Chapter 1

Overview of Matrix Algebra

Usually one of the most important tools in understanding whether two objects are close is that of the norm. If the said objects are vectors then we use the vector norm to compute their distance. Let us define this tool,

Definition 1. *A vector norm, denoted by $\|\cdot\|$, is a function which has vectors as input and produces a number as output. In other words $\|\cdot\| \in \mathbb{R}^n \to \mathbb{R}$. The vector norm has the following properties,*

$$\begin{array}{ll}
a) & \|\mathbf{x}\| \geq 0 \quad \textit{for all} \quad \mathbf{x} \in \mathbb{R}^n \\
b) & \|\mathbf{x}\| = 0 \quad \textit{if and only if} \quad \mathbf{x} \equiv 0. \\
c) & \|k\mathbf{x}\| = |k|\|\mathbf{x}\| \quad \textit{for all } k \in \mathbb{R} \textit{ and } \mathbf{x} \in \mathbb{R}^n \\
d) & \|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \quad \textit{for all} \quad \mathbf{x}, \mathbf{y} \in \mathbb{R}^n
\end{array}$$

There are several different kinds of vector norms. We most often use the following types of norms in our calculations:

Definition 2. *The Euclidean or l_2 norm,*

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}$$

and the maximum or l_∞ norm,

$$\|\mathbf{x}\|_\infty = \max_{1 \leq i \leq n} |x_i|$$

or the l_1 norm

$$\|\mathbf{x}\|_1 = \sum_{i=1}^{n} |x_i|$$

Example
Calculate all three: the l_1, Euclidean and maximum norm for the vector, $x = [1, -3, 2]$.
Solution
The l_1 norm is,

$$\|\mathbf{x}\|_1 = |1| + |-3| + |2| = 6$$

the Euclidean norm is,

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{n} x_i^2} = \sqrt{x_1^2 + x_2^2 + x_3^2} = \sqrt{1 + (-3)^2 + 2^2} = \sqrt{14}$$

while the maximum norm gives,

$$\|\mathbf{x}\|_\infty = \max_{1 \le i \le n} |x_i| = \max\{|x_1|, |x_2|, |x_3|\} = \max\{1, 3, 2\} = 3$$

Therefore using these norms we can now describe distances between vectors as follows,

$$\|\mathbf{x} - \mathbf{y}\|_2 = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad \text{or} \quad \|\mathbf{x} - \mathbf{y}\|_\infty = \max_{1 \le i \le n} |x_i - y_i|$$

As a result we are now also in position to understand the concept of *convergence* for vectors. Convergence is a concept which we will use later in our numerical calculations and will allow us to know how close we are to the solution of a given problem. To do this we essentially need to measure, usually via some kind of norm, the distance from our current approximate solution to the true solution.

Definition 3. *We say that a vector sequence $\{\mathbf{x}_n\}_{n=0}^{\infty}$ converges to another vector $\mathbf{x}$ in $\mathbb{R}^n$ with respect to given norm $\|\cdot\|$ if given any $\epsilon > 0$ there exist an integer $N(\epsilon)$ such that*

$$\|\mathbf{x}_n - \mathbf{x}\| \le \epsilon \quad \textit{for all} \quad n \ge N(\epsilon)$$

The following result can now be shown:

Theorem 4. *The sequence of vectors $\{\mathbf{x}_n\}_{n=0}^{\infty}$ converges to $\mathbf{x}$ in $\mathbb{R}^n$ with respect to the $\|\cdot\|_\infty$ norm if and only if*

$$\lim_{n \to \infty} x_n(i) = x(i) \quad \textit{for all } i = 1, 2, \ldots, n$$

Similarly we can prove the following ordering between different norms,

Theorem 5. *For $\mathbf{x} \in \mathbb{R}^n$ we have,*

$$\|\mathbf{x}\|_\infty \le \|\mathbf{x}\|_2 \le \sqrt{n}\|\mathbf{x}\|_\infty$$

There are a couple of well known inequalities which apply to vectors. The triangle inequality,

$$\|\mathbf{x} + \mathbf{y}\| \le \|\mathbf{x}\| + \|\mathbf{y}\|$$

and the Cauchy-Schwarz inequality,

$$\left| \sum_{i=1}^{n} x_i y_i \right| \le \|\mathbf{x}\|_2 \|\mathbf{y}\|_2 \tag{1.1}$$

Example

Verify the Cauchy-Schwarz inequality for the following vectors,

$$\mathbf{x} = [1, 2, 3] \quad \text{and} \quad \mathbf{y} = [0, 2, 1]$$

Solution
According to the Cauchy-Schwarz inequality the left hand side of (1.1) gives,

$$|0+4+3| = 7$$

while the right hand side of (1.1),

$$\sqrt{1+4+9}\sqrt{0+4+1} = \sqrt{14}\sqrt{5} = \sqrt{70} \approx 8.36$$

Thus the right hand side is bigger as expected.

We now generalize these results. First we define the following general p norm,

$$\|\mathbf{x}\|_p = \sqrt[p]{|x_1|^p + \cdots + |x_n|^p}$$

Thus the following two inequalities hold,

$$\begin{array}{lrcl} \underline{\text{Holder}}\text{ ineq.} & |\mathbf{x}^T\mathbf{y}| & \leq & \|\mathbf{x}\|_p\|\mathbf{y}\|_q \quad \text{where} \quad \frac{1}{q}+\frac{1}{p}=1 \\ \underline{\text{Minkowski}} & \|\mathbf{x}+\mathbf{y}\|_p & \leq & \|x\|_p + \|y\|_p \end{array}$$

Here $\mathbf{x}^T$ denotes the transpose of vector $\mathbf{x}$.

Now that we have put together the basic necessary tools regarding vectors we can start discussing how to handle matrices. Matrices are just rows of vectors. As such we can essentially use the same tools, from vectors, apply them to matrices. So the matrix norm is defined via,

Definition 6. *Suppose that A and B denote $n \times n$ size matrices and k a constant. Then the matrix norm is the real valued function $\|\cdot\|$ has the following properties*

$$\begin{array}{ll} a) & \|A\| \geq 0 \\ b) & \|A\| = 0 \quad \textit{if and only if} \quad A \equiv 0. \\ c) & \|kA\| = |k|\|A\| \quad \textit{for all } k \in \mathbb{R} \\ d) & \|A+B\| \leq \|A\| + \|B\| \\ e) & \|AB\| \leq \|A\|\|B\| \end{array}$$

Based on the already presented vector norms we can in fact define new matrix norms. The following result can be shown,

Theorem 7. *If $\|\cdot\|$ is any vector norm then*

$$\|A\| = \max_{\|\mathbf{x}\|=1} \|Ax\|$$

is the corresponding induced matrix norm.

Therefore we can now refer to the following types of matrix norms,

$$\|A\|_2 = \max_{\|\mathbf{x}\|_2=1} \|A\mathbf{x}\|_2 \quad \text{or} \quad \|A\|_\infty = \max_{\|\mathbf{x}\|_\infty=1} \|A\mathbf{x}\|_\infty$$

Note that in fact these norms are not very easy to calculate since you must examine all possible vector of length 1 and find the maximum possible result. We present instead a result which allows us to easily calculate one of these matrix norms,

Theorem 8. *Suppose that A is an $m \times n$ matrix. Then the $\|A\|_\infty$ norm is calculated as,*

$$\|A\|_\infty = \max_{1 \le i \le n} \sum_{j=1}^{m} |a_{ij}|$$

Similarly the $\|A\|_1$ is found from,

$$\|A\|_1 = \max_{1 \le j \le m} \sum_{i=1}^{n} |a_{ij}|$$

Let us look at an example using this norm:
Example
Suppose the following matrix is given,

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

Calculate $\|A\|_\infty$.
Solution
The norm is simply found by simply adding all elements in each row and obtaining the maximum result from there,

$$\begin{aligned} |1| + |2| &= 3 \\ |3| + |4| &= 7 \end{aligned}$$

Thus

$$\|A\|_\infty = \max\{3, 7\} = 7,$$

while (check this!)

$$\|A\|_1 = 6.$$

1.0.1 Eigenvalues, eigenvectors and condition number

As you noticed it was not easy to obtain the matrix norm for some cases such as $\|A\|_2$. For this kind of norm we can develop another method which will allow us to obtain the value of the matrix norm in an easy way. This method relies on eigenvalues and eigenvectors.

First some definitions.

Definition 9. *Suppose that A is an $n \times n$ square matrix. Then the following polynomial*

$$p(\lambda) = \det(A - \lambda I)$$

is the so called characteristic polynomial.

Note that $p(\lambda)$ is an nth degree polynomial

Definition 10. *Suppose that the characteristic polynomial p as defined above. The zeros of p are called eigenvalues for the matrix A. If for a given eigenvalue λ we have that $(A - \lambda I)x = 0$ with $x \neq 0$ then x is called an eigenvector corresponding to the eigenvalue λ.*

Let us look at a simple example on how to obtain the eigenvalues and eigenvectors for a given matrix A.

Example

Suppose the following matrix is given,

$$A = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$$

Find the eigenvalues and eigenvectors for A.

Solution

To find the eigenvalues we solve the following characteristic polynomial $p(\lambda)$ based on Definition 9,

$$(1-\lambda)(4-\lambda)+2=0$$

The polynomial simplifies to $\lambda^2 - 5\lambda + 6 = 0$ and the solutions are $\lambda_1 = 2$ and $\lambda_2 = 3$. To find the corresponding eigenvectors we must solve the following matrix system for each eigenvalue,

$$\begin{bmatrix} 1-\lambda & 1 \\ -2 & 4-\lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where here $[v_1, v_2]$ represented the unknown eigenvector. For $\lambda_1 = 2$ we must solve the following system,

$$\begin{bmatrix} -1 & 1 \\ -2 & 2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

which gives that

$$v_1 = v_2$$

Thus an eigenvector for $\lambda_1 = 2$ is $[1, 1]$. Similarly the eigenvector for $\lambda_2 = 3$ is $[1, 2]$

Definition 11. *The spectral radius $\rho(A)$ of a given matrix A is given by,*

$$\rho(A) = \max |\lambda|$$

where λ corresponds to the eigenvalues of A.

Now we are finally in position to provide an alternate method for evaluating the matrix norm $\|A\|_2$ based on the eigenvalues of A.

Theorem 12. *Suppose that A is an $n \times n$ matrix. Then*

$$\begin{aligned} \sqrt{\rho(A^t A)} &= \|A\|_2 && (1.2) \\ \rho(A) &\leq \|A\| \quad \textit{for any matrix norm} && (1.3) \\ \frac{1}{\min |\lambda|} &= \|A^{-1}\|_2 \quad \textit{for } A \textit{ symmetric} && (1.4) \end{aligned}$$

Let us see this result in more detail through a numerical example.

Example Given the following matrix A, compute the $\|A\|_2$ norm,

$$A = \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix}$$

Solution Let us outline this procedure. We first calculate A^tA and then find its eigenvalues then according to the theorem above the square root of the maximum eigenvalue will be equal to $\|A\|_2$. Thus we start by first calculating,

$$A^tA = \begin{bmatrix} 0 & 2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 2 & 1 \end{bmatrix} = \begin{bmatrix} 4 & 2 \\ 2 & 2 \end{bmatrix}$$

The eigenvalues for this matrix are $\lambda_1 = 3 + \sqrt{5}$ and $\lambda_2 = 3 - \sqrt{5}$. Thus

$$\|A\|_2 = \sqrt{3 + \sqrt{5}}$$

Another very important quantity for matrices is the condition number. The condition number of a matrix describes how "good" that matrix is. For instance matrices that are not invertible have condition number ∞ which is considered as bad as possible. On the other hand the smaller the condition number the better the matrix behaves in our calculations. The condition number is defined through,

Definition 13. *Suppose $A \in \mathbb{R}^{n,n}$ and that A is nonsingular. Then the condition number of A is given by,*

$$cond(A) = \|A\| \|A^{-1}\|$$

Note that if we use the Euclidean norm and further assume that A is symmetric then the condition number is given from the eigenvalues of A as follows,

$$\text{cond}(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\sqrt{\rho(A^tA)}}{\min|\lambda|} = \frac{\max|\lambda|}{\min|\lambda|}$$

The above calculation is true but, at least here, we do not show the details. Can you show why this is true based on what you learned so far?