UGC 12:55 W. LCB 225

Arctic

Math 2270-1

Notes of 11/11/19

- Quick review: Two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$ are **orthogonal** if

$$\mathbf{u} \bullet \mathbf{v} = 0.$$

  They are **orthonormal** if they are orthogonal, and are also unit vectors:

$$\|\mathbf{u}\| = \|\mathbf{v}\| = 1.$$

  An **orthogonal set**

$$\{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$$

  is a set whose vectors are pairwise orthogonal, i.e.,

$$i \neq j \quad \Longrightarrow \quad \mathbf{u}_i \bullet \mathbf{u}_j = 0.$$

  An orthogonal set is an **orthonormal set** if it is orthogonal and all of its vectors are unit vectors. An orthogonal (**orthonormal**) **basis** is a basis that is also an **orthogonal** (**orthonormal**) set.

- Orthogonal Bases are nice. Suppose

$$\mathrm{B} = \{\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_p\}$$

  is an orthogonal basis of a subspace W of $\mathbb{R}^n$ and $\mathbf{y}$ is in $\mathbb{R}^n$. Then

$$\hat{\mathbf{y}} = \sum_{i=1}^{p} \frac{\mathbf{y} \bullet \mathbf{u}_i}{\mathbf{u}_i \bullet \mathbf{u}_i} \mathbf{u}_i \tag{1}$$

is the **orthogonal projection of y onto** W**.** In the special case that $\mathbf{y}$ is in W then of course $\mathbf{y} = \hat{\mathbf{y}}$:

$$\mathbf{y} = \sum_{i=1}^{\cancel{p}\ n} \frac{\mathbf{y} \bullet \mathbf{u}_i}{\mathbf{u}_i \bullet \mathbf{u}_i} \mathbf{u}_i \qquad (2)$$

- In the special case that B is orthonormal the formulas (1) and (2) simplify to

$$\hat{\mathbf{y}} = \sum_{i=1}^{p} (\mathbf{y} \bullet \mathbf{u}_i) \mathbf{u}_i. \qquad (3)$$

and

$$\overset{\cancel{\wedge}}{\mathbf{y}} = \sum_{i=1}^{\cancel{p}\ n} (\mathbf{y} \bullet \mathbf{u}_i) \mathbf{u}_i. \qquad (4)$$

- Any vector $\mathbf{w}$ can be changed to a unit vector $\mathbf{v}$ in the same direction by dividing $\mathbf{w}$ by its norm:

$$\mathbf{v} = \frac{\mathbf{w}}{\|\mathbf{w}\|}.$$

- Today's topic is how to construct an orthogonal or orthonormal basis.

6.4 The Gram-Schmidt Process

- According to the wikipedia, the Gram-Schmidt Process is named after Jorgen Pedersen Gram (1850-1916) and Erhard Schmidt (1876-1959), but was already known to Pierre-Simon Laplace (1749-1827).
- Given a linear space W, how do we construct an orthogonal or orthonormal basis?
- The Gram-Schmidt Process is one way to answer that question.
- It's based on the assumption that we already have some basis of W and uses that basis to construct an orthogonal basis.
- We start with Example 2 from the textbook (page 356).
- Let

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}$$

  and

$$W = \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}.$$

  Construct an orthogonal basis n

$$\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$$

  of W.

- Actually we are going to do a little more: we are going to get a set of orthogonal vectors $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ such that

$$\begin{aligned} \operatorname{span}\{\mathbf{x}_1\} &= \operatorname{span}\{\mathbf{v}_1\}, \\ \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2\} &= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}, \\ \operatorname{span}\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\} &= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}. \end{aligned} \tag{5}$$

- The order of the basis vectors matters. We have a sequence of nested vector spaces, and nested bases of those spaces.

- Recall that the requirement that the $\mathbf{v}_i$ be orthogonal means that

$$\mathbf{v}_1 \bullet \mathbf{v}_2 = \mathbf{v}_1 \bullet \mathbf{v}_3 = \mathbf{v}_2 \bullet \mathbf{v}_3 = 0. \tag{6}$$

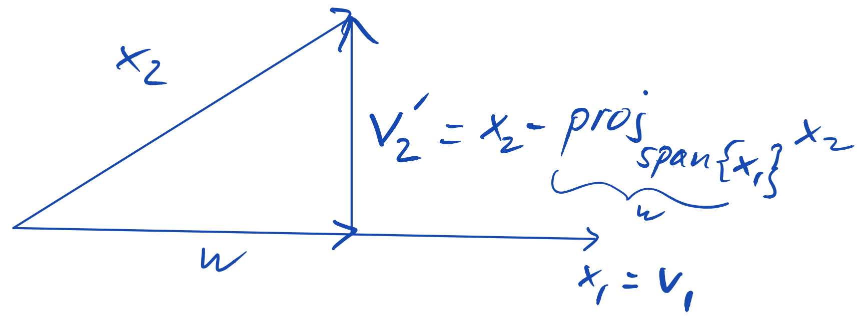
- Clearly, the requirements (5) and (6) do not determine the vectors $\mathbf{v}_1$, $\mathbf{v}_2$, and $\mathbf{v}_3$ uniquely. We can multiply them with non-zero constants and the properties (5) and (6) will still be satisfied.

- A common requirement is that the $\mathbf{v}_i$ be **unit vectors**, i.e., $\|\mathbf{v}_i\| = 1$.

- The textbook also suggest the possibility of making the entries of the $\mathbf{v}_i$ integer to simplify hand calculations. For this example, let's follow that suggestion.

$$x_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix} \quad x_2 = \begin{bmatrix}0\\1\\1\\1\end{bmatrix} \quad x_3 = \begin{bmatrix}0\\0\\1\\1\end{bmatrix}$$

$$V_1 = x_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$



$$V_2' = x_2 - \underbrace{\frac{x_2 \cdot V_1}{V_1 \cdot V_1} V_1}_{w}$$

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

$$V_2 = \begin{bmatrix}-3\\1\\1\\1\end{bmatrix} \qquad V_2 \cdot V_1 = \begin{bmatrix}-3\\1\\1\\1\end{bmatrix} \cdot \begin{bmatrix}1\\1\\1\\1\end{bmatrix} = 0$$

$$V_3' = x_3 - \text{proj}_{\text{span}\{V_1, V_2\}} x_3$$

$$= \begin{bmatrix}0\\0\\1\\1\end{bmatrix} - \frac{V_1 \cdot x_3}{V_1 \cdot V_1} V_1 - \frac{V_2 \cdot x_3}{V_2 \cdot V_2} V_2$$

$$= \begin{bmatrix}0\\0\\1\\1\end{bmatrix} - \underset{1/2}{\frac{2}{4}}\begin{bmatrix}1\\1\\1\\1\end{bmatrix} - \underset{1/6}{\frac{2}{12}}\begin{bmatrix}-3\\1\\1\\1\end{bmatrix}$$

$$= \begin{bmatrix} -1/2 + 1/2 \\ -1/2 - 1/6 \\ 1 - 1/2 - 1/6 \\ 1 - 1/2 - 1/6 \end{bmatrix} = \begin{bmatrix} 0 \\ -2/3 \\ 1/3 \\ 1/3 \end{bmatrix} \qquad v_3 = \begin{bmatrix} 0 \\ -2 \\ 1 \\ 1 \end{bmatrix}$$

- The general process is described by **Theorem 11**, page 357, textbook: Given a basis

$$\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_p\}$$

for a non-zero subspace W of $\mathbb{R}^n$, define

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_p &= \mathbf{x}_p - \frac{\mathbf{x}_p \bullet \mathbf{v}_1}{\mathbf{v}_1 \bullet \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_p \bullet \mathbf{v}_2}{\mathbf{v}_2 \bullet \mathbf{v}_2}\mathbf{v}_2 - \ldots - \frac{\mathbf{x}_p \bullet \mathbf{v}_{p-1}}{\mathbf{v}_{p-1} \bullet \mathbf{v}_{p-1}}\mathbf{v}_{p-1}
\end{aligned}$$

Then $\{\mathbf{v}_1, \ldots, \mathbf{v}_p\}$ is an orthogonal basis for W. In addition,

$$\text{span}\,\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} = \text{span}\,\{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \quad \text{for} \quad k = 1, 2, \ldots, p.$$

- This is version 1 of the Gram-Schmidt Process.

- Version 2 is just a more compact notation for the process. For $k = 1, \ldots, p$ define

$$\mathbf{v}_k = \mathbf{x}_k - \sum_{i=1}^{k-1} \frac{\mathbf{x}_k \bullet \mathbf{v}_i}{\mathbf{v}_i \bullet \mathbf{v}_i} \mathbf{v}_i. \tag{7}$$

- This works even for $k = 1$ if you use the convention that the empty sum is zero.
- In most references you will find a description of the Gram-Schmidt process that defines **orthonormal vectors** $\mathbf{v}_1, \mathbf{v}_2, \ldots$. This means that the $\mathbf{v}_i$ are orthogonal, and that they are also unit vectors. In the usual description of the Gram Schmidt process the $\mathbf{v}_i$ are normalized as soon as they are computed. Thus the denominators $\mathbf{v}_i \bullet \mathbf{v}_i$ equal 1, and disappear.
- This gives rise to Version 3 of the Gram-Schmidt Process:

  For $k = 1, \ldots, p$ define

$$\left\{ \begin{array}{lcl} \mathbf{w}_k & = & \mathbf{x}_k - \displaystyle\sum_{i=1}^{k-1} (\mathbf{x}_k \bullet \mathbf{v}_i) \mathbf{v}_i \\ \mathbf{v}_k & = & \dfrac{\mathbf{w}_k}{\|\mathbf{w}_k\|} \end{array} \right. \tag{8}$$

The QR Factorization

- Orthonormal bases are particularly important if they form the columns of a matrix.
- Definition: A square matrix Q is **orthogonal** if its columns form an **orthonormal** set.
- This means that

$$Q^TQ = I,$$

i.e., Q is invertible, and

$$Q^{-1} = Q^T.$$

(see textbook, page 346.)

An orthogonal matrix should really be called **orthonormal**, but it's not. There is no such thing as an orthonormal matrix.

- Examples of orthogonal matrices include:
  - The identity matrix
  - A rotation matrix:

$c = \cos\theta$

$s = \sin\theta$

$$\begin{bmatrix} c & s \\ -s & c \end{bmatrix} = R^T$$

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} c^2+s^2 & sc-sc \\ cs-sc & s^2+c^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

  - A permutation matrix.
  - A **Householder Reflection**, i.e., a matrix of the form

$$H = I - 2\mathbf{u}\mathbf{u}^T$$

$$H^T = (I - 2uu^T)^T = I^T - 2(uu^T)^T$$
$$= I - 2uu^T$$
$$= H$$

H is symmetric

$Hx \quad x=u \quad Hu = Iu - 2u\underbrace{u^T u}_{1}$

$= u - 2u = -u$

$u^T v = 0 \qquad v = \left(\text{span}\{u\}\right)^{\perp}$

$Hv = Iv - 2uu^T v = v$

$x = \alpha u + v \qquad v^T u = 0$

$Hx = \alpha Hu + Hv$

$= -\alpha u + v$

$H^T H = (I - 2uu^T)^T (I - 2uu^T)$

$= I^T I - 4uu^T + 4u\underbrace{u^T u}_{1}u^T$

$= I$

where $\mathbf{u}$ is a unit vector.

- It is of course easy to come up with a non-square matrix whose columns form an orthonormal set. For example, pick a subset of the columns of an orthogonal matrix. The columns are still orthonormal.

However, there is no generally accepted name for that kind of matrix. So every time we use a matrix like this we have to describe it as in

- **Theorem 12**, page 359, textbook. If A is and $m \times n$ matrix with linearly independent columns, then A can be factored a

$$A = QR \qquad [a_1, a_2, \dots, a_n] = [q_1, q_2, \dots, q_n]$$

where Q is an $m \times n$ matrix whose columns form an orthonormal basis for $\mathrm{Col}(A)$ and R is an $n \times n$ upper triangular invertible matrix with positive entries on its diagonal.

$$\begin{bmatrix} x & x & & \\ 0 & x & & \\ 0 & 0 & \ddots & \\ 0 & 0 & & x \end{bmatrix}$$

- Let's first revisit Example 1.

R triangular
Q orthogonal
(has orthonormal columns)

- The ideas apply in general. The matrix Q can be constructed in several different ways, including the Gram-Schmidt process. Once we have Q we can construct R by the observation that

$$A = QR \qquad \Longleftrightarrow \qquad R = Q^T A.$$

But read the numerical notes on page 360 of the textbook.