Lecture # 3
Orthogonal Matrices and Matrix Norms

We repeat the definition an orthogonal set and orthornormal set.

Definition 1 *A set of k vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k\}$, where each $\mathbf{u}_i \in \mathbf{R}^n$, is said to be an orthogonal with respect to the inner product $(\cdot, \cdot)$ if $(\mathbf{u}_i, \mathbf{u}_j) = 0$ for $i \neq j$. The set is said to be orthonormal if it is orthogonal and $(\mathbf{u}_i, \mathbf{u}_i) = 1$ 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 $U = (\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k) \in \mathbf{R}^{n\times k}$ 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 2×2 orthogonal matrices*

$$\begin{aligned}
U &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\
U &= \sqrt{0.5}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \\
U &= \begin{pmatrix} 0.8 & 0.6 \\ 0.6 & -0.8 \end{pmatrix} \\
U &= \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}
\end{aligned}$$

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

$$\begin{aligned}
\|U\mathbf{x}\|_2^2 &= (U\mathbf{x})^T(U\mathbf{x}) \\
&= \mathbf{x}^T U^T U\mathbf{x} \\
&= \mathbf{x}^T\mathbf{x} \\
&= \|\mathbf{x}\|_2^2
\end{aligned}$$

So

$$\|U\mathbf{x}\|_2 = \|\mathbf{x}\|_2.$$

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 ∞-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* $\mathbf{R}^n$. *Then for all* $\mathbf{x},\mathbf{y} \in \mathbf{R}^n$ *we have*

$$|(\mathbf{x},\mathbf{y})| \leq (\mathbf{x},\mathbf{x})^{1/2}\,(\mathbf{y},\mathbf{y})^{1/2}\,. \tag{1}$$

Moreover, equality in (*1*) *holds if and only if* $\mathbf{x} = \alpha\mathbf{y}$ *for some* $\alpha \in \mathbf{R}$.

In the Euclidean inner product, this is written

$$|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|_2\|\mathbf{y}\|_2.$$

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

Definition 3 *The angle* θ *between the two nonzero vectors* $\mathbf{x},\mathbf{y} \in \mathbf{R}^n$ *with respect to an inner product* $(\cdot,\cdot)$ *is given by*

$$\cos\,\theta = \frac{(\mathbf{x},\mathbf{y})}{(\mathbf{x},\mathbf{x})^{1/2}\,(\mathbf{y},\mathbf{y})^{1/2}}. \tag{2}$$

With respect to the Euclidean inner product, this reads

$$\cos\,\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}.$$

The one-norm and the ∞-norm share two inequalities similar to the Cauchy–Schwarz inequality.

$$\begin{aligned} |\mathbf{x}^T\mathbf{y}| &\leq \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty, \\ |\mathbf{x}^T\mathbf{y}| &\leq \|\mathbf{x}\|_\infty\|\mathbf{y}\|_1, \end{aligned}$$

but these do not lead to any reasonable definition of angle between vectors.

Example 2

$$\mathbf{x} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}.$$

We have

$$\|\mathbf{x}\|_2 = \sqrt{2}, \quad \|\mathbf{y}\|_2 = \sqrt{10}, \quad \mathbf{x}^T\mathbf{y} = 2$$
$$\cos\,\theta = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2} = \sqrt{1/5}.$$

Thus

$$\theta = 1.1071^R = 63.43^\circ$$

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

$$\begin{aligned} \|\mathbf{x}\|_2\|\mathbf{y}\|_2 &= \sqrt{20}, \\ \|\mathbf{x}\|_1\|\mathbf{y}\|_\infty &= 2 \cdot 3 = 6 \\ \|\mathbf{x}\|_\infty\|\mathbf{y}\|_1 &= 1 \cdot 4 = 4. \end{aligned}$$

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* $\mathbf{R}^{m\times n}$ *is a function* $\|\cdot\|$ *mapping* $\mathbf{R}^{m\times n}$ *into* $\mathbf{R}$ *satisfying the following three axioms*

1. $\|X\| \geq 0;$
   $\|X\| = 0$ *if and only if* $X = 0, \quad X \in \mathbf{R}^{m\times n}$
2. $\|\alpha X\| = |\alpha|\|X\| \quad X \in \mathbf{R}^{m\times n}, \alpha \in \mathbf{R}$
3. $\|X + Y\| \leq \|X\| + \|Y\| \quad X, Y \in \mathbf{R}^{m\times n}.$

This definition is isomorphic to the definition of a vector norm on $\mathbf{R}^{mn}$. For example, the Frobenius norm defined by

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

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

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

$$\|X\|_\alpha = \sup_{\mathbf{y}\neq 0} \frac{\|X\mathbf{y}\|_\alpha}{\|\mathbf{y}\|_\alpha}. \tag{4}$$

It is a simple matter to show that

$$\|X\|_\alpha = \max_{\|\mathbf{y}\|_\alpha=1} \|X\mathbf{y}\|_\alpha. \tag{5}$$

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

$$\|X\|_\alpha = \|X\mathbf{y}^*\|_\alpha \tag{6}$$

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

$$\|X\mathbf{y}\|_\alpha \leq \|X\|_\alpha \|\mathbf{y}\|_\alpha \tag{7}$$

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

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

$$\|I_n\| = 1. \tag{8}$$

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

$$\|X\|_1 = \max_{1\leq j\leq n} \sum_{i=1}^{m} |x_{ij}|. \tag{9}$$

If j_{max} is the index of a column such that

$$\|X\|_1 = \sum_{i=1}^{m} |x_{i,j_{max}}|$$

then $\mathbf{y}^* = \mathbf{e}_{j_{max}}$, the corresponding column of the identity matrix.

The ∞-norm formula is

$$\|X\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^{n} |x_{ij}|. \tag{10}$$

If i_{max} is the index of a row such that

$$\|X\|_\infty = \sum_{j=1}^{n} |x_{i_{max},j}|$$

then the vector $\mathbf{y}^* = (y_1^*, \ldots, y_n^*)^T$ with components

$$y_j^* = \text{sign}(x_{i_{max},j})$$

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 $\mathbf{y}^*$ 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 $X \in \mathbf{R}^{m \times n}$ and $Y \in \mathbf{R}^{n \times s}$ consider $\|XY\|_\alpha$. We have that

$$\|XY\|_\alpha = \max_{\|\mathbf{z}\|_\alpha = 1} \|XY\mathbf{z}\|_\alpha \leq \max_{\|\mathbf{z}\|_\alpha = 1} \|X\|_\alpha \|Y\mathbf{z}\|_\alpha$$

$$= \|X\|_\alpha \max_{\|\mathbf{z}\|_\alpha = 1} \|Y\mathbf{z}\|_\alpha = \|X\|_\alpha \|Y\|_\alpha.$$

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 ∞-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* $\mathbf{R}^{m\times n}$ *given by*

$$\|X\|_\beta = \max_{(i,j)} |x_{ij}|.$$

This is simply the ∞*-norm applied to* X *written out as vector in* $\mathbf{R}^{mn}$*. 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

$$\begin{aligned} \|X\|_1 &= 27, \quad \mathbf{y}_1^* = \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \\ \|X\|_\infty &= 29. \quad \mathbf{y}_\infty^* = \begin{pmatrix} -1 \\ 1 \\ 1 \end{pmatrix}, \\ \|X\|_F &= 31.70, \quad \|X\|_2 = 25.46. \end{aligned}$$

The "magic vector" in the two-norm is (to the digits displayed)

$$\mathbf{y}_2^* = \begin{pmatrix} -0.18943 \\ 0.97256 \\ 0.13508 \end{pmatrix}$$

This will not always be true*, but notice that its sign pattern is the same as* $\mathbf{y}_\infty^*$ *and that its largest component corresponds to the non-zero component of* $\mathbf{y}_1^*$*.*