Unit 7: Gram-Schmidt

Lecture

7.1. For vectors in the linear space $\mathbb{R}^n$, the **dot product** is defined as $v \cdot w = \sum_i v_i w_i$. More generally, in the linear space $M(n,m)$ there is a natural dot product $v \cdot w = \text{tr}(v^T w)$, where tr is the trace, the sum of the diagonal entries. It is the sum $\sum_{i,j} v_{ij} w_{ij}$. The dot product allows to compute **length** $|v| = \sqrt{v \cdot v}$ and **angles** α between two vectors defined by the equation $v \cdot w = |v||w|\cos(\alpha)$. If the relation $v \cdot w = 0$ holds, the vectors v and w are called **orthogonal**.

7.2. A collection of pairwise orthogonal vectors $\{v_1, v_2, \cdots, v_n\}$ in $\mathbb{R}^n$ is linearly independent because $a_1 v_1 + \cdots a_n v_n = 0$ implies that $v_k \cdot (a_1 v_1 + \cdots a_n v_n) = a_k v_k \cdot v_k = a_k |v_k|^2 = 0$ and so $a_k = 0$. A collection of n orthogonal vectors therefore automatically forms a basis.

7.3. Definition. A basis is called **orthonormal** if all vectors have length 1 and are orthogonal. Why do we like to have an orthogonal basis? One reason is that an orthogonal basis looks like the standard basis. Another reason is that rotations preserving a space V or orthogonal projections onto a space V are easier to describe if an orthogonal basis is known on V. Let's look at projections as we will need them to produce an orthonormal basis. Remember that the **projection** of a vector x onto a unit vector v is $(v \cdot x)v$. We can now give the matrix of a projection onto a space V if we know an orthonormal basis in V:

> **Lemma:** If $\mathcal{B} = \{v_1, v_2, \cdots, v_n\}$ is an orthonormal basis in V, then the projection P onto V satisfies $Px = (v_1 \cdot x)v_1 + \cdots + (v_n \cdot x)v_n$.

Proof. By Pythagoras, $(x - Px) \cdot x = |x|^2 - (v_1 \cdot x)^2 - \cdots - (v_n \cdot x)^2 = 0$, so that $x - Px$ is perpendicular to x.

Let Q be the matrix containing the basis v_k as columns. We can rewrite the result as $\boxed{P = QQ^T}$. We write Q because it is not a $n \times n$ matrix like S. The matrix Q contains the basis of the subspace V and not the basis of the entire space. We will see next week a more general formula for P which also holds if the vectors are not perpendicular.

7.4. Let $v_1, \dots, v_n$ be a basis in V. Let $w_1 = v_1$ and $u_1 = w_1/|w_1|$. The **Gram-Schmidt process** recursively constructs from the already constructed orthonormal set $u_1, \dots, u_{i-1}$ which spans a linear space V_{i-1} the new vector $w_i = (v_i - \text{proj}_{V_{i-1}}(v_i))$ which is orthogonal to V_{i-1}, and then normalizes w_i to get $u_i = w_i/|w_i|$. Each vector w_i is orthonormal to the linear space V_{i-1}. The vectors $\{u_1, \dots, u_n\ \}$ form then an orthonormal basis in V.

7.5. The formulas can be written as

$$v_1 = |w_1|u_1 = r_{11}u_1$$

...

$$v_i = (u_1 \cdot v_i)u_1 + \cdots + (u_{i-1} \cdot v_i)u_{i-1} + |w_i|u_i = r_{1i}u_1 + \cdots + r_{ii}u_i$$

...

$$v_n = (u_1 \cdot v_n)u_1 + \cdots + (u_{n-1} \cdot v_n)u_{n-1} + |w_n|u_n = r_{1n}u_1 + \cdots + r_{nn}u_n\ .$$

In matrix form this means

$$A = \begin{bmatrix} | & | & \cdot & | \\ v_1 & v_2 & \cdot & v_n \\ | & | & \cdot & | \end{bmatrix} = \begin{bmatrix} | & | & \cdot & | \\ u_1 & u_2 & \cdot & u_n \\ | & | & \cdot & | \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdot & r_{1n} \\ 0 & r_{22} & \cdot & r_{2n} \\ 0 & 0 & \cdot & r_{nn} \end{bmatrix} = QR\ ,$$

where A and Q are $m \times n$ matrices and R is a $n \times n$ matrix with

$$\boxed{r_{ij} = v_j \cdot u_i, \text{ for } i < j \text{ and } v_{ii} = |w_i|}$$

We have just seen:

Theorem: A matrix A with linearly independent columns v_i can be decomposed as $A = QR$, where Q has orthonormal column vectors and where R is an upper triangular square matrix with the same number of columns than A. The matrix Q has the orthonormal vectors u_i in the columns.

7.6. The recursive process was stated first by Erhard Schmidt (1876-1959) in 1907. The essence of the formula was already in a 1883 paper by J.P.Gram in 1883 which Schmidt mentions in a footnote. The process seems to already have been anticipated by Laplace (1749-1827) and was also used by Cauchy (1789-1857) in 1836.

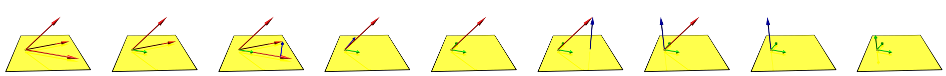

FIGURE 1.

EXAMPLES

7.7. Problem. Use Gram-Schmidt on $\{v_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, v_2 = \begin{bmatrix} 1 \\ 3 \\ 0 \end{bmatrix}, v_3 = \begin{bmatrix} 1 \\ 2 \\ 5 \end{bmatrix}\}$.

Solution. 1. $w_1 = \frac{v_1}{|v_1|} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $u_1 = w_1$.

2. $w_2 = (v_2 - \text{proj}_{V_1}(v_2)) = v_2 - (u_1 \cdot v_2)u_1 = \begin{bmatrix} 0 \\ 3 \\ 0 \end{bmatrix}$. $u_2 = \frac{w_2}{|w_2|} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

3. $w_3 = (v_3 - \text{proj}_{V_2}(v_3)) = v_3 - (u_1 \cdot v_3)u_1 - (u_2 \cdot v_3)u_2 = \begin{bmatrix} 0 \\ 0 \\ 5 \end{bmatrix}$, $u_3 = \frac{w_3}{|w_3|} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

7.8. From $v_1 = |v_1|u_1$, $v_2 = (u_1 \cdot v_2)u_1 + |w_2|u_2$, and $v_3 = (u_1 \cdot v_3)u_1 + (u_2 \cdot v_3)u_2 + |w_3|u_3$, we get the QR decomposition

$$A = \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 & 1 \\ 0 & 3 & 2 \\ 0 & 0 & 5 \end{bmatrix} = QR .$$

7.9. One reason why we are interested in orthogonality is that in statistics, "orthogonal" means "uncorrelated". **Data** are often also arranged in matrices as relational databases. Let's take the matrices $v_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$ and $v_2 = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix}$. They span a two dimensional plane in the linear space $M(2,2)$ of 2×2 matrices. We want to have an orthogonal set of vectors in that plane. Now, how do we do that? We can use Gram-Schmidt in the same way as with vectors in $\mathbb{R}^n$. One possibility is to write the matrices as vectors like $v_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \end{bmatrix}^T$ and $v_2 = \begin{bmatrix} 0 & 3 & 3 & 0 \end{bmatrix}^T$ and proceed with vectors. But we can also remain within matrices and do the Gram-Schmidt procedure in $M(2,2)$. Let us do that. The first step is to normalize the first vector. We get $u_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}/2$. The second step is to produce $w_2 = v_2 - (u_1 \cdot v_2)u_1 = \begin{bmatrix} 0 & 3 \\ 3 & 0 \end{bmatrix} - 3\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}/2 = \begin{bmatrix} -3/2 & 3/2 \\ 3/2 & -3/2 \end{bmatrix}$. Now, we have to normalize this to get $u_2 = w_2/|w_2| = w_2/3 = \begin{bmatrix} -1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}$. Now, $\mathcal{B} = \{u_1, u_2\}$ is an orthonormal basis in the space X spanned by $\{v_1, v_2\}$.

7.10.

Theorem: If $S^T = S^{-1}$ the map $T : A \to S^{-1}AS$ is an orthogonal transformation from $M(n,n) \to M(n,n)$.

Proof. The dot product between $A_1 = S^{-1}AS$ and $B_1 = S^{-1}BS$ is equal to the one between A and B:

$$\text{tr}(A_1^T B_1) = \text{tr}((S^{-1}AS)^T S^{-1}BS) = \text{tr}((S^{-1}A^T SS^{-1}BS) = \text{tr}(S^{-1}A^T BS) = \text{tr}(A^T B) .$$

□

We have used in the last step that similar matrices always have the same trace. We prove this later. For 2×2 matrices we can check it by brute force:

```
A={{a,b},{c,d}};S={{p,q},{r,s}};
Simplify[Tr[Inverse[S].A.S]==Tr[A]]
```

Homework

This homework is due on Tuesday, 2/20/2019.

Problem 7.1: Perform the Gram-Schmidt process on the three vectors $\{\begin{bmatrix} 5 \\ 5 \\ 5 \\ 5 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 4 \\ 2 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}\}$ and then write down the QR decomposition.

Problem 7.2: a) Find an orthonormal basis of the plane $x + y + z = 0$ and form the projection matrix $P = QQ^T$.
b) Find an orthonormal basis of the hyper plane $x_1 + x_2 + x_3 + x_4 + x_5 = 0$ in $\mathbf{R}^5$.

Problem 7.3: a) Produce an orthonormal basis of the kernel of

$$A = \begin{bmatrix} 1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & 1 & 1 \end{bmatrix} .$$

b) Write down an orthonormal basis for the image of A.

Problem 7.4: Find the QR factorization of the following matrices

$$A = \begin{bmatrix} 2 & 4 \\ 1 & 1 \end{bmatrix}, B = \begin{bmatrix} 0 & -3 & 0 \\ 0 & 0 & 0 \\ 7 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}, C = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, D = \begin{bmatrix} 12 & 5 \\ -5 & 12 \end{bmatrix} .$$

Problem 7.5: Find the QR factorization of the following matrices (D was the quaternion matrix you have derived earlier)

$$A = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 0 & 5 \end{bmatrix} ,$$

$$B = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 4 & 0 \\ 0 & 0 & 0 & 4 \end{bmatrix}, C = \begin{bmatrix} 5 & 12 \\ 12 & -5 \end{bmatrix} .$$

$$D = \begin{bmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{bmatrix}$$