Recitation 4. March 30

Focus: orthogonal bases, orthogonal matrices, the Gram-Schmidt process, QR factorization

A basis $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$ of a vector space V is called **orthogonal** if:

$$\boldsymbol{q}_i \perp \boldsymbol{q}_j = 0 \quad \Leftrightarrow \quad \boldsymbol{q}_i \cdot \boldsymbol{q}_j = 0 \quad \Leftrightarrow \quad \boldsymbol{q}_i^T \boldsymbol{q}_j = [0]$$

for all $1 \leq i \neq j \leq n$. The basis is called **orthonormal** if it is orthogonal and:

$$||\boldsymbol{q}_i|| = 1 \quad \Leftrightarrow \quad \boldsymbol{q}_i \cdot \boldsymbol{q}_i = 1 \quad \Leftrightarrow \quad \boldsymbol{q}_i^T \boldsymbol{q}_i = [1]$$

for all $1 \leq i \leq n$. A square matrix is called **orthogonal** if its columns form an orthonormal basis, i.e.:

$$Q^T Q = I \quad \Leftrightarrow \quad Q^{-1} = Q^T$$

If Q is a rectangular matrix, the second condition above does not make sense, but $Q^T Q = I$ does and precisely means that the columns of Q are orthonormal vectors. Still, the term "orthogonal matrix" is only applied to square matrices.

Why are matrices with orthonormal columns important? We know that in order to write down the projection matrix onto a subspace V, we need to construct a matrix A whose columns are a basis of V, and then the projection matrix takes the form $P_V = A(A^T A)^{-1} A^T$. This formula is simplified if the basis is taken to be orthonormal (i.e. A has orthonormal columns) because in this case $A^T A = I$ and we don't need to compute any inverses to write down P_V.

Therefore, it's important to have a method to produce orthonormal bases of subspaces, and the **Gram-Schmidt** process precisely does that. The setup is that you have a basis $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$, and you want to transform it into an orthonormal basis $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$. At the i–th step, your basis will take the form $\boldsymbol{q}_1, ..., \boldsymbol{q}_{i-1}, \boldsymbol{v}_i, ..., \boldsymbol{v}_n$ and the goal is to change $\boldsymbol{v}_i$ into some length 1 vector $\boldsymbol{q}_i$ which is perpendicular to $\boldsymbol{q}_1, ..., \boldsymbol{q}_{i-1}$. The way to do so is a two-step process:

- Subtract from $\boldsymbol{v}_i$ a linear combination of $\boldsymbol{q}_1, ..., \boldsymbol{q}_{i-1}$, so that the result is orthogonal to these $i-1$ vectors:

$$\boxed{\boldsymbol{w}_i = \boldsymbol{v}_i - \text{proj}_{\boldsymbol{q}_1} \boldsymbol{v}_i - \cdots - \text{proj}_{\boldsymbol{q}_{i-1}} \boldsymbol{v}_i = \boldsymbol{v}_i - \boldsymbol{q}_1(\boldsymbol{q}_1 \cdot \boldsymbol{v}_i) - \cdots - \boldsymbol{q}_{i-1}(\boldsymbol{q}_{i-1} \cdot \boldsymbol{v}_i)}$$

- Divide $\boldsymbol{w}_i$ by its length, so the result will be a length 1 vector:

$$\boxed{\boldsymbol{q}_i = \frac{\boldsymbol{w}_i}{||\boldsymbol{w}_i||}}$$

Let A be the matrix whose columns are $\boldsymbol{v}_1, ..., \boldsymbol{v}_n$, and let Q be the matrix whose columns are $\boldsymbol{q}_1, ..., \boldsymbol{q}_n$ produced by Gram-Schmidt. We have $Q^T Q = I$ because the columns of Q are orthonormal, by construction. Moreover, we have:

$$\boxed{A = QR}$$

where R is an upper triangular square matrix (in practice, R is a product of elimination and diagonal matrices, according to the steps in the Gram-Schmidt process).

Application of Gram-Schmidt: how to compute a basis for the orthogonal complement $V^\perp$ of a given k-dimensional vector space $V \subset \mathbb{R}^n$? Take an arbitrary basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_k$ of V, and complete it to an arbitrary basis $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ of $\mathbb{R}^n$. Applying Gram-Schmidt to this basis will give you an orthonormal basis $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_n$ of $\mathbb{R}^n$. The first k of these vectors (namely $\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k$) give a basis of V, and the last $n-k$ of these vectors (namely $\boldsymbol{q}_{k+1}, \ldots, \boldsymbol{q}_n$) give a basis of $V^\perp$.

1. Are the following statements true or false? Give arguments in each case.

    - If V and W are orthogonal subspaces, the only vector they have in common is the zero vector.

    - If V and W are orthogonal subspaces, then $V^\perp$ and $W^\perp$ are orthogonal.

Solution:

• TRUE: any vector from V is orthogonal to any vector from W. So if there were a vector $\boldsymbol{a} \in V \cap W$, this would require $\boldsymbol{a} \perp \boldsymbol{a} \Rightarrow ||\boldsymbol{a}|| = 0 \Rightarrow \boldsymbol{a} = 0$.

• FALSE: for example if V and W are perpendicular lines (passing through the origin) in three-dimensional space. Then their orthogonal complements $V^\perp$ and $W^\perp$ would be planes (passing through the origin) in three-dimensional space, so they would intersect in a line. By the previous bullet, this means that $V^\perp$ and $W^\perp$ could not be orthogonal.

However, if V and W are orthogonal complements instead of just orthogonal, then the second bullet would be true: this is because $W = V^\perp$ implies $W^\perp = V$ (try and think of a quick argument) and so $V^\perp = W \perp W^\perp$.

2. Prove that if A and B are orthogonal matrices of the same size, then AB is also orthogonal.

Solution: A and B being orthogonal means that $A^T A = B^T B = I$. Meanwhile:

$$(AB)^T AB = B^T A^T AB = B^T B = I$$

so AB is also orthogonal.

3. Let $\boldsymbol{q}_1, ..., \boldsymbol{q}_k \in \mathbb{R}^n$ be orthonormal vectors. Compute the projection matrix onto the subspace generated by $\boldsymbol{q}_1, ..., \boldsymbol{q}_k$, simplifying the answer as much as possible.

Solution: Let $Q = \left[\ \boldsymbol{q}_1 \ \middle| \ \dots \ \middle| \ \boldsymbol{q}_k\ \right]$ and the assumption states that $Q^T Q = I$. The projection matrix onto the column space of Q is:

$$P_{C(Q)} = Q(Q^T Q)^{-1} Q^T = QQ^T = \begin{bmatrix} \boldsymbol{r}_1 \cdot \boldsymbol{r}_1 & \dots & \boldsymbol{r}_1 \cdot \boldsymbol{r}_n \\ \vdots & \ddots & \vdots \\ \boldsymbol{r}_n \cdot \boldsymbol{r}_1 & \dots & \boldsymbol{r}_n \cdot \boldsymbol{r}_n \end{bmatrix}$$

where $\boldsymbol{r}_1, ..., \boldsymbol{r}_n$ denote the rows of the matrix Q.

4. Use Gram-Schmidt to compute the QR factorization of the matrix:

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

Solution: Let $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, $\boldsymbol{v}_3$ be the columns of A, and let us apply Gram-Schmidt to them. The first step is to renormalize $\boldsymbol{v}_1$ in order for it to have length 1:

$$\boldsymbol{q}_1 = \frac{\boldsymbol{v}_1}{2} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

The next step is to modify $\boldsymbol{v}_2$ so that it becomes perpendicular to $\boldsymbol{q}_1$:

$$\boldsymbol{w}_2 = \boldsymbol{v}_2 - \boldsymbol{q}_1(\boldsymbol{q}_1 \cdot \boldsymbol{v}_2) = \boldsymbol{v}_2 - 3\boldsymbol{q}_1 = \frac{1}{2} \begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

Note that $\boldsymbol{w}_2$ already has length 1, so we set:

$$\boldsymbol{q}_2 = \boldsymbol{w}_2$$

Finally, we modify $\boldsymbol{w}_2$ so that it becomes perpendicular to $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$:

$$\boldsymbol{w}_3 = \boldsymbol{v}_3 - \boldsymbol{q}_1(\boldsymbol{q}_1 \cdot \boldsymbol{v}_3) - \boldsymbol{q}_2(\boldsymbol{q}_2 \cdot \boldsymbol{v}_3) = \boldsymbol{v}_3 - 4\boldsymbol{q}_1 - \boldsymbol{q}_2 = \frac{3}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

We divide $\boldsymbol{w}_3$ by its length so that it has length 1:

$$\boldsymbol{q}_3 = \frac{\boldsymbol{w}_3}{3} = \frac{1}{2}\begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix}$$

Let:

$$Q = \left[\ \boldsymbol{q}_1 \mid \boldsymbol{q}_2 \mid \boldsymbol{q}_3\ \right] = \frac{1}{2}\begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Then the steps in the Gram-Schmidt process read:

$$AD_1^{\left(\frac{1}{2}\right)} E_{12}^{(-3)} D_2^{(1)} E_{13}^{(-4)} E_{23}^{(-1)} D_3^{\left(\frac{1}{3}\right)} = Q$$

Let us move all the diagonal and elimination matrices to the right-hand side by multiplying with their inverses:

$$A = Q\, \underbrace{D_3^{(3)} E_{23}^{(1)} E_{13}^{(4)} D_2^{(1)} E_{12}^{(3)} D_1^{(2)}}_{\text{call this } R}$$

So we got the $A = QR$ factorization, and the R–matrix is explicitly given by:

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