MATH 21B, MARCH 2: GRAM-SCHMIDT, QR DECOMPOSITION, AND ORTHOGONAL MATRICES

Gram-Schmidt Orthonormalization: Given a set of vectors $\vec{v}_1, \ldots, \vec{v}_n$ in $\mathbb{R}^m$, *Gram-Schmidt* is an algorithm to generate a set of *orthonormal* vectors $\vec{u}_1, \ldots, \vec{u}_n$ which have the *same span.* $\vec{u}_i$ is constructed from $\vec{v}_i$ recursively (starting with $\vec{u}_1$) by letting

$$\vec{w}_i = \vec{v}_i - \mathrm{proj}_{V_{i-1}}(\vec{v}_i) \quad ; \quad \vec{u}_i = \vec{w}_i/||\vec{w}_i||$$

where V_{i-1} is the space spanned by $\vec{u}_1, \ldots, \vec{u}_{i-1}$.

(1) Use Gram-Schmidt to orthonormalize the following sets of vectors.

(a) $\vec{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \vec{v}_2 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}, \vec{v}_3 = \begin{bmatrix} 6 \\ 4 \\ 4 \end{bmatrix}$.

Solution 1: The first step of Gram-Schmidt is to normalize the first vector to be a unit vector.

$$\vec{u}_1 = \frac{\vec{v}_1}{||\vec{v}_1||} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

Now we replace $\vec{v}_2$ by the part orthogonal to the line spanned by $\vec{u}_1$. In class, we called this replacement $\vec{v}_2^{\perp}$, so I'll use that notation here, although it is often written as $\vec{w}_2$ elsewhere.

$$\vec{v}_2^{\perp} = \vec{v}_2 - \mathrm{proj}_{\langle \vec{u}_1 \rangle}(\vec{v}_2) = \vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\vec{u}_1$$

$$= \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$$

Now we normalize.

$$\vec{u}_2 = \frac{\vec{v}_2^{\perp}}{||\vec{v}_2^{\perp}||} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

Finally, we perform the same steps to find $\vec{v}_3^{\perp}$ and $\vec{u}_3$.

$$\vec{v}_3^{\perp} = \vec{v}_3 - \mathrm{proj}_{\langle \vec{u}_1, \vec{u}_2 \rangle}(\vec{v}_3) = \vec{v}_3 - (\vec{u}_1 \cdot \vec{v}_3)\vec{u}_1 - (\vec{u}_2 \cdot \vec{v}_3)\vec{u}_2$$

$$= \begin{bmatrix} 6 \\ 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 4 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 4 \end{bmatrix}$$

$$\vec{u}_3 = \frac{\vec{v}_3^{\perp}}{||\vec{v}_3^{\perp}||} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Thus, we get the orthonormal basis

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

Solution 2: It was asked in class what happens if we went through the vectors in a different order, i.e. say if $\vec{v}_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 6 \\ 4 \\ 4 \end{bmatrix}$, $\vec{v}_3 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix}$. I'll do this example to show that we get a different orthonormal basis. The first vector comes out the same.

$$\vec{u}_1 = \frac{\vec{v}_1}{||\vec{v}_1||} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

$$\vec{v}_2^{\perp} = \vec{v}_2 - (\vec{u}_1 \cdot \vec{v}_2)\vec{u}_1 = \begin{bmatrix} 6 \\ 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 6 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \\ 4 \end{bmatrix}$$

$$\vec{u}_2 = \frac{\vec{v}_2^{\perp}}{||\vec{v}_2^{\perp}||} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$$

$$\vec{v}_3^{\perp} = \vec{v}_3 - (\vec{u}_1 \cdot \vec{v}_3)\vec{u}_1 - (\vec{u}_2 \cdot \vec{v}_3)\vec{u}_2 = \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 3/2 \\ 3/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3/2 \\ -3/2 \end{bmatrix}$$

$$\vec{u}_3 = \frac{\vec{v}_3^{\perp}}{||\vec{v}_3^{\perp}||} = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

and so we get the orthonormal basis

$$\vec{u}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \vec{u}_2 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \vec{u}_3 = \begin{bmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$$

(b) $\vec{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, $\vec{v}_2 = \begin{bmatrix} 3 \\ -1 \\ -1 \\ 3 \end{bmatrix}$, $\vec{v}_3 = \begin{bmatrix} 1 \\ 3 \\ 1 \\ -1 \end{bmatrix}$.

Solution: This was actually the last problem on worksheet 9 - the solution is in the solutions file for that one.

QR decomposition: Let $m \geq n$. Any $m \times n$ matrix A whose columns are linearly independent can be written in the form $A = QR$

$$\begin{bmatrix} | & | & | \\ \vec{v}_1 & \cdots & \vec{v}_n \\ | & | & | \end{bmatrix} = \begin{bmatrix} | & | & | \\ \vec{u}_1 & \cdots & \vec{u}_n \\ | & | & | \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix}$$

where Q is an $m \times n$ matrix whose columns are *orthonormal* and R is an $n \times n$ matrix which is *upper triangular*. The vectors $\vec{u}_1, \ldots, \vec{u}_n$ are obtained from $\vec{v}_1, \ldots, \vec{v}_n$ by *Gram-Schmidt orthonormalization.*

If you resolve the matrix multiplication in the matrix equation above, you get the equations on the left below, which I've rearranged into the equations on the right side.

$$\begin{aligned} \vec{v}_1 = r_{11}\vec{u}_1 &\implies \vec{u}_1 = \frac{\vec{v}_1}{r_{11}} \\ \vec{v}_2 = r_{12}\vec{u}_1 + r_{22}\vec{u}_2 &\implies \vec{u}_2 = \frac{\vec{v}_2 - r_{12}\vec{u}_1}{r_{22}} \\ \vec{v}_3 = r_{13}\vec{u}_1 + r_{23}\vec{u}_2 + r_{33}\vec{u}_3 &\implies \vec{u}_3 = \frac{\vec{v}_3 - r_{13}\vec{u}_1 + r_{23}\vec{u}_2}{r_{33}} \\ &\vdots \end{aligned}$$

Each of the expressions on the right side should look a lot like the manipulations you do in Gram-Schmidt. For example, the first one says that $r_{11} = ||\vec{v}_1||$ is the factor you use to normalize the first vector. The second equation says that $r_{12} = \vec{u}_1 \cdot \vec{v}_2$ is the number you calculate in order to form $\vec{v}_2^{\perp}$, the part perpendicular to $\vec{u}_1$, and $r_{22} = ||\vec{v}_2^{\perp}||$ is the factor you use to normalize the second vector. In general,

$$\boxed{r_{ij} = \vec{u}_i \cdot \vec{v}_j} \quad (i < j) \quad ; \quad \boxed{r_{jj} = ||\vec{v}_j^{\perp}||}$$

The reason you don't have any r_{ij}'s with $i > j$ is because the computation of $v_j^{\perp}$ relies on only the previous vectors $\vec{u}_i$ with $i < j$. These equations allow you to write down the columns of Q and R successively as you perform Gram-Schmidt.

(2) For each set of vectors in (1), write down the QR decomposition of $A = \begin{bmatrix} | & | & | \\ \vec{v}_1 & \vec{v}_2 & \vec{v}_3 \\ | & | & | \end{bmatrix}$.

Solution:

$$\begin{bmatrix} 2 & 2 & 6 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 2 & 6 \\ 0 & 3 & 4 \\ 0 & 0 & 4 \end{bmatrix}$$

When we did Gram-Schmidt on the vectors in the other order,

$$\begin{bmatrix} 2 & 6 & 2 \\ 0 & 4 & 3 \\ 0 & 4 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 2 & 6 & 2 \\ 0 & 4\sqrt{2} & \frac{3\sqrt{2}}{2} \\ 0 & 0 & \frac{3\sqrt{2}}{2} \end{bmatrix}$$

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

Transpose: Let A be an $m \times n$ matrix. Then the *transpose* A^T is an $n \times n$ matrix whose ij-th entry is $(A^T)_{ij} = A_{ji}$. The transpose satisfies the following properties:

- $(AB)^T = B^T A^T$.
- $\vec{v}^T \vec{w}$ is the dot product $\vec{v} \cdot \vec{w}$.
- $\vec{x} \cdot A\vec{y} = A^T \vec{x} \cdot \vec{y}$
- $(A^T)^T = A$
- If A is invertible, $(A^T)^{-1} = (A^{-1})^T$
- $\text{rank}(A) = \text{rank}(A^T)$

(3) Let $A = \begin{bmatrix} 1 & 3 \\ 2 & 1 \\ 7 & 4 \\ 5 & 2 \end{bmatrix}$. Find a matrix B whose kernel is the orthogonal complement of $\text{im}(A)$. What can you say about $\text{rank}(A)$ and $\text{rank}(B)$?

Solution: $B = A^T = \begin{bmatrix} 1 & 2 & 7 & 5 \\ 3 & 1 & 4 & 2 \end{bmatrix}$ has the desired kernel. This is because the kernel of B, is the space of vectors orthogonal to both rows of B (i.e., both columns of A), while the image of A is the span of the columns of A.

The kernel of B has dimension $4 - \dim(\text{im}A)$ (because they are orthogonal complements in $\mathbb{R}^4$, and so the image of B has dimension $4 - (4 - \dim(\text{im}A)) = \dim(\text{im}A)$. Therefore, $\text{im}A$ and $\text{im}B$ have the same dimension - in this case, that dimension is 2.

Orthogonal Matrices: A matrix A is called *orthogonal* if it is square (say $n \times n$) and its columns form an orthonormal basis for $\mathbb{R}^n$. Equivalently, A is orthogonal if $A^T A = I_n$. (why?)

(4) Which of the following transformations are orthogonal? What geometric transformation does each represent? (One is rotation, one is reflection, one is projection, and one is shear.)

$$\frac{1}{5}\begin{bmatrix} 3 & -4 \\ 4 & 3 \end{bmatrix} \qquad \frac{1}{3}\begin{bmatrix} 2 & -2 & -1 \\ -2 & -1 & -2 \\ -1 & -2 & 2 \end{bmatrix} \qquad \frac{1}{9}\begin{bmatrix} 5 & -4 & -2 \\ -4 & 5 & -2 \\ -2 & -2 & 8 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 \\ 0 & 1 \end{bmatrix}$$

Solution: From left to right, they are Rotation, Reflection, Projection, Shear. We can deduce this as follows. First, it is pretty clear that the last one is a shear. For the other three possibilities, notice that rotations and reflections *send any orthonormal basis to another orthonormal basis* - that is, they are *orthogonal transformations*. However, projections do not (since they send some nonzero vectors to the zero vector). It is easy to check that the first two matrices have all columns orthonormal, and thus, the third matrix is a projection. Finally, to distinguish between the first two, remember that if A is a reflection, then A^2 is the identity matrix. The second matrix squares to the identity, the first does not.

(5) Suppose A is an orthogonal $n \times n$ matrix, and $\vec{x}, \vec{y}$ are any vectors in $\mathbb{R}^n$. Argue that $(A\vec{x}) \cdot (A\vec{y}) = \vec{x} \cdot \vec{y}$. Can you give a geometric interpretation of this fact?

Solution: This is saying that A *preserves all lengths and angles.* Here's the way I think about why this is true. Let's suppose that

$$\vec{x} = x_1\vec{e}_1 + \ldots + x_n\vec{e}_n$$
$$\vec{y} = y_1\vec{e}_1 + \ldots + y_n\vec{e}_n$$

i.e., $\vec{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$ and $\vec{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$. Then $\vec{x} \cdot \vec{y} = x_1y_1 + \ldots + x_ny_n$. Meanwhile,

$$A\vec{x} = x_1\vec{u}_1 + \ldots + x_n\vec{u}_n$$
$$A\vec{y} = y_1\vec{u}_1 + \ldots + y_n\vec{u}_n$$

and so $(A\vec{x}) \cdot (A\vec{y}) = x_1y_1 + \ldots + x_ny_n$.

Another, more immediate proof, is to use one of the properties of the transpose:

$$(A\vec{x}) \cdot (A\vec{y}) = (A^TA\vec{x}) \cdot \vec{y} = (I\vec{x}) \cdot \vec{y} = \vec{x} \cdot \vec{y}$$