Slide 22-Orthogonality IV

MAT2040 Linear Algebra

SSE, CUHK(SZ)

Gram-Schmidt process

Note that the linearly independent set $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ may not the orthogonal set.

Question: Can we make a linearly independent set $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ into an orthonormal set $\{\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m\}$ while keeping the same span $(\mathbf{Span}(\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m) = \mathbf{Span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_m))$?

Idea is to use the projection and project it into the subspace and the remaining vector will be orthogonal to the subspace.

Lemma 22.1 (Projection onto a subspace) Let S be a subspace of the inner product space V and $\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ is the orthonormal basis for S, for any $\mathbf{x} \in V$. And let $\mathbf{p}$ be the projection vector of $\mathbf{x}$ onto S ($\mathbf{x} - \mathbf{p} \perp S$), then $\mathbf{p}$ is uniquely determined by

$$\mathbf{p} = \sum_{i=1}^{m} < \mathbf{x}, \mathbf{u}_i > \mathbf{u}_i$$

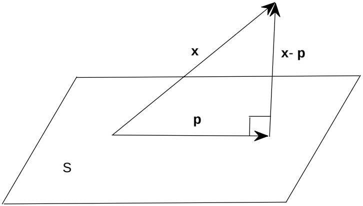


Figure: Projection of $\mathbf{x} \in V$ onto subspace S.

Proof. Since $\mathbf{p} \in S$, then write

$$\mathbf{p} = \sum_{i=1}^{m} c_i \mathbf{u}_i$$

In addition, $\mathbf{x} - \mathbf{p} \perp S$, thus

$$< \mathbf{x} - \sum_{j=1}^{m} c_j \mathbf{u}_j, \mathbf{u}_i >= 0, \ \ i = 1, \cdots, m$$

$$< \mathbf{x}, \mathbf{u}_i > - \sum_{j=1}^{m} c_j < \mathbf{u}_j, \mathbf{u}_i >= 0$$

Thus

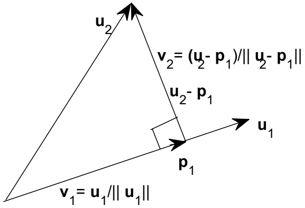
$$c_i =< \mathbf{x}, \mathbf{u}_i >, \ \ i = 1, \cdots, m$$

$$\mathbf{p} = \sum_{i=1}^{m} < \mathbf{x}, \mathbf{u}_i > \mathbf{u}_i$$

and $\mathbf{x} - \mathbf{p} \perp S$

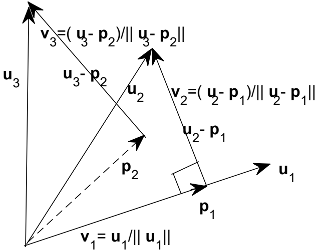
Question: Given an linearly independent set $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ in an inner product vector space V, how can we find an orthonormal set $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ such that $\mathrm{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m) = \mathrm{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_m)$?

Thinking geometrically for $m = 2$ as the following figure:



$\mathbf{v}_1 = \mathbf{u}_1/\| \mathbf{u}_1 \| \qquad \mathbf{v}_2 = (\mathbf{u}_2 - \mathbf{p}_1)/\| \mathbf{u}_2 - \mathbf{p}_1 \|$

Thinking geometrically for $m = 3$ as following figure:



$\mathbf{p}_1 = \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1 \qquad \mathbf{p}_2 = \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2$

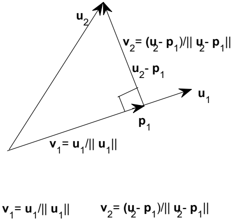
$\mathbf{v}_1 = \mathbf{u}_1 / \|\mathbf{u}_1\| \qquad \mathbf{v}_2 = (\mathbf{u}_2 - \mathbf{p}_1)/\|\mathbf{u}_2 - \mathbf{p}_1\| \qquad \mathbf{v}_3 = (\mathbf{u}_3 - \mathbf{p}_2)/\|\mathbf{u}_3 - \mathbf{p}_2\|$

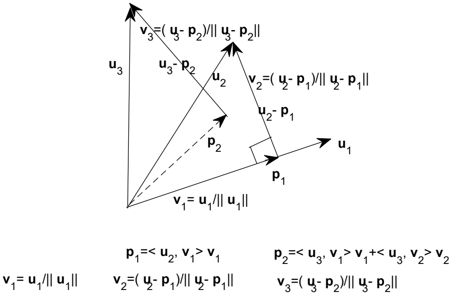
Gram-Schmidt Process

(**Gram-Schmidt Process**)

Step 1: normalize $\mathbf{u}_1$ to get $\mathbf{v}_1$, i.e., $\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|}$

Step 2: project $\mathbf{u}_2$ onto $\mathrm{Span}(\mathbf{v}_1)$ to get $\mathbf{p}_1 = \langle \mathbf{u}_2, \mathbf{v}_1 \rangle \mathbf{v}_1$, then $\mathbf{r}_1 = \mathbf{u}_2 - \mathbf{p}_1 \perp \mathrm{Span}(\mathbf{u}_1)$. Set $\mathbf{v}_2 = \frac{\mathbf{r}_1}{\|\mathbf{r}_1\|} = \frac{\mathbf{u}_2 - \mathbf{p}_1}{\|\mathbf{u}_2 - \mathbf{p}_1\|}$, then $\{\mathbf{v}_1, \mathbf{v}_2\}$ are orthonormal set and $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2)$.





Step 3: project $\mathbf{u}_3$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$ to get $\mathbf{p}_2 = \langle \mathbf{u}_3, \mathbf{v}_1 \rangle \mathbf{v}_1 + \langle \mathbf{u}_3, \mathbf{v}_2 \rangle \mathbf{v}_2$, then $\mathbf{u}_3 - \mathbf{p}_2 \perp \mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2)$, set $\mathbf{v}_3 = \frac{\mathbf{u}_3 - \mathbf{p}_2}{\|\mathbf{u}_3 - \mathbf{p}_2\|}$, then $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ are orthonormal set and $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3) = \mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$.

$$\vdots$$

Step m: project $\mathbf{u_m}$ onto $\mathrm{Span}(\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_{m-1})$ to get

$$\mathbf{p}_{m-1} = < \mathbf{u}_m, \mathbf{v}_1 > \mathbf{v}_1 + < \mathbf{u}_m, \mathbf{v}_2 > \mathbf{v}_2 + \cdots + < \mathbf{u}_m, \mathbf{v}_{m-1} > \mathbf{v}_{m-1}$$

Then $\mathbf{r}_{m-1} = \mathbf{u}_m - \mathbf{p}_{m-1} \perp \mathrm{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_{m-1})$ and set $\mathbf{v}_m = \frac{\mathbf{r}_{m-1}}{\|\mathbf{r}_{m-1}\|} = \frac{\mathbf{u}_m - \mathbf{p}_{m-1}}{\|\mathbf{u}_m - \mathbf{p}_{m-1}\|}$. Then $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ is an orthonormal set and $\mathrm{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m) = \mathrm{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_m)$

Theorem 22.2 The set $\{\mathbf{v}_1, \cdots, \mathbf{v}_m\}$ constructed by the above Gram-Schmidt process from linearly independent set $\{\mathbf{u}_1, \cdots, \mathbf{u}_m\}$ is an orthonormal set and $\mathrm{Span}(\mathbf{u}_1, \cdots, \mathbf{u}_m) = \mathrm{Span}(\mathbf{v}_1, \cdots, \mathbf{v}_m)$.

Proof Skipped. See Steven's book p267.

Example 22.3 Let

$$S = \{\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3\} = \left\{ \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 4 \\ 4 \\ -1 \end{bmatrix}, \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix} \right\} \subseteq \mathbb{R}^4$$

In $\mathbb{R}^n$, the standard inner product is the scalar product $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{y}^T \mathbf{x}$.

Now find the orthonormal basis for $\mathrm{Span}(\mathbf{u}_1, \mathbf{u}_2, \mathbf{u}_3)$.

Step 1: $\mathbf{v}_1 = \frac{\mathbf{u}_1}{\|\mathbf{u}_1\|} = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix}$.

Step 2: calculate

$$\begin{aligned}
\mathbf{u}_2' &= \mathbf{u}_2 - < \mathbf{u}_2, \mathbf{v}_1 > \mathbf{v}_1 \\
&= \begin{bmatrix} -1 \\ 4 \\ 4 \\ -1 \end{bmatrix} - \left(\begin{bmatrix} -1 & 4 & 4 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \right) \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \\
&= \begin{bmatrix} -\frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \\ -\frac{5}{2} \end{bmatrix}
\end{aligned}$$

then

$$\mathbf{v}_2 = \frac{\mathbf{u}'_2}{\| \mathbf{u}'_2 \|} = \frac{1}{5} \begin{bmatrix} -\frac{5}{2} \\ \frac{5}{2} \\ \frac{5}{2} \\ -\frac{5}{2} \end{bmatrix} = \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

Step3: calculate

$$\begin{aligned}\mathbf{u}_3' =& \mathbf{u}_3 - < \mathbf{u}_3, \mathbf{v}_1 > \mathbf{v}_1 - < \mathbf{u}_3, \mathbf{v}_2 > \mathbf{v}_2 \\ =& \begin{bmatrix} 4 \\ -2 \\ 2 \\ 0 \end{bmatrix} - \left(\begin{bmatrix} 4 & -2 & 2 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \right) \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \\ & - \left(\begin{bmatrix} 4 & -2 & 2 & 0 \end{bmatrix} \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} \right) \begin{bmatrix} -\frac{1}{2} \\ \frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix} = \begin{bmatrix} 2 \\ -2 \\ 2 \\ -2 \end{bmatrix}\end{aligned}$$

then

$$\mathbf{v}_3 = \frac{\mathbf{u}'_3}{\| \mathbf{u}'_3 \|} = \frac{1}{4} \begin{bmatrix} 2 \\ -2 \\ 2 \\ -2 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}$$

QR decomposition

Theorem 22.5 (**QR decomposition**) Let $A \in \mathbb{R}^{m\times n}$ be an $m \times n$ real matrix and rank(A)=n (column vectors are linearly independent), then A can be factorized as $A = QR$, where Q is an $m \times n$ matrix with orthonormal column vectors and R is an upper triangular $n \times n$ matrix with all positive diagonal elements.

Proof. Suppose that $A = [\mathbf{a}_1, \cdots, \mathbf{a}_n]$ and $\{\mathbf{q}_1, \cdots, \mathbf{q}_n\}$ is the orthnormal set obtained from $\{\mathbf{a}_1, \cdots, \mathbf{a}_n\}$ by the following Gram-Schmidt process (see step 1 and step 2).

Gram-Schmidt process

Step 1: $\mathbf{q}_1 = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|}, \quad \mathbf{q}_1^T\mathbf{a}_1 =< \mathbf{a}_1, \mathbf{q}_1 >=\| \mathbf{a}_1 \|> 0$

QR decomposition

Step 2: For $j = 2, \cdots, n$

1. Let

$$\mathbf{r}_{j-1} = \mathbf{a}_j - \mathbf{p}_{j-1}$$

where $\mathbf{p}_{j-1} = < \mathbf{a}_j, \mathbf{q}_1 > \mathbf{q}_1 + \cdots + < \mathbf{a}_j, \mathbf{q}_{j-1} > \mathbf{q}_{j-1}$ is the projection of $\mathbf{a}_j$ onto $\mathrm{Span}\{\mathbf{q}_1, \cdots, \mathbf{q}_{j-1}\} = \mathrm{Span}\{\mathbf{a}_1, \cdots, \mathbf{a}_{j-1}\}$

2. Let $\mathbf{q}_j = \frac{\mathbf{r}_{j-1}}{\|\mathbf{r}_{j-1}\|}$, then

$\mathbf{a}_j = < \mathbf{a}_j, \mathbf{q}_1 > \mathbf{q}_1 + \cdots + < \mathbf{a}_j, \mathbf{q}_{j-1} > \mathbf{q}_{j-1} + \| \mathbf{r}_{j-1} \| \mathbf{q}_j$.

QR decomposition

The above relations can be rewritten as

$$\begin{aligned}
\mathbf{a}_1 &= \| \mathbf{a}_1 \| \mathbf{q}_1 \\
\mathbf{a}_2 &= < \mathbf{a}_2, \mathbf{q}_1 > \mathbf{q}_1 + \|\mathbf{r}_1\|\mathbf{q}_2 \\
\mathbf{a}_3 &= < \mathbf{a}_3, \mathbf{q}_1 > \mathbf{q}_1 + < \mathbf{a}_3, \mathbf{q}_2 > \mathbf{q}_2 + \|\mathbf{r}_2\|\mathbf{q}_3 \\
&\vdots \\
\mathbf{a}_n &= < \mathbf{a}_n, \mathbf{q}_1 > \mathbf{q}_1 + \cdots + < \mathbf{a}_n, \mathbf{q}_{n-1} > \mathbf{q}_{n-1} + \|\mathbf{r}_{n-1}\|\mathbf{q}_n
\end{aligned}$$

Thus,

$$\mathbf{q}_j^T \mathbf{a}_j = < \mathbf{a}_j, \mathbf{q}_j > = \| \mathbf{r}_{j-1} \| > 0 \quad (j = 2, \cdots, n).$$

and

$$\mathbf{q}_j \perp \mathrm{Span}(\mathbf{q}_1, \cdots, \mathbf{q}_{j-1}) = \mathrm{Span}(\mathbf{a}_1, \cdots, \mathbf{a}_{j-1}), (j = 2, \cdots, n)$$

QR decomposition

$$A = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n] = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n] \begin{bmatrix} \|\mathbf{a}_1\| & < \mathbf{q}_1, \mathbf{a}_2 > & \dots & < \mathbf{q}_1, \mathbf{a}_n > \\ 0 & \|\mathbf{r}_1\| & \dots & < \mathbf{q}_2, \mathbf{a}_n > \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \|\mathbf{r}_{n-1}\| \end{bmatrix}$$

$$\triangleq QR$$

Here $Q = [\mathbf{q}_1, \mathbf{q}_2, \cdots, \mathbf{q}_n]$, and

$$R = \begin{bmatrix} \mathbf{q}_1^T \mathbf{a}_1 & \mathbf{q}_1^T \mathbf{a}_2 & \dots & \mathbf{q}_1^T \mathbf{a}_n \\ 0 & \mathbf{q}_2^T \mathbf{a}_2 & \dots & \mathbf{q}_2^T \mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{q}_n^T \mathbf{a}_n \end{bmatrix}$$

since $\mathbf{q}_1^T \mathbf{a}_1 =< \mathbf{a}_1, \mathbf{q}_1 >= \|\mathbf{a}_1\|, \quad \mathbf{q}_j^T \mathbf{a}_j =< \mathbf{a}_j, \mathbf{q}_j >=\| \mathbf{r}_{j-1} \|> 0 \quad (j = 2, \cdots, n)$.

QR decomposition

In fact, from $A = QR$, one has

$$R = Q^T A = \begin{bmatrix} \mathbf{q}_1^T \mathbf{a}_1 & \mathbf{q}_1^T \mathbf{a}_2 & \dots & \mathbf{q}_1^T \mathbf{a}_n \\ \mathbf{q}_2^T \mathbf{a}_1 & \mathbf{q}_2^T \mathbf{a}_2 & \dots & \mathbf{q}_2^T \mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{q}_n^T \mathbf{a}_1 & \mathbf{q}_n^T \mathbf{a}_2 & \dots & \mathbf{q}_n^T \mathbf{a}_n \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{q}_1^T \mathbf{a}_1 & \mathbf{q}_1^T \mathbf{a}_2 & \dots & \mathbf{q}_1^T \mathbf{a}_n \\ 0 & \mathbf{q}_2^T \mathbf{a}_2 & \dots & \mathbf{q}_2^T \mathbf{a}_n \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{q}_n^T \mathbf{a}_n \end{bmatrix}$$

where

$$\mathbf{q}_j \perp \mathrm{Span}(\mathbf{q}_1, \cdots, \mathbf{q}_{j-1}) = \mathrm{Span}(\mathbf{a}_1, \cdots, \mathbf{a}_{j-1}), (j = 2, \cdots, n)$$

are used.

Theorem: If A is an $m \times n$ matrix of rank n, then the least square solution of $A\mathbf{x} = \mathbf{b}$ is given by $\hat{\mathbf{x}} = R^{-1}Q^T\mathbf{b}$, where Q and R are the matrices obtained from the QR decomposition of A.

Proof. Since $A = QR$, then $A^TA\hat{x} = A^T\mathbf{b}$. $A^TA = R^TQ^TQR = R^TR$, $A^T\mathbf{b} = R^TQ^T\mathbf{b}$.
$A\mathbf{x} = \mathbf{b}$ is equivalent to

$$R^TR\hat{x} = R^TQ^T\mathbf{b}$$

Since R is invertible, thus

$$R\hat{x} = Q^T\mathbf{b}$$

Then

$$\hat{x} = R^{-1}Q^T\mathbf{b}$$

Appendix: Gram-Schmidt Process on a functional space (an inner product space)

Example 22.7 For subspace **Span**$\{1, x, x^2\} \subseteq C[-1, 1]$, find the orthonormal basis for **Span**$\{1, x, x^2\}$, where the inner product and norm is defined as:

$$< f, g >= \int_{-1}^{1} f(x)g(x)dx, \quad \| f \|^2 = \int_{-1}^{1} |f(x)|^2 dx$$

Now it can be verify that

$$< 1, x >= 0, \quad < x, x^2 >= 0, \quad < 1, x^2 >= \frac{2}{3}$$

$$\mathbf{p}_1 = x - \frac{< x, 1 >}{< 1, 1 >} = x$$

$$\mathbf{p}_2 = x^2 - \frac{< x^2, 1 >}{< 1, 1 >} - \frac{< x^2, x >}{< x, x >} x = x^2 - \frac{1}{3}$$

$$\mathbf{q}_1 = \frac{1}{\| 1 \|} = \frac{1}{\sqrt{2}}$$

$$\mathbf{q}_2 = \frac{\mathbf{p}_1}{\| \mathbf{p}_1} \| = \frac{x}{\sqrt{\frac{2}{3}}}$$

$$\mathbf{q}_3 = \frac{\mathbf{p}_2}{\| \mathbf{p}_2 \|} = \frac{x^2 - \frac{1}{3}}{\sqrt{\frac{8}{45}}}$$

$\{\frac{1}{\sqrt{2}}, \quad \frac{x}{\sqrt{\frac{2}{3}}}, \quad \frac{x^2-\frac{1}{3}}{\sqrt{\frac{8}{45}}}\}$ are the orthonormal basis for **Span**$\{1, x, x^2\}$.