma_r^2}$
3. Nuclear Norm

1.8.1 Spectral Norm

We know that the vector norm for a vector $\vec{x}$ is nothing but $\vec{x}^T\vec{x}$. We take this property.

$$
\begin{aligned}
Max\|A\|_2^2 &= Max\frac{\|Ax\|_2^2}{\|x\|_2^2} \\
&= Max\frac{x^TA^TAx}{x^Tx} \\
&= Max\{\lambda_i(S)\} = \lambda_1 = \sigma_1^2
\end{aligned}
$$

1.8.2 Frobenius Norm

The Frobenius norm for a matrix M, $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ is the equation $\sqrt{a_{11}^2 + a_{12}^2 + a_{21}^2 + a_{22}^2}$

1.8.3 Nuclear Norm

The nuclear norm is the sum of the singular values of a matrix A.

For an identity matrix,

- The Spectral Norm is 1
- The Frobenius Norm is $\sqrt{n}$
- The Nuclear Norm is n

For an orthogonal matrix,

- The Spectral Norm is 1
- The Frobenius Norm is $\sqrt{n}$
- The Nuclear Norm is n

1.9 Best Low Rank Matrix

We say that a matrix is the best approximation of another matrix, based on the Frobenius Norm.. For a singular value decomposition $A = U\Sigma V^T$, if we assume that the singular values are arranged in descending order... We can select the singular value range where the values are significant contributors to the final matrix.

We can then reduce the size of U, Σ and V^T into a smaller matrix B, based on the number of singular values chosen which would give the best approximation.

> **Theorem 1.9.1.** Let $A = U\Sigma V^T$ where $\Sigma : \sigma_1 \geq \sigma_2 \geq \cdots \sigma_n$, then B = $U_{m\times m}\Sigma V^T n \times n$ is a best rank-k approx. to A. Where, S is a diagonal matrix of $n \times n$ where $s_i = \sigma_i(i = 1 \cdots k)$ else $s_i = 0$, by best B is a solution to $min_B\|A - B\|F$ where rank(B) = k