CME 200 Workshop 5

October 25, 2013

Warm up.

Let A be a 3×3 skew-symmetric matrix, i.e. $A^T = -A$. Find $\det(A)$. (Hint: use properties of determinants.) What if $A \in \mathbb{R}^{4\times4}$? Generalize your result to $A \in \mathbb{R}^{n\times n}$.

Solution.
A general 3×3 skew-symmetric can be written as,

$$A = \begin{bmatrix} 0 & a & b \\ -a & 0 & c \\ -b & -c & 0 \end{bmatrix}$$

We use properties of determinants (specifically, for $A \in \mathbb{R}^{n\times n}$, $\det(A) = \det(A^T)$ and $\det(kA) = k^n \det(A)$), to find $\det(A)$,

$$\det(A) = \det(A^T) = \det(-A) = (-1)^3 \det(A) = -\det(A).$$

Hence, $\det(A) = 0$.
Now, if $A \in \mathbb{R}^{4\times4}$, the same derivation gives,

$$\det(A) = \det(A^T) = \det(-A) = (-1)^4 \det(A) = \det(A).$$

Thus, we can no longer deduce (based on the same derivation) that $\det(A)$ is zero. In fact, we can come up with a simple example of $A \in \mathbb{R}^{4\times4}$, for which the determinant is not zero,

$$A = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}, \qquad \det(A) = 1.$$

In general, the determinant of a skew-symmetric matrix $A \in \mathbb{R}^{n\times n}$ is given by,

$$\det(A) = \begin{cases} 0 & n \text{ odd} \\ d & n \text{ even,} \end{cases}$$

for some $d \in \mathbb{R}$.

Cauchy-Schwarz inequality.

For any vectors $\vec{a},\ \vec{b} \in \mathbb{R}^n$, the following inequality holds,

$$|\vec{a}^T\vec{b}| \leq \|\vec{a}\|_2\|\vec{b}\|_2 \qquad \textbf{(Cauchy-Schwarz)}.$$

Proof.
First, if either of the vectors $\vec{a}$, $\vec{b}$ is a zero vector, then the Cauchy-Schwarz inequality is satisfied trivially. Thus, we assume that neither $\vec{a}$, nor $\vec{b}$ is zero.
Taking $\vec{c} = \|\vec{a}\|_2^2\vec{b} - (\vec{a}^T\vec{b})\vec{a}$, we derive,

$$0 \le \|\vec{c}\|_2^2 = \vec{c}^T\vec{c} = (\|\vec{a}\|_2^2\vec{b} - (\vec{a}^T\vec{b})\vec{a})^T(\|\vec{a}\|_2^2\vec{b} - (\vec{a}^T\vec{b})\vec{a})$$
$$= \|\vec{a}\|_2^4\|\vec{b}\|_2^2 - 2\|\vec{a}\|_2^2(\vec{a}^T\vec{b})^2 + \|\vec{a}\|_2^2(\vec{a}^T\vec{b})^2 = \|\vec{a}\|_2^4\|\vec{b}\|_2^2 - \|\vec{a}\|_2^2(\vec{a}^T\vec{b})^2.$$

Rearranging the inequality and dividing out by $\|\vec{a}\|_2^2$ (which is non-zero since $\vec{a}$ is non-zero), we obtain

$$(\vec{a}^T\vec{b})^2 \le \|\vec{a}\|_2^2\|\vec{b}\|_2^2.$$

Taking the square root, we get the desired Cauchy-Schwarz inequality

$$|\vec{a}^T\vec{b}| \le \|\vec{a}\|_2\|\vec{b}\|_2.$$

[Note: the choice of vector $\vec{c}$ appears to come from nowhere. In our proof we treated this choice as being "magical" and it just happened to give us exactly what we were looking for. There is a proof that derives this vector $\vec{c}$ by looking at $f(\gamma) = \|\vec{a} - \gamma\vec{b}\|_2^2$. However, it requires solving an optimization problem of the function f with respect to γ which we preferred to avoid in favour of a straightforward algebraic derivation.]

Norm compatibility.

Recall that matrix norms fall into two groups: induced norms and non-induced norms (cf. lecture notes 1.5) Given a vector norm $\|\cdot\|$ (e.g. the Euclidean 2-norm $\|\cdot\|_2$, maximum norm $\|\cdot\|_\infty$, L_1 norm $\|\cdot\|_1$, etc.), we define the induced matrix norm (with respect to $\|\cdot\|$) to be

$$\|A\| = \max_{\vec{x}\ne 0} \frac{\|A\vec{x}\|}{\|\vec{x}\|}.$$

On the other hand, Frobenius matrix norm is an example of a non-induced matrix norm (since we cannot represent it using the formula above with respect to some vector norm).
Now, from the definition of the induced norm, we can easily see that $\|A\vec{x}\| \le \|A\|\|\vec{x}\|$,

$$\|A\| = \max_{\vec{x}\ne 0} \frac{\|A\vec{x}\|}{\|\vec{x}\|} \ge \frac{\|A\vec{x}\|}{\|\vec{x}\|}.$$

Here we will prove that the same inequality holds, if instead of the induced norm, we consider a Frobenius matrix norm and the vector 2-norm.

Problem: Consider $A \in \mathbb{R}^{m\times n}$. Using Cauchy-Schwarz inequality, show

$$\|A\vec{x}\|_2 \le \|A\|_F\|\vec{x}\|_2.$$

We say that the Frobenius matrix norm $\|\cdot\|_F$ and the vector 2-norm $\|\cdot\|_2$ are compatible.

Solution.
Recall the definition of the Frobenius norm,

$$\|A\|_F^2 = \sum_i\sum_j |x_{ij}|^2 = \sum_i \vec{r}_i^T\vec{r}_i = \sum_i \|\vec{r}_i\|_2^2.$$

Using Cauchy-Schwartz inequality, $|\vec{x}^T\vec{y}| \le \|\vec{x}\|_2\|\vec{y}\|_2$, we derive,

$$\|A\vec{x}\|_2^2 = \sum_i (\vec{r}_i^T\vec{x})^2 \le \sum_i \|\vec{r}_i\|_2^2\|\vec{x}\|_2^2 = \left(\sum_i \|\vec{r}_i\|_2^2\right)\|\vec{x}\|_2^2 = \|A\|_F^2\|\vec{x}\|_2^2.$$

Full QR vs. Skinny QR.

Find the QR decomposition of A using Gram-Schmidt orthogonalization.

$$A = \begin{bmatrix} 1 & -1 & 4 & 0 \\ 1 & 4 & -2 & 5 \\ 1 & 4 & 2 & 5 \\ 1 & -1 & 0 & 0 \end{bmatrix}.$$

1. $\vec{q}_1 = \vec{a}_1 / \|\vec{a}_1\|_2$.

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad r_{11} = \|\vec{a}_1\|_2 = 2,$$

$$\vec{q}_1 = \frac{\vec{a}_1}{r_{11}} = \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

2. $\vec{w}_2 = \vec{a}_2 - \vec{q}_1(\vec{q}_1^T \vec{a}_2)$
$\vec{q}_2 = \frac{\vec{w}_2}{\|\vec{w}_2\|_2}$.

$$\vec{a}_2 = \begin{bmatrix} -1 \\ 4 \\ 4 \\ -1 \end{bmatrix}, \quad r_{12} = \vec{q}_1^T \vec{a}_2 = 3, \quad \vec{w}_2 = \vec{a}_2 - \vec{q}_1 r_{12} = \begin{bmatrix} -1 \\ 4 \\ 4 \\ -1 \end{bmatrix} - \begin{bmatrix} 3/2 \\ 3/2 \\ 3/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} -5/2 \\ 5/2 \\ 5/2 \\ -5/2 \end{bmatrix}, \quad r_{22} = \|\vec{w}_2\|_2 = 5,$$

$$\vec{q}_2 = \frac{\vec{w}_2}{r_{22}} = \begin{bmatrix} -1/2 \\ 1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$$

3. $\vec{w}_3 = \vec{a}_3 - \vec{q}_1(\vec{q}_1^T \vec{a}_3) - \vec{q}_2(\vec{q}_2^T \vec{a}_3)$
$\vec{q}_3 = \frac{\vec{w}_3}{\|\vec{w}_3\|_2}$.

$$\vec{a}_3 = \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix}, \quad r_{13} = \vec{q}_1^T \vec{a}_3 = 2, \quad r_{23} = \vec{q}_2^T \vec{a}_3 = -2,$$

$$\vec{w}_3 = \vec{a}_3 - \vec{q}_1 r_{13} - \vec{q}_2 r_{23} = \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix} - \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1 \\ -1 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 2 \\ -2 \end{bmatrix}, \quad r_{33} = \|\vec{w}_3\|_2 = 4,$$

$$\vec{q}_3 = \frac{\vec{w}_3}{r_{33}} = \begin{bmatrix} 1/2 \\ -1/2 \\ 1/2 \\ -1/2 \end{bmatrix}$$

4. $\vec{w}_4 = \vec{a}_4 - \vec{q}_1(\vec{q}_1^T\vec{a}_4) - \vec{q}_2(\vec{q}_2^T\vec{a}_4) - \vec{q}_3(\vec{q}_3^T\vec{a}_4)$
   $\vec{q}_4 = \frac{\vec{w}_4}{\|\vec{w}_4\|_2}$.

$$r_{14} = \vec{q}_1^T\vec{a}_4 = 5, \quad r_{24} = \vec{q}_2^T\vec{a}_4 = 5, \quad r_{34} = \vec{q}_3^T\vec{a}_4 = 0,$$

$$\vec{w}_4 = \vec{a}_4 - \vec{q}_1 r_{14} - \vec{q}_2 r_{24} - \vec{q}_3 r_{34} = \vec{0}, \quad r_{44} = \|\vec{w}_4\|_2 = 0,$$

   Since $\vec{w}_4 = \vec{0}$, we know that A is singular and it's column space is spanned by the first 3 orthonormal vectors $\vec{q}_1,\ \vec{q}_2,\ \vec{q}_3$.

Putting r_{ij}'s into matrix form, we have

$$R = \begin{bmatrix} 2 & 3 & 2 & 5 \\ 0 & 5 & -2 & 5 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

We can now form a "skinny" QR decomposition by dropping the last row of R,

$$\tilde{Q} = \begin{bmatrix} 1/2 & -1/2 & 1/2 \\ 1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 \end{bmatrix} \quad \tilde{R} = \begin{bmatrix} 2 & 3 & 2 & 5 \\ 0 & 5 & -2 & 5 \\ 0 & 0 & 4 & 0 \end{bmatrix}$$

$$A = \tilde{Q}\tilde{R} \qquad \text{(Skinny QR)}$$

For the full QR, we need a fourth vector $\vec{q}$ that will be orthonormal to the first three. One way to find such vector is to take a random vector $\vec{a}$ and subtract the components of $\vec{q}_1,\ \vec{q}_2,$ and $\vec{q}_3$. If the result is non-zero, then (after normalizing) we found the fourth vector $\vec{q}_4$. We pick $\vec{e}_3 = [0\ 0\ 1\ 0]^T$ as our "random" vector.

$$\vec{w} = \vec{e}_3 - \vec{q}_1(\vec{q}_1^T\vec{e}_3) - \vec{q}_2(\vec{q}_2^T\vec{e}_3) - \vec{q}_3(\vec{q}_3^T\vec{e}_3) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} - \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix} - \begin{bmatrix} -1/4 \\ 1/4 \\ 1/4 \\ -1/4 \end{bmatrix} - \begin{bmatrix} 1/4 \\ -1/4 \\ 1/4 \\ -1/4 \end{bmatrix} = \begin{bmatrix} -1/4 \\ -1/4 \\ 1/4 \\ 1/4 \end{bmatrix},$$

$\|\vec{w}\|_2 = 1/2$

$$\vec{q}_4 = \vec{w}/\|\vec{w}\|_2 = \begin{bmatrix} -1/2 \\ -1/2 \\ 1/2 \\ 1/2 \end{bmatrix}.$$

Putting $\vec{q}_i$'s into matrix Q and r_{ij}'s into R, we get the desired QR decomposition, Thus, we get the full QR decomposition,

$$Q = \begin{bmatrix} 1/2 & -1/2 & 1/2 & -1/2 \\ 1/2 & 1/2 & -1/2 & -1/2 \\ 1/2 & 1/2 & 1/2 & 1/2 \\ 1/2 & -1/2 & -1/2 & 1/2 \end{bmatrix} \quad R = \begin{bmatrix} 2 & 3 & 2 & 5 \\ 0 & 5 & -2 & 5 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A = QR \qquad \text{(Full QR)}$$

Assignment 3 Review.

Problem 3 (HW3).
Consider a matrix product AB, where A is $m \times n$ and B is $n \times p$. Show that the column space of AB is contained in the column space of A.
Definition. A vector space U is *contained* in another vector space V (denoted as $U \subseteq V$) if every vector $\vec{u} \in U$ ($\vec{u}$ in vector space U) is also in V.
Definition. We say that two vector spaces are *identical* (equal) if $U \subseteq V$ **and** $V \subseteq U$.
(e.g. V is identical to itself since $V \subseteq V$ and $V \subseteq V$.)

Solution.
The most concise proof for this exercise follows by recalling that the column space of A is the same as the range of A (denoted $R(A)$).
Now, if $\vec{y} \in R(AB)$, it means that there exists $\vec{x} \in \mathbb{R}^p$ such that $AB\vec{x} = \vec{y}$ (i.e. $\vec{y}$ is a linear combination of the columns of AB). Using associativity of matrix multiplication, we have,

$$\vec{y} = AB\vec{x} = A(B\vec{x}) = A\vec{w}.$$

This shows that for a vector $\vec{y} \in \mathbb{R}^m$, there is a vector $\vec{w} \in \mathbb{R}^n$, namely $\vec{w} = B\vec{x}$, such that $A\vec{w} = \vec{y}$. This means that $\vec{y}$ can be expressed as a linear combination of the columns of A, i.e. $\vec{y} \in R(A)$.
Since for any vector $\vec{y} \in R(AB)$ we showed that $\vec{y}$ is also in $R(A)$, by the definition above, $R(AB) \subseteq R(A)$.

Problem 6 (HW3).
(a) The nonzero column vectors $\vec{u}$ and $\vec{v}$ have n elements. An $n \times n$ matrix A is given by $A = \vec{u}\vec{v}^T$.
(**Note:** $\vec{u}\vec{v}^T$ is different from the innerproduct $\vec{v}^T\vec{u}$). Show that the rank of A is 1.

Solution.
We argue here by looking at the column space of A. For any vector $\vec{x} \in R^n$, we have

$$\vec{y} = A\vec{x} = \vec{u}\vec{v}^T\vec{x} = (\vec{v}^T\vec{x})\vec{u} = \alpha\vec{u},$$

where $\alpha = \vec{v}^T\vec{x}$ is a scalar. Since $\vec{x}$ was any arbitrary vector in $\mathbb{R}^n$, we know that all vectors in the range space of A are of the form $\vec{y} = \alpha\vec{u}$ with $\vec{u} \neq \vec{0}$, which means that $\text{rank}(A) = \dim(R(A)) = 1$.

Problem 5 (HW3). Let V and W be 3 dimensional subspaces of $\mathbb{R}^5$. Show that V and W must have at least one nonzero vector in common.

Solution.
We prove by contradiction. Let bases of V and W be $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ and $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$, respectively. We assume that V and W are disjoint (that is V and W have no vectors in common). Then any non-trivial linear combination of $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ cannot be expressed as a linear combination of $\{\vec{w}_1, \vec{w}_2, \vec{w}_3\}$. Equivalently, the equation, $A\vec{c} = \vec{0}$,

$$c_1\vec{v}_1 + c_2\vec{v}_2 + c_3\vec{v}_3 + c_4\vec{w}_1 + c_5\vec{w}_2 + c_6\vec{w}_3 = 0$$

can only have a trivial solution $\vec{c} = [c_1, c_2, c_3, c_4, c_5, c_6]^T = \vec{0}$.
Now form a matrix $A = [\vec{v}_1, \vec{v}_2, \vec{v}_3, \vec{w}_1, \vec{w}_2, \vec{w}_3]$. Since A is a 5×6 matrix, its null space is non-trivial, so the linear equation $A\vec{x} = \vec{0}$ must have a non-trivial solution. This contradiction proves that V and W cannot be disjoint.