Lecture # 4
Matrix Norms, Orthogonality, and Least Squares

Matrix Norms

We will also need norms for matrices.

Definition 1 *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. $\|X + Y\| \leq \|X\| + \|Y\| \quad X, Y \in \mathbf{R}^{m\times n}$.

The one-norm formula is

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

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

$$\|X\|_\infty = \max_{1\leq i\leq m} \sum_{j=1}^{n} |x_{ij}|. \tag{2}$$

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

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

is a vector that achieves the maximum. Note that $\|X\|_\infty = \|X^T\|_1$.

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

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

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

The matrix two-norm does not have a formula like (1) or (2). Moreover, computing the vector $\mathbf{y}^*$ is a nontrivial task that we will discuss later.

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 $\mathbf{R}^{mn}$.

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

$$\|XY\|_\alpha \le \|X\|_\alpha\|Y\|_\alpha. \tag{4}$$

A norm $\|\cdot\|_\alpha$ (or really family of norms) that satisfies the property (4) 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 (4).

Example 1 *Consider*

$$X = \begin{pmatrix} 3 & -2 & 1 \\ 10 & 0 & -16 \\ -3 & 25 & 1 \end{pmatrix}.$$

It is easily verified that

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

$$\begin{aligned}
\|X\|_F &= = 31.70, \quad \|X\|_2 = 25.46, \\
\mathbf{y}_2^* &= \begin{pmatrix} 0.1894 \\ -0.9726 \\ -0.1351 \end{pmatrix}
\end{aligned}$$

Some inequalities for matrix norms (which we will not prove).

$$(mn)^{-\frac{1}{4}} \left(\|X\|_1 \|X\|_\infty\right)^{1/2} \le \|X\|_2 \le \left(\|X\|_1 \|X\|_\infty\right)^{1/2}, \tag{5}$$

$$\max\left\{\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{m}}\right\} \|X\|_F \le \|X\|_2 \le \|X\|_F, \tag{6}$$

$$\|XY\|_F \le \|X\|_2 \|Y\|_F, \tag{7}$$

$$\|XY\|_F \le \|X\|_F \|Y\|_2. \tag{8}$$

An example of a matrix norm that is not *consistent* is given below.

Example 2 *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.

Also, for any diagonal matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda_2 & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \lambda_{n-1} & 0 \\ \cdots & \cdots & \cdots & 0 & \lambda_n \end{pmatrix},$$

we have

$$\|\Lambda\|_p = \max_{1\le i\le n} |\lambda_i|.$$

Of course,

$$\|\Lambda\|_F = \left(\sum_{i=1}^{n} |\lambda_i|^2\right)^{1/2}.$$

Effects of Orthogonality

First, for any left orthogonal (or orthogonal) matrix $U \in \mathbf{R}^{m\times n}$,

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

This properties extends to the matrix two-norm and to the Frobenius norm as stated in the following lemma.

Lemma 1 *Let $X \in \mathbf{R}^{m\times n}$, $Y \in \mathbf{R}^{n\times k}$ and $\mathbf{y} \in \mathbf{R}^n$. Let $U \in \mathbf{R}^{m\times m}$ and $V \in \mathbf{R}^{n\times n}$ be orthogonal matrices, and let $Z \in \mathbf{R}^{m\times n}$ be a left orthogonal matrix. Then*

$$\|V\mathbf{y}\|_2 = \|\mathbf{y}\|_2, \quad \|Z\mathbf{y}\|_2 = \|\mathbf{y}\|_2, \tag{9}$$

$$\|UXV\|_2 = \|X\|_2, \quad \|UXV\|_F = \|X\|_F, \tag{10}$$

$$\|ZY\|_2 = \|Y\|_2, \quad \|ZY\|_F = \|Y\|_F, \tag{11}$$

$$\|U\|_2 = \|V\|_2 = \|Z\|_2 = 1, \tag{12}$$

$$\|U\|_F = \sqrt{m}, \quad \|V\|_F = \|Z\|_F = \sqrt{n}. \tag{13}$$

The properties (9)–(11) are called *orthogonal invariance.*

Least Squares Problems Let $X \in \mathbf{R}^{m\times n}, m \ge n$ be such that $\operatorname{rank}(X) = n$. That is,

$$X\mathbf{y} = 0, \quad \textbf{iff } \mathbf{y} = 0.$$

Assume that X has a Q–R decomposition, that is,

$$X = Q \begin{matrix} n \\ m-n \end{matrix} \begin{pmatrix} R \\ 0 \end{pmatrix},$$

where $Q \in \mathbf{R}^{m \times m}$ is orthogonal and $R \in \mathbf{R}^{n \times n}$ is upper triangular and nonsingular. We show later how to compute this, but, for now, we just want to understand WHY we should compute it! An upper triangular matrix is just one such

$$R = \begin{pmatrix} r_{11} & r_{12} & \cdots & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & \cdots & r_{2n} \\ 0 & 0 & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & r_{n-1,n-1} & r_{n-1,n} \\ 0 & 0 & \cdots & 0 & r_{nn} \end{pmatrix},$$

thus equations of the form

$$R\mathbf{y} = \mathbf{c}$$

may be solved with back substitution.

We want to solve

$$\min_{\mathbf{y} \in \mathbf{R}^n} \|\mathbf{b} - X\mathbf{y}\|_2^2 \tag{14}$$

We use orthogonal invariance to solve this problem.

$$\begin{aligned} \|\mathbf{b} - X\mathbf{y}\|_2^2 &= \|Q^T(\mathbf{b} - X\mathbf{y})\|_2^2 \\ &= \|Q^T(\mathbf{b} - X\mathbf{y})\|_2^2 \\ &= \|Q^T\mathbf{b} - Q^T X\mathbf{y}\|_2^2 \end{aligned}$$

If we let

$$\mathbf{c} = \begin{matrix} n \\ m-n \end{matrix} \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} = Q^T\mathbf{b},$$

then

$$\begin{aligned} \|\mathbf{b} - X\mathbf{y}\|_2^2 &= \|Q^T\mathbf{b} - Q^T X\mathbf{y}\|_2^2 \\ &= \| \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} - \begin{pmatrix} R \\ 0 \end{pmatrix} \mathbf{y}\|_2^2 \\ &= \| \begin{pmatrix} \mathbf{c}_1 - R\mathbf{y} \\ \mathbf{c}_2 \end{pmatrix} \|_2^2 \\ &= \|\mathbf{c}_1 - R\mathbf{y}\|_2^2 + \|\mathbf{c}_2\|_2^2 \end{aligned}$$

The value of $\mathbf{y}$ has no influence on the second term. The first term is clearly minimized by the solution to

$$R\mathbf{y}_{LS} = \mathbf{c}_1.$$

Moreover (this is an exercise for you),

$$\mathbf{r}_{LS} = \mathbf{b} - X\mathbf{y}_{LS} = Q \begin{pmatrix} 0 \\ \mathbf{c}_2 \end{pmatrix}.$$

.