16. Lecture 16 5.2 The Gram-Schmidt Orthogonalization Process

Here we will learn a process for constructing an orthonormal basis for subspace W of $\mathbf{R}^m$. Starting from any basis $\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$ for W we construct an orthonormal basis $\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$. We will construct the $\mathbf{u}_i$'s inductively so that $\{\mathbf{u}_1,\cdots,\mathbf{u}_k\}$ are orthonormal and

$$\text{Span}(\mathbf{u}_1,\cdots,\mathbf{u}_k) = \text{Span}(\mathbf{v}_1,\cdots,\mathbf{v}_k)$$

for $k = 1,...,n$. To begin the process, let

$$\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\ \mathbf{v}_1$$

Then $\text{Span}(\mathbf{u}_1) = \text{Span}(\mathbf{v}_1)$, since $\mathbf{u}_1$ is a multiple of $\mathbf{v}_1$ and $\|\mathbf{u}_1\| = 1$.
Let $\mathbf{p}_1$ be the projection of $\mathbf{v}_2$ onto $\text{Span}(\mathbf{v}_1) = \text{Span}(\mathbf{u}_1)$, i.e.

$$\mathbf{p}_1 = (\mathbf{v}_2\cdot\mathbf{u}_1)\,\mathbf{u}_1, \qquad \mathbf{v}_2 - \mathbf{p}_1 \in \text{Span}(\mathbf{u}_1)^\perp$$

Then $\mathbf{v}_2 - \mathbf{p}_1 \neq \mathbf{0}$ since $\mathbf{v}_2 \notin \text{Span}(\mathbf{u}_1)$. If we set

$$\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2 - \mathbf{p}_1\|}\ (\mathbf{v}_2 - \mathbf{p}_1)$$

then $\mathbf{u}_2$ is a unit vector orthogonal to $\text{Span}(\mathbf{u}_1)$ and $\text{Span}(\mathbf{u}_1,\mathbf{u}_2) = \text{Span}(\mathbf{v}_1,\mathbf{v}_2)$.
To construct $\mathbf{u}_3$ let $\mathbf{p}_3$ be the projection of $\mathbf{v}_3$ into $\text{Span}\{\mathbf{u}_1,\mathbf{u}_2\}$:

$$\mathbf{p}_2 = (\mathbf{v}_3\cdot\mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{v}_3\cdot\mathbf{u}_2)\,\mathbf{u}_2$$

and set

$$\mathbf{u}_3 = \frac{1}{\|\mathbf{v}_3 - \mathbf{p}_2\|}\ \big(\mathbf{v}_3 - \mathbf{p}_2\big)$$

In general we define $\mathbf{u}_k$ recursively by

$$\mathbf{u}_{k+1} = \frac{1}{\|\mathbf{v}_{k+1} - \mathbf{p}_k\|}\ \big(\mathbf{v}_{k+1} - \mathbf{p}_k\big)$$

where

$$\mathbf{p}_k = (\mathbf{v}_{k+1}\cdot\mathbf{u}_1)\,\mathbf{u}_1 + \cdots + (\mathbf{v}_{k+1}\cdot\mathbf{u}_k)\,\mathbf{u}_k$$

is the projection of $\mathbf{v}_{k+1}$ onto $\text{Span}(\mathbf{u}_1,\cdots,\mathbf{u}_k)$. This procedure, called the **Gram-Schmidt orthogonalization process** yields an orthonormal basis $\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$ for W.

Ex Find an orthonormal basis for the plane $F = \{\mathbf{x} \in \mathbf{R}^3;\ x_1 + x_2 + x_3 = 0\}$.
Sol $\mathbf{v}_1 = (1,-1,0)^T$ and $\mathbf{v}_2 = (1,0,-1)^T$ are two vectors in the plane.
First let $\mathbf{u}_1 = \frac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix}$. Then let $\mathbf{p}_1 = (\mathbf{v}_2\cdot\mathbf{u}_1)\mathbf{u}_1 = \frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix} = \frac{1}{2}\begin{bmatrix}1\\-1\\0\end{bmatrix}$.
Since $\mathbf{v}_2 - \mathbf{p}_1 = \frac{1}{2}\begin{bmatrix}1\\1\\-2\end{bmatrix}$ we get $\mathbf{u}_2 = \frac{1}{\|\mathbf{v}_2 - \mathbf{p}_1\|}\big(\mathbf{v}_2 - \mathbf{p}_1\big) = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\1\\-2\end{bmatrix}$.

QR factorization

One can also use the Gram-Schmidt process to obtain the so called QR factorization of a matrix $A = QR$, where the column vectors of Q are orthonormal and R is upper triangular. In fact if M is an $m \times n$ matrix such that the n column vectors of $M = \big[\mathbf{v}_1 \cdots \mathbf{v}_n\big]$ form a basis for a subspace W of $\mathbf{R}^m$ we can perform the Gram-Schmidt process on these to obtain an orthonormal basis $\{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ such that $\text{Span}\big(\mathbf{u}_1, \cdots, \mathbf{u}_k\big) = \text{Span}\big(\mathbf{v}_1, \cdots, \mathbf{v}_k\big)$, for $k = 1, ..., n$. Hence for some constants r_{ij}

$$\mathbf{v}_k = r_{1k}\mathbf{u}_1 + \cdots + \mathbf{r}_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots 0\mathbf{u}_n, \qquad k = 1, ..., n.$$

Let R be the upper triangular matrix with column vectors defined by

$$R = \big[\mathbf{r}_1 \cdots \mathbf{r}_n\big], \qquad \text{where} \qquad \mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and let $Q = \big[\mathbf{u}_1 \cdots \mathbf{u}_n\big]$. Then

$$Q\mathbf{r}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k = \mathbf{v}_k$$

and hence

$$QR = \big[Q\mathbf{r}_1 \cdots Q\mathbf{r}_n\big] = \big[\mathbf{v}_1 \cdots \mathbf{v}_n\big] = M.$$

Note that in principle one can calculate what R is from the Gram-Schmidt process:

$$r_{11}\mathbf{u}_1 = \mathbf{v}_1, \qquad r_{11} = \|\mathbf{v}_1\|,$$

$$r_{22}\mathbf{u}_2 = \mathbf{v}_2 - \mathbf{p}_1, \quad \mathbf{p}_1 = r_{12}\mathbf{u}_1, \quad r_{12} = \langle \mathbf{v}_2, \mathbf{u}_1 \rangle, \quad r_{22} = \|\mathbf{v}_2 - \mathbf{p}_1\|,$$

and so on

$$r_{kk}\mathbf{u}_k = \mathbf{v}_k - \mathbf{p}_{k-1}, \quad \mathbf{p}_{k-1} = r_{1k}\mathbf{u}_1 + \cdots + r_{(k-1)k}\mathbf{u}_{k-1}, \quad r_{\ell k} = \langle \mathbf{v}_k, \mathbf{u}_\ell \rangle, \quad r_{kk} = \|\mathbf{v}_k - \mathbf{p}_{k-1}\|.$$

However, it is simpler to get R just from using that $M = QR$ and that $Q^TQ = I$ so that

$$R = Q^TQR = Q^TM.$$

That $Q^TQ = I$ follows from that the columns of Q are orthonormal:

$$Q^TQ = \begin{bmatrix} - & \mathbf{u}_1^T & - \\ - & \mathbf{u}_2^T & - \\ & \vdots & \\ - & \mathbf{u}_n^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \mathbf{u}_1 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_n \\ \mathbf{u}_2 \cdot \mathbf{u}_1 & \mathbf{u}_2 \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_2 \cdot \mathbf{u}_n \\ \vdots & \vdots & & \vdots \\ \mathbf{u}_n \cdot \mathbf{u}_1 & \mathbf{u}_n \cdot \mathbf{u}_2 & \cdots & \mathbf{u}_n \cdot \mathbf{u}_n \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{bmatrix}$$

Ex Find the QR factorization of $M = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 3 \end{bmatrix}$. **Sol** Use Gram Schmidt on the columns of $M = [\mathbf{v}_1\, \mathbf{v}_2]$ to find an orthonormal basis $\{\mathbf{u}_1, \mathbf{u}_2\}$ and from that construct $Q = [\mathbf{u}_1\, \mathbf{u}_2]$.

We have $\mathbf{u}_1 = \dfrac{1}{\|\mathbf{v}_1\|}\mathbf{v}_1 = \dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ and with $\mathbf{p}_1 = (\mathbf{v}_2 \cdot \mathbf{u}_1)\mathbf{u}_1 = \dfrac{4}{\sqrt{2}}\dfrac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} = 2\begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$ we have $\mathbf{v}_2 - \mathbf{p}_1 = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$ so $\mathbf{u}_2 = \dfrac{1}{\|\mathbf{v}_2 - \mathbf{p}_1\|}\big(\mathbf{v}_2 - \mathbf{p}_1\big) = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$.

It follows that $Q = \begin{bmatrix} 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 0 \\ 0 & 1 \end{bmatrix}$ and $R = Q^TM = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 2\sqrt{2} \\ 0 & 3 \end{bmatrix}$.

Summary

Starting from any basis $\{\mathbf{v}_1,\cdots,\mathbf{v}_n\}$ for W we construct an orthonormal basis $\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$. We will construct the $\mathbf{u}_i$'s inductively so that $\{\mathbf{u}_1,\cdots,\mathbf{u}_k\}$ are orthonormal and

$$\mathrm{Span}(\mathbf{u}_1,\cdots,\mathbf{u}_k) = \mathrm{Span}(\mathbf{v}_1,\cdots,\mathbf{v}_k)$$

for $k = 1,...,n$. To begin the process, let

$$\mathbf{u}_1 = \frac{1}{\|\,\mathbf{v}_1\|}\ \mathbf{v}_1$$

Then $\mathrm{Span}(\mathbf{u}_1) = \mathrm{Span}(\mathbf{v}_1)$, since $\mathbf{u}_1$ is a multiple of $\mathbf{v}_1$ and $\|\,\mathbf{u}_1\| = 1$.
Let $\mathbf{p}_1$ be the projection of $\mathbf{v}_2$ onto $\mathrm{Span}(\mathbf{v}_1) = \mathrm{Span}(\mathbf{u}_1)$, i.e.

$$\mathbf{p}_1 = (\mathbf{v}_2\cdot\mathbf{u}_1)\,\mathbf{u}_1, \qquad \mathbf{v}_2 - \mathbf{p}_1 \in \mathrm{Span}(\mathbf{u}_1)^\perp$$

Then $\mathbf{v}_2 - \mathbf{p}_1 \neq \mathbf{0}$ since $\mathbf{v}_2 \notin \mathrm{Span}(\mathbf{u}_1)$. If we set

$$\mathbf{u}_2 = \frac{1}{\|\,\mathbf{v}_2 - \mathbf{p}_1\|}\ (\mathbf{v}_2 - \mathbf{p}_1)$$

then $\mathbf{u}_2$ is a unit vector orthogonal to $\mathrm{Span}(\mathbf{u}_1)$ and $\mathrm{Span}(\mathbf{u}_1,\mathbf{u}_2) = \mathrm{Span}(\mathbf{v}_1,\mathbf{v}_2)$.
To construct $\mathbf{u}_3$ let $\mathbf{p}_3$ be the projection of $\mathbf{v}_3$ into $\mathrm{Span}\{\mathbf{u}_1,\mathbf{u}_2\}$:

$$\mathbf{p}_2 = (\mathbf{v}_3\cdot\mathbf{u}_1)\,\mathbf{u}_1 + (\mathbf{v}_3\cdot\mathbf{u}_2)\,\mathbf{u}_2$$

and set

$$\mathbf{u}_3 = \frac{1}{\|\,\mathbf{v}_3 - \mathbf{p}_2\|}\ \big(\mathbf{v}_3 - \mathbf{p}_2\big)$$

In general we define $\mathbf{u}_k$ recursively by

$$\mathbf{u}_{k+1} = \frac{1}{\|\,\mathbf{v}_{k+1} - \mathbf{p}_k\|}\ \big(\mathbf{v}_{k+1} - \mathbf{p}_k\big)$$

where

$$\mathbf{p}_k = (\mathbf{v}_{k+1}\cdot\mathbf{u}_1)\,\mathbf{u}_1 + \cdots + (\mathbf{v}_{k+1}\cdot\mathbf{u}_k)\,\mathbf{u}_k$$

is the projection of $\mathbf{v}_{k+1}$ onto $\mathrm{Span}(\mathbf{u}_1,\cdots,\mathbf{u}_k)$. This procedure, called the **Gram-Schmidt orthogonalization process** yields an orthonormal basis $\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$ for W.

One can also use the Gram-Schmidt process to obtain the so called QR factorization of a matrix $A = QR$, where the column vectors of Q are orthonormal and R is upper triangular. In fact if M is an $m \times n$ matrix such that the n column vectors of $M = \big[\mathbf{v}_1\cdots\mathbf{v}_n\big]$ form a basis for a subspace W of $\mathbf{R}^m$ we can perform the Gram-Schmidt process on these to obtain an orthonormal basis $\{\mathbf{u}_1,\cdots,\mathbf{u}_n\}$ such that $\mathrm{Span}\big(\mathbf{u}_1,\cdots,\mathbf{u}_k\big) = \mathrm{Span}\big(\mathbf{v}_1,\cdots,\mathbf{v}_k\big)$, for $k = 1,...,n$. Hence for some constants r_{ij}

$$\mathbf{v}_k = r_{1k}\mathbf{u}_1 + \cdots + \mathbf{r}_{kk}\mathbf{u}_k + 0\mathbf{u}_{k+1} + \cdots 0\mathbf{u}_n, \qquad k = 1,...,n.$$

Let R be the upper triangular matrix with column vectors defined by

$$R = \big[\mathbf{r}_1\cdots\mathbf{r}_n\big], \qquad \text{where} \qquad \mathbf{r}_k = \begin{bmatrix} r_{1k} \\ \vdots \\ r_{kk} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

and let $Q = \big[\mathbf{u}_1\cdots\mathbf{u}_n\big]$. Then

$$Q\mathbf{r}_k = r_{1k}\mathbf{u}_1 + \cdots + r_{kk}\mathbf{u}_k = \mathbf{v}_k$$

and hence

$$QR = \big[Q\mathbf{r}_1\cdots Q\mathbf{r}_n\big] = \big[\mathbf{v}_1\cdots\mathbf{v}_n\big] = M.$$

Note that in principle one can calculate what R is from the Gram-Schmidt process.
However, it is simpler to get R just from using that $M = QR$ and that $Q^TQ = I$ so that

$$R = Q^TQR = Q^TM.$$