Lecture # 3
Orthogonal Matrices and Matrix Norms

We repeat the definition an orthogonal set and orthonormal set.

**Definition 1** A set of $k$ vectors $\{u_1, u_2, \ldots, u_k\}$, where each $u_i \in \mathbb{R}^n$, is said to be an orthogonal with respect to the inner product $(\cdot, \cdot)$ if $(u_i, u_j) = 0$ for $i \neq j$. The set is said to be orthonormal if it is orthogonal and $(u_i, u_i) = 1$ for $i = 1, 2, \ldots, k$.

The definition of an orthogonal matrix is related to the definition for vectors, but with a subtle difference.

**Definition 2** The matrix $U = (u_1, u_2, \ldots, u_k) \in \mathbb{R}^{n \times k}$ whose columns form an orthonormal set is said to be left orthogonal. If $k = n$, that is, $U$ is square, then $U$ is said to be an orthogonal matrix.

Note that the columns of (left) orthogonal matrices are orthonormal, not merely orthogonal. Square complex matrices whose columns form an orthonormal set are called **unitary**.

**Example 1** Here are some common $2 \times 2$ orthogonal matrices:

\[
U = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad U = \sqrt{0.5} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad U = \begin{pmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{pmatrix}, \quad U = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}
\]

Let $x \in \mathbb{R}^n$ then

\[
\|Ux\|_2^2 = (Ux)^T(Ux) = x^T U^T U x = x^T x = \|x\|_2^2
\]
So
\[ \|Ux\|_2 = \|x\|_2. \]
This property is called orthogonal invariance, it is an important and useful property of the two norm and orthogonal transformations. That is, orthogonal transformations DO NOT AFFECT the two-norm, there is no comparable property for the one-norm or \(\infty\)-norm.

The Cauchy-Schwarz inequality given below is quite useful for all inner products.

**Lemma 1 (Cauchy-Schwarz inequality)** Let \((\cdot, \cdot)\) be an inner product in \(\mathbb{R}^n\). Then for all \(x, y \in \mathbb{R}^n\) we have
\[ |(x, y)| \leq (x, x)^{1/2} (y, y)^{1/2}. \] (1)
Moreover, equality in (1) holds if and only if \(x = \alpha y\) for some \(\alpha \in \mathbb{R}\).

In the Euclidean inner product, this is written
\[ |x^T y| \leq \|x\|_2 \|y\|_2. \]

This inequality leads to the following definition of the angle between two vectors relative to an inner product.

**Definition 3** The angle \(\theta\) between the two nonzero vectors \(x, y \in \mathbb{R}^n\) with respect to an inner product \((\cdot, \cdot)\) is given by
\[ \cos \theta = \frac{(x, y)}{(x, x)^{1/2} (y, y)^{1/2}}. \] (2)

With respect to the Euclidean inner product, this reads
\[ \cos \theta = \frac{x^T y}{\|x\|_2 \|y\|_2}. \]

The one-norm and the \(\infty\)-norm share two inequalities similar to the Cauchy-Schwarz inequality.
\[ |x^T y| \leq \|x\|_1 \|y\|_{\infty}, \]
\[ |x^T y| \leq \|x\|_{\infty} \|y\|_1, \]
but these do not lead to any reasonable definition of angle between vectors.
Example 2

\[ x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad y = \begin{pmatrix} -1 \\ 3 \end{pmatrix}. \]

We have

\[ \|x\|_2 = \sqrt{2}, \quad \|y\|_2 = \sqrt{10}, \quad x^T y = 2 \]

\[ \cos \theta = \frac{x^T y}{\|x\|_2 \|y\|_2} = \frac{1}{\sqrt{5}}. \]

Thus

\[ \theta = 1.1071^R = 63.43^\circ \]

is the angle between the two vectors. The three upper bounds on the dot product are

\[ \|x\|_2 \|y\|_2 = \sqrt{20}, \]
\[ \|x\|_1 \|y\|_\infty = 2 \cdot 3 = 6 \]
\[ \|x\|_\infty \|y\|_1 = 1 \cdot 4 = 4. \]

It is possible to come up with examples where each of these three is the smallest bound (or the largest one).

Matrix Norms

We will also need norms for matrices.

**Definition 4** A norm in \( \mathbb{R}^{m \times n} \) is a function \( \| \cdot \| \) mapping \( \mathbb{R}^{m \times n} \) into \( \mathbb{R} \) satisfying the following three axioms

1. \( \|X\| \geq 0; \)
   \( \|X\| = 0 \) if and only if \( X = 0 \), \( X \in \mathbb{R}^{m \times n} \)

2. \( \|\alpha X\| = |\alpha| \|X\| \quad X \in \mathbb{R}^{m \times n}, \alpha \in \mathbb{R} \)

3. \( \|X + Y\| \leq \|X\| + \|Y\| \quad X, Y \in \mathbb{R}^{m \times n}. \)
This definition is isomorphic to the definition of a vector norm on $\mathbb{R}^{mn}$. For example, the Frobenius norm defined by

$$\|X\|_F = \left( \sum_{i=1}^{m} \sum_{j=1}^{n} x_{ij}^2 \right)^{1/2}$$  \hspace{1cm} (3)

is isomorphic to the two-norm on $\mathbb{R}^{mn}$.

Since $X$ represents a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^m$, it is appropriate to define the induced norm $\| \cdot \|_\alpha$ on $\mathbb{R}^{m \times n}$ by

$$\|X\|_\alpha = \sup_{y \neq 0} \frac{\|Xy\|_\alpha}{\|y\|_\alpha}.$$ \hspace{1cm} (4)

It is a simple matter to show that

$$\|X\|_\alpha = \max_{\|y\|_\alpha = 1} \|Xy\|_\alpha.$$ \hspace{1cm} (5)

Note that the maximum is taken over a closed, bounded set, thus we have that

$$\|X\|_\alpha = \|Xy^*\|_\alpha$$ \hspace{1cm} (6)

for some $y^*$ such that $\|y^*\|_\alpha = 1$. The above definition leads to the very useful bound

$$\|Xy\|_\alpha \leq \|X\|_\alpha \|y\|_\alpha$$ \hspace{1cm} (7)

where equality occurs for every vector of the form $\gamma y^*, \gamma \in \mathbb{R}$.

For any induced norm $\| \cdot \|$, the identity matrix $I_n$ for $\mathbb{R}^{n \times n}$ satisfies

$$\|I_n\| = 1.$$ \hspace{1cm} (8)

However, for the Frobenius norm

$$\|I_n\|_F = \sqrt{n},$$

thus it is not an induced norm for any vector norm.

For the one-norm and the $\infty$-norm there are formulas for the corresponding matrix norms and for a vector $y^*$ satisfying (6). The one-norm formula is

$$\|X\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^{m} |x_{ij}|.$$ \hspace{1cm} (9)
If \( j_{\text{max}} \) is the index of a column such that
\[
\|X\|_1 = \sum_{i=1}^{m} |x_{i,j_{\text{max}}}|\]
then \( y^* = e_{j_{\text{max}}} \), the corresponding column of the identity matrix.

The \( \infty \)-norm formula is
\[
\|X\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |x_{ij}|. \tag{10}
\]

If \( i_{\text{max}} \) is the index of a row such that
\[
\|X\|_\infty = \sum_{j=1}^{n} |x_{i_{\text{max}},j}|\]
then the vector \( y^* = (y^*_1, \ldots, y^*_n)^T \) with components
\[
y^*_j = \text{sign}(x_{i_{\text{max}},j})
\]
satisfies (6). Note that \( \|X\|_\infty = \|X^T\|_1 \).

The matrix two-norm does not have a formula like (9) or (10) and all other formulations are really equivalent to (4). Moreover, computing the vector \( y^* \) in (6) is a nontrivial task that we will discuss later.

The induced norms have a convenient property that is important in understanding matrix computations. For \( X \in \mathbb{R}^{m \times n} \) and \( Y \in \mathbb{R}^{n \times s} \) consider
\[
\|XY\|_\alpha.
\]
We have that
\[
\|XY\|_\alpha = \max_{\|z\|_\alpha = 1} \|XYz\|_\alpha \leq \max_{\|z\|_\alpha = 1} \|X\|_\alpha \|Yz\|_\alpha \leq \|X\|_\alpha \max_{\|z\|_\alpha = 1} \|Yz\|_\alpha = \|X\|_\alpha \|Y\|_\alpha.
\]
Thus
\[
\|XY\|_\alpha \leq \|X\|_\alpha \|Y\|_\alpha. \tag{11}
\]
A norm \( \| \cdot \|_\alpha \) (or really family of norms) that satisfies the property (11) is said to be consistent. Since they are induced norms the two-norm, one-norm, and the \( \infty \)-norm are all consistent. The Frobenius norm also satisfies (11).

An example of a matrix norm that is not consistent is given below.
Example 3 Consider the norm \( \| \cdot \|_\beta \) on \( \mathbb{R}^{m \times n} \) given by
\[
\|X\|_\beta = \max_{(i,j)} |x_{ij}|.
\]
This is simply the \( \infty \)-norm applied to \( X \) written out as vector in \( \mathbb{R}^{mn} \). For \( m = n = 2 \), consider
\[
X = Y = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.
\]
Note that
\[
XY = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix}
\]
and thus \( \|XY\|_\beta = 2 > \|X\|_\beta \|Y\|_\beta = 1 \). Clearly, \( \| \cdot \|_\beta \) is not consistent.

Henceforth, we use only consistent norms.

Now we give a numerical example with our four most used norms.

Example 4 Consider
\[
X = \begin{pmatrix} 3 & -2 & 1 \\ 10 & 0 & -16 \\ -3 & 25 & 1 \end{pmatrix}.
\]

It is easily verified that
\[
\|X\|_1 = 27, \quad y_1^* = e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix},
\]
\[
\|X\|_\infty = 29, \quad y_\infty^* = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix},
\]
\[
\|X\|_F = 31.70, \quad \|X\|_2 = 25.46.
\]

The “magic vector” in the two-norm is (to the digits displayed)
\[
y_2^* = \begin{pmatrix} -0.18943 \\ 0.97256 \\ 0.13508 \end{pmatrix}
\]

This will not always be true, but notice that its sign pattern is the same as \( y_\infty^* \) and that its largest component corresponds to the non-zero component of \( y_1^* \).