## Gram Schmidt Process

This *suggests* that if we begin with an arbitrary basis $\{\mathbf{x}_1, \ \mathbf{x}_2, \ \ldots, \ \mathbf{x}_k\}$ for $W$ then we might be able to "orthogonalize" one vector at a time. Let us start with a simple example.

### Example

Find an orthogonal basis for $W = \text{span}\,(\mathbf{x}_1, \ \mathbf{x}_2)$ where

$$\mathbf{x}_1 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}.$$

**Solution:** Let $\mathbf{v}_1 = \mathbf{x}_1$. We need to find a vector which is orthogonal to $\mathbf{v}_1$ but also lies in $W$. The perpendicular component of $\mathbf{x}_2$ with respect to $\mathbf{x}_1$ will be orthogonal to $\mathbf{x}_1$ and lie in $W$ (why?). Thus we set

$$\mathbf{v}_2 = \text{perp}_{\mathbf{x}_1}(\mathbf{x}_2) = \mathbf{x}_2 - \text{proj}_{\mathbf{x}_1}(\mathbf{x}_2).$$

Now $\{\mathbf{v}_1, \ \mathbf{v}_2\}$ will span $W$ and thus form an orthogonal basis for $W$. This projection is onto a *1-dimensional subspace* and so

## Gram Schmidt Process

$$\begin{aligned} \mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{x}_1}{\mathbf{x}_1 \cdot \mathbf{x}_1}\mathbf{x}_1 \\ &= \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} - \frac{-2}{5}\begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{5} \\ 1 \\ \frac{2}{5} \end{bmatrix} \end{aligned}$$

As a check

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{5} \\ 1 \\ \frac{2}{5} \end{bmatrix} = 0.$$

Note that the output from this method depends on the *order* of the original basis. If we had taken

# Gram Schmidt Process

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}$$

then we would have obtained a different orthogonal basis (check!).

Now let us suppose $W$ is a three dimensional subspace with $W = \text{span}\left(\mathbf{x}_1,\ \mathbf{x}_2,\ \mathbf{x}_3\right)$. We proceed as above by setting

$$\begin{aligned} \mathbf{v}_1 &= \mathbf{x}_1 \\ \mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 . \end{aligned}$$

We now have $\text{span}\left(\mathbf{v}_1,\ \mathbf{v}_2\right) = \text{span}\left(\mathbf{x}_1,\ \mathbf{x}_2\right) = W_2$, say, and $\mathbf{v}_2$ orthogonal to $\mathbf{v}_1$. We need a third vector $\mathbf{v}_3$ which lies in $W$ which is orthogonal to *both* $\mathbf{v}_1$ and $\mathbf{v}_2$ (that is, to $W_2$). Thus

$$\mathbf{v}_3 = \text{perp}_{W_2}\left(\mathbf{x}_3\right) = \mathbf{x}_3 - \text{proj}_{W_2}\left(\mathbf{x}_3\right).$$

# Gram Schmidt Process

But $\{\mathbf{v}_1,\ \mathbf{v}_2\}$ form an orthogonal basis for $W_2$ and so

$$\text{proj}_{W_2}\left(\mathbf{x}_3\right) = \text{proj}_{\mathbf{v}_1}\left(\mathbf{x}_3\right) + \text{proj}_{\mathbf{v}_2}\left(\mathbf{x}_3\right) = \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 + \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2$$

(since we have not required $\{\mathbf{v}_1,\ \mathbf{v}_2\}$ to form an orthonormal set). Thus

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2$$

will be orthogonal to $\mathbf{v}_1$ and $\mathbf{v}_2$. Explicitly, we have

$$\begin{aligned} \mathbf{v}_3 \cdot \mathbf{v}_1 &= \mathbf{x}_3 \cdot \mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 \cdot \mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 \cdot \mathbf{v}_1 \\ &= \mathbf{x}_3 \cdot \mathbf{v}_1 - \mathbf{x}_3 \cdot \mathbf{v}_1 = 0 \\ \mathbf{v}_3 \cdot \mathbf{v}_2 &= \mathbf{x}_3 \cdot \mathbf{v}_2 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 \cdot \mathbf{v}_2 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2 \cdot \mathbf{v}_2 \\ &= \mathbf{x}_3 \cdot \mathbf{v}_2 - \mathbf{x}_3 \cdot \mathbf{v}_2 = 0 \end{aligned}$$

## Gram Schmidt Process

This construction generalises to give

**Theorem 11.12 (Gram-Schmidt Process)**

*Let* $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$ *for a basis for a subspace* $W$ *of* $\mathbf{R}^n$. *Define*

$$\begin{aligned}
\mathbf{v}_1 &= \mathbf{x}_1 \\
\mathbf{v}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 \\
\mathbf{v}_3 &= \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 \\
&\vdots \\
\mathbf{v}_k &= \mathbf{x}_k - \frac{\mathbf{x}_k \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_k \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\mathbf{v}_2 - \cdots - \frac{\mathbf{x}_k \cdot \mathbf{v}_{k-1}}{\mathbf{v}_{k-1} \cdot \mathbf{v}_{k-1}}\mathbf{v}_{k-1}
\end{aligned}$$

*and*

$$W_i = \text{span}\left(\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\right).$$

## Gram Schmidt Process

Then, for each $i = 1, 2, \ldots, k$, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$ is an orthogonal basis for $W_i$. In particular $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is an orthogonal basis for $W_k = W$.

Proof: Induction based upon the argument above.

Note that it is the previously computed $\mathbf{v}_i$ that are on the right hand side of these expressions (and **not** the $\mathbf{x}_i$).

# Gram Schmidt Process

## Example

Find an orthogonal basis for $W = \text{span}(\mathbf{x}_1, \ \mathbf{x}_2, \ \mathbf{x}_3)$ where

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

**Solution:** Applying the Gram-Schmidt Process, we have

$$\mathbf{v}_1 = \mathbf{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \frac{3}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix}$$

# Gram Schmidt Process

and

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1} \mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2} \mathbf{v}_2$$

$$= \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{2}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{\frac{2}{4}}{\frac{12}{16}} \frac{1}{4} \begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{6} \begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix}$$

# Gram Schmidt Process

The Gram-Schmidt process is frequently messy to complete. For hand calculations, one can scale to vectors $\mathbf{v}_i$ to avoid fractions. Thus, in the above example, we could have chosen

$$\mathbf{v}_2' = \begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix}$$

and then

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2' = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} - \frac{2}{4}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{2}{12}\begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix}.$$

We would then scale $\mathbf{v}_3$ to give

$$\mathbf{v}_3' = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix}.$$

# Gram Schmidt Process

The set $\{\mathbf{v}_1, \ \mathbf{v}_2', \ \mathbf{v}_3'\}$ also forms an orthogonal basis for $W$. To obtain an orthonormal basis, we normalize each vector in any orthogonal basis. Thus

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

$$\mathbf{q}_2 = \frac{1}{\|\mathbf{v}_2'\|}\mathbf{v}_2' = \frac{\sqrt{3}}{6}\begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix}$$

$$\mathbf{q}_3 = \frac{1}{\|\mathbf{v}_3'\|}\mathbf{v}_3' = \frac{\sqrt{6}}{6}\begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix}$$

is an orthonormal basis for $W$.

## Gram Schmidt Process

When the Gram-Schmidt Process is implemented on a computer, there is almost always round off error. This round off error leads to a loss of orthogonality of the vectors $\mathbf{q}_i$. To minimize this, the vectors $\mathbf{v}_i$ are normalize to give $\mathbf{q}_i$ as soon as they are computed rather than at the end of the computation. The remaining vectors $\mathbf{x}_j$ are modified to be orthogonal to $\mathbf{q}_i$. This procedure is known as the **Modified Gram-Schmidt Process**. For hand computations, where we are using exact arithmetic, the issue of round off error does not arise. In this case we wish to avoid dealing with fractions and square roots as long as possible and so we delay scaling until the end.

One important use of the Gram-Schmidt Process is to contruct a basis containing a given vector (or set of vectors).

## Gram Schmidt Process

### Example

Find an orthogonal basis for $\mathbf{R}^3$ that contains

$$\mathbf{v}_1 = \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}.$$

**Solution:** We first need to find any basis for $\mathbf{R}^3$ that contains $\mathbf{v}_1$. Let

$$\mathbf{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

then $\{\mathbf{v}_1,\ \mathbf{x}_2,\ \mathbf{x}_3\}$ will form a basis for $\mathbf{R}^3$. We now apply the Gram-Schmidt Process to this basis. Thus

# Gram Schmidt Process

$$\mathbf{v}_2 = \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} - \frac{2}{14}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} = \frac{1}{7}\begin{bmatrix} 5 \\ 1 \\ -3 \end{bmatrix}$$

and so we choose

$$\mathbf{v}_2' = \begin{bmatrix} 5 \\ 1 \\ -3 \end{bmatrix}.$$

Similarly, we have

$$\mathbf{v}_3 = \mathbf{x}_3 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\mathbf{v}_1 - \frac{\mathbf{x}_3 \cdot \mathbf{v}_2'}{\mathbf{v}_2' \cdot \mathbf{v}_2'}\mathbf{v}_2' = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} - \frac{-1}{14}\begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix} - \frac{1}{35}\begin{bmatrix} 5 \\ 1 \\ -3 \end{bmatrix} = \frac{3}{10}\begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}$$

# Gram Schmidt Process

and so we choose

$$\mathbf{v}_3' = \begin{bmatrix} 0 \\ 3 \\ 1 \end{bmatrix}.$$

Thus $\{\mathbf{v}_1, \ \mathbf{v}_2', \ \mathbf{v}_3'\}$ is an orthogonal basis for $\mathbf{R}^n$ containing $\mathbf{v}_1$.

Let $A$ be a $m \times n$ matrix with linearly independent columns (thus $m \geqslant n$ and $\text{rk}\,(A) = n$). If we apply the Gram-Schmidt process to $A$ we obtain a very useful factorization of $A$ into a product of a matrix $Q$ whose columns form an orthonormal set and an upper triangular matrix $R$. This factorization is called **QR factorization** and it is extremely useful in many numerical problems from the computation of eigenvalues to least squares approximation.

# QR FACTORIZATION

Let

$$A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$$

and $\{\mathbf{q}_1, \ \mathbf{q}_2, \ \ldots, \ \mathbf{q}_n\}$ be the set of orthonormal vectors obtained from $\{\mathbf{a}_1, \ \mathbf{a}_2, \ \ldots, \ \mathbf{a}_n\}$ by the Gram-Schmidt Process with normalization. From the Gram-Schmidt Theorem $W_1 = \text{span}\,(\mathbf{a}_1) = \text{span}\,(\mathbf{q}_1)$ and so

$$\mathbf{a}_1 = r_{11}\,\mathbf{q}_1$$

for some scalar $r_{11}$. Similarly $W_2 = \text{span}\,(\mathbf{a}_1, \ \mathbf{a}_2) = \text{span}\,(\mathbf{q}_1, \mathbf{q}_2)$ and so

$$\mathbf{a}_2 = r_{12}\,\mathbf{q}_1 + r_{22}\,\mathbf{q}_2$$

for some scalars $r_{12}$ and $r_{22}$. Continuing, we obtain

$$\mathbf{a}_i = r_{1i}\,\mathbf{q}_1 + r_{2i}\,\mathbf{q}_2 + \cdots + r_{ii}\,\mathbf{q}_i$$

for $i = 1, \ 2, \ \ldots, \ n$.

# QR FACTORIZATION

Thus

$$\begin{aligned} A &= \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ 0 & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_{nn} \end{bmatrix} = Q\,R. \end{aligned}$$

$Q$ has orthonormal columns (if $A$ is square then $Q$ is an orthogonal matrix) and $\text{col}\,(A) = \text{col}\,(Q)$. If $r_{ii} = 0$ then $\mathbf{a}_i \in W_{i-1}$ and so $\{\mathbf{a}_1, \ \mathbf{a}_2, \ \ldots, \ \mathbf{a}_i\}$ would not be linearly independent. Therefore $r_{ii} \neq 0$ and so $R$ is upper triangular with non-zero diagonal elements. Thus $R$ is invertible. The entries in $R$ are computed during the Gram-Schmidt Process. At stage $i$ of the Gram-Schmidt Process, we have

$$\mathbf{v}_i = \mathbf{a}_i - \frac{\mathbf{a}_i \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1 - \frac{\mathbf{a}_i \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\,\mathbf{v}_2 - \cdots - \frac{\mathbf{a}_i \cdot \mathbf{v}_{i-1}}{\mathbf{v}_{i-1} \cdot \mathbf{v}_{i-1}}\,\mathbf{v}_{i-1}$$

# QR Factorization

and so

$$\mathbf{a}_i = \mathbf{v}_i + \frac{\mathbf{a}_i \cdot \mathbf{v}_1}{\mathbf{v}_1 \cdot \mathbf{v}_1}\,\mathbf{v}_1 + \frac{\mathbf{a}_i \cdot \mathbf{v}_2}{\mathbf{v}_2 \cdot \mathbf{v}_2}\,\mathbf{v}_2 + \cdots + \frac{\mathbf{a}_i \cdot \mathbf{v}_{i-1}}{\mathbf{v}_{i-1} \cdot \mathbf{v}_{i-1}}\,\mathbf{v}_{i-1}.$$

Now

$$\mathbf{q}_j = \frac{1}{\|\mathbf{v}_j\|}\mathbf{v}_j$$

and so

$$\frac{\mathbf{a}_i \cdot \mathbf{v}_j}{\mathbf{v}_j \cdot \mathbf{v}_j}\,\mathbf{v}_j = \frac{\|\mathbf{v}_j\|\,(\mathbf{a}_i \cdot \mathbf{q}_j)}{\mathbf{v}_j \cdot \mathbf{v}_j}\,\|\mathbf{v}_j\|\,\mathbf{q}_j = (\mathbf{a}_i \cdot \mathbf{q}_j)\,\mathbf{q}_j.$$

Therefore

$$\mathbf{a}_i = \|\mathbf{v}_i\|\,\mathbf{q}_i + (\mathbf{a}_i \cdot \mathbf{q}_1)\,\mathbf{q}_1 + (\mathbf{a}_i \cdot \mathbf{q}_2)\,\mathbf{q}_2 + \cdots + (\mathbf{a}_i \cdot \mathbf{q}_{i-1})\,\mathbf{q}_{i-1}$$

and thus

$$r_{ji} = \begin{cases} \mathbf{a}_i \cdot \mathbf{q}_j & \text{for } j = 1,\ 2,\ \ldots,\ i-1 \\ \|\mathbf{v}_i\| & \text{for } j = i. \end{cases}$$

# QR Factorization

Since $\mathbf{q}_j$ for $j = 1,\ 2,\ \ldots,\ i-1$ have been computed in previous steps, we can compute all these coefficients at step $i$ of the Gram-Schmidt Process. This method also produces an upper triangular matrix $R$ which has strictly positive entries on the diagonal (not just non-zero).

Note that, since $Q$ has orthonormal columns, $Q^T Q = I$ and so

$$R = Q^T Q R = Q^T A.$$

Moreover, since $\text{col}(A) = \text{col}(Q)$,

$$\text{proj}_{\text{col}(A)}(\mathbf{x}) = \text{proj}_{\text{col}(Q)}(\mathbf{x}) = Q\,Q^T\,\mathbf{x}.$$

# QR Factorization

## Example

Find a QR factorization for

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}.$$

**Solution:** The columns of $A$ are the vectors we used in the first example of the Gram-Schmidt process. In that example, we found

$$\mathbf{q}_1 = \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{q}_2 = \frac{\sqrt{3}}{6}\begin{bmatrix} 1 \\ 1 \\ 1 \\ -3 \end{bmatrix} \quad \text{and} \quad \mathbf{q}_3 = \frac{\sqrt{6}}{6}\begin{bmatrix} 1 \\ 1 \\ -2 \\ 0 \end{bmatrix}.$$

# QR Factorization

Thus

$$Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \mathbf{q}_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{2} & \frac{\sqrt{3}}{6} & \frac{\sqrt{6}}{6} \\ \frac{1}{2} & \frac{\sqrt{3}}{6} & \frac{\sqrt{6}}{6} \\ \frac{1}{2} & \frac{\sqrt{3}}{6} & -\frac{\sqrt{6}}{3} \\ \frac{1}{2} & -\frac{\sqrt{3}}{2} & 0 \end{bmatrix}.$$

R could be found directly from the Gram-Schmidt Process or

$$R = Q^T A = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{6} & \frac{\sqrt{3}}{6} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{6}}{6} & \frac{\sqrt{6}}{6} & -\frac{\sqrt{6}}{3} & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 2 & \frac{3}{2} & 1 \\ 0 & \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{3} \\ 0 & 0 & \frac{\sqrt{6}}{3} \end{bmatrix}.$$

# QR Factorization

The QR decomposition can be generalized for any matrix (not just for matrices whose columns are linearly independent). However, if the columns of $A$ are not linearly independent, then some of the diagonal elements of $R$ are $0$. MATLAB has its `qr` command to compute QR decompositions. Thus

```
A=[1 1 1;1 1 1; 1 1 0; 1 0 0];
>> [Q, R]=qr(A,0)
Q =
   -0.5000   -0.2887    0.4082
   -0.5000   -0.2887    0.4082
   -0.5000   -0.2887   -0.8165
   -0.5000    0.8660         0
R =
   -2.0000   -1.5000   -1.0000
         0   -0.8660   -0.5774
         0         0    0.8165
```

MATLAB has chosen its normalizations differently from what we did above. This explains the differences in signs. Note that

# QR Factorization

```
>> [Q, R]=qr(A)
Q =
   -0.5000   -0.2887    0.4082   -0.7071
   -0.5000   -0.2887    0.4082    0.7071
   -0.5000   -0.2887   -0.8165   -0.0000
   -0.5000    0.8660         0    0.0000
R =
   -2.0000   -1.5000   -1.0000
         0   -0.8660   -0.5774
         0         0    0.8165
         0         0         0
```

In this case, MATLAB has computed a fourth vector

$$\mathbf{q}_4 = \frac{\sqrt{2}}{2}\begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

# QR Factorization

so that $Q$ is orthogonal (that is $\{\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4\}$ will form an orthonormal basis for $\mathbf{R}^4$). This necessitates adding a row of zeros to $R$, since the columns of $A$ do not depend on $\mathbf{q}_4$, to achieve $A = Q\,R$.

```
>> Q*R
ans =
    1.0000    1.0000    1.0000
    1.0000    1.0000    1.0000
    1.0000    1.0000         0
    1.0000    0.0000    0.0000
```

# QR Factorization

If we have the QR decomposition of $A$ then we can quickly compute the solution of $A\,\mathbf{x} = \mathbf{b}$ since $A\,\mathbf{x} = Q\,R\,\mathbf{x}$ and so

$$R\,\mathbf{x} = Q^T\,Q\,R\,\mathbf{x} = Q^T\,A\,\mathbf{x} = Q^T\,\mathbf{b}.$$

Since $R$ is upper triangular, the system $R\,\mathbf{x} = Q^T\,\mathbf{b}$ can be solved by back substitution. This approach avoids row reduction. Round off error is a major concern in row reduction since we have ensure that off diagonal elements are zero. Any round off error can dramatic decrease the accuracy of the solution. The QR approach is more stable numerically. Of course, with exact arithmetic, this problem does not occur.

# OVERDETERMINED SYSTEMS

MATLAB will give a solution to an overdetermined system of equations, $A\,\mathbf{x} = \mathbf{b}$; that is, where $A$ has more rows than columns (more equations that unknowns). For such a system, $\mathbf{b}$ is unlikely to be in $\operatorname{col}(A)$ and therefore the system is unlikely to have a solution. However if we project $\mathbf{b}$ onto $\operatorname{col}(A)$ then we get the system

$$A\,\mathbf{x} = \operatorname{proj}_{\operatorname{col}(A)}(\mathbf{b})$$

which will have a solution (note that if $\mathbf{b} \in \operatorname{col}(A)$ then this system is identical to the original system). If we find the QR decomposition of $A$ then this equation takes the form

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

or

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

This is the so-called *least squares solution* for an overdetermined system and has the same form as a system which is consistent.

# OVERDETERMINED SYSTEMS

The residual (or "error") in the least squares solution is

$$\mathbf{r} = \mathbf{b} - A\mathbf{x} = \mathbf{b} - \operatorname{proj}_{\operatorname{col}(A)}(\mathbf{b}) = \operatorname{perp}_{\operatorname{col}(A)}(\mathbf{b}).$$

The choice of the least squares solution $\mathbf{x}$ minimizes the length of this vector. If we choose any vector $\mathbf{y}$ to "be" the solution then the length of the residual would be

$$\|\mathbf{b}-A\,\mathbf{y}\|^2 = \|\mathbf{b}-A\,\mathbf{x}+A\,\mathbf{x}-A\,\mathbf{y}\|^2 = \|\mathbf{r}\|^2+\|A(\mathbf{x}-\mathbf{y})\|^2+2\,\mathbf{r}\cdot A(\mathbf{x}-\mathbf{y}).$$

Now $A(\mathbf{x}-\mathbf{y}) \in \operatorname{col}(A)$ and $\mathbf{r}$ is orthogonal to anything in $\operatorname{col}(A)$ and so

$$\|\mathbf{b}-A\,\mathbf{y}\|^2 = \|\mathbf{r}\|^2 + \|A(\mathbf{x}-\mathbf{y})\|^2 \geqslant \|\mathbf{r}\|^2.$$

Therefore the error in choosing $\mathbf{y}$ is at least as bad as that for the least squares solution. Furthermore, $\mathbf{y}$ minimizes the residual if and only if $\mathbf{x}-\mathbf{y} \in \operatorname{null}(A)$. Thus $\operatorname{null}(A)$ gives the ambiguity in the solution that minimizes the residual. If $\operatorname{nullity}(A) = \dim(\operatorname{null}(A)) = 0$ then the least squares solution is the *unique* solution that minimizes the residual.

## Overdetermined Systems

**Example**

Find the (least squares) solution for $A\,\mathbf{x} = \mathbf{b}$ where

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \\ 2 & 1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ -1 \end{bmatrix}.$$

**Solution:** Apply Gram Schmidt we have

$$\mathbf{q}_1 = \frac{1}{3}\begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix}, \quad r_{11} = 3$$

and

$$\mathbf{v}_2 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - \left( \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} \cdot \mathbf{q}_1 \right) \mathbf{q}_1 = \begin{bmatrix} 1 \\ 3 \\ 1 \end{bmatrix} - 3\,\mathbf{q}_1 = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Therefore



$$\mathbf{q}_2 = \frac{\sqrt{2}}{2}\begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}, \quad r_{22} = \sqrt{2}, \quad r_{12} = 3$$

and so $A = Q\,R$ where

$$Q = \begin{bmatrix} \frac{1}{3} & 0 \\ \frac{2}{3} & \frac{\sqrt{2}}{2} \\ \frac{2}{3} & -\frac{\sqrt{2}}{2} \end{bmatrix}, \quad R = \begin{bmatrix} 3 & 3 \\ 0 & \sqrt{2} \end{bmatrix}.$$

The least square solution will be

$$\left[R \mid Q^T\,\mathbf{b}\right] = \left[\begin{array}{cc|c} 3 & 3 & -\frac{1}{3} \\ 0 & \sqrt{2} & 0 \end{array}\right];$$

that is

$$\mathbf{x} = \begin{bmatrix} -\frac{1}{9} \\ 0 \end{bmatrix}.$$

# Orthogonal Complements

From our discussion on projections, we have seen that, for a symmetric matrix $A$, col$(A)$ is orthogonal to null$(A)$. What happens more generally? Can we get an insight into the geometry of the subspaces associated with a matrix $A$?

**Definition 11.13 (Orthogonal Complement)**

*Let $W$ be a subspace of $\mathbf{R}^n$. A vector $\mathbf{v} \in \mathbf{R}^n$ is **orthogonal to** $W$ if $\mathbf{v}$ is orthogonal to **every** vector in $W$. The set of all vectors orthogonal to $W$ is called the orthogonal complement of $W$, denoted by $W^{\perp}$. Thus*

$$W^{\perp} = \{\mathbf{v} \in \mathbf{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \quad \text{for all } \mathbf{w} \in W\}.$$

For a symmetric matrix $A$, col$(A)$ is the orthogonal complement of null$(A)$.

# Orthogonal Complements

**Theorem 11.14**

*Let $W$ be a subspace of $\mathbf{R}^n$.*

(a) $W^{\perp}$ *is a subspace of* $\mathbf{R}^n$.

(b) $(W^{\perp})^{\perp} = W$.

(c) $W \cap W^{\perp} = \{\mathbf{0}\}$.

(d) *If* $W = \text{span}(\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_k)$ *then* $\mathbf{v} \in W^{\perp}$ *if and only if* $\mathbf{v} \cdot \mathbf{w}_i = 0$ *for each* $i = 1, 2, \ldots, k$.

Proof: Exercise

Note that

$$\text{perp}_W(\mathbf{v}) \in W^{\perp}$$

since the perpendicular component is orthogonal to any vector in $W$.

# Orthogonal Complements

**Example**

Find the orthogonal complement of the plane

$$W = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x + y - z = 0 \right\}.$$

**Solution:** We have already seen that $\mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$ and $\mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$ forms a basis for $W$. Thus, for $\mathbf{v} = \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in W^{\perp}$, we require $\mathbf{v} \cdot \mathbf{w}_1 = x + z = 0$ and $\mathbf{v} \cdot \mathbf{w}_2 = y + z = 0$. Thus

# Orthogonal Complements

$$W^{\perp} = \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} : x + z = 0,\ y + z = 0 \right\} = \left\{ \begin{bmatrix} t \\ t \\ -t \end{bmatrix} : t \in \mathbf{R} \right\}$$

is the line whose direction is given by $\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}$.

**Theorem 11.15 (Orthogonal Decomposition)**

*Let $W$ be a subspace of $\mathbf{R}^n$ and $\mathbf{v} \in \mathbf{R}^n$. Then there exists unique vectors $\mathbf{w} \in W$ and $\mathbf{w}^{\perp} \in W^{\perp}$ such that*

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^{\perp}.$$

*Furthermore*

$$\mathbf{w} = \mathrm{proj}_W(\mathbf{v}) \quad \textit{and} \quad \mathbf{w}^{\perp} = \mathrm{perp}_W(\mathbf{v}).$$

Proof: Clearly

$$\mathbf{w} = \text{proj}_W(\mathbf{v}) \in W, \qquad \mathbf{w}^\perp = \text{perp}_W(\mathbf{v}) \in W^\perp$$

and

$$\mathbf{v} = \mathbf{w} + \mathbf{w}^\perp.$$

It only remains to show that this choice is unique. Suppose $\mathbf{v} = \mathbf{w}_1 + \mathbf{w}_1^\perp$ with $\mathbf{w}_1 \in W$ and $\mathbf{w}_1^\perp \in W^\perp$. Then $\mathbf{w} + \mathbf{w}^\perp = \mathbf{w}_1 + \mathbf{w}_1^\perp$ and so

$$\mathbf{w} - \mathbf{w}_1 = \mathbf{w}^\perp - \mathbf{w}_1^\perp.$$

However $\mathbf{w} - \mathbf{w}_1 \in W$ and $\mathbf{w}^\perp - \mathbf{w}_1^\perp \in W^\perp$. By part (c) of the above theorem, the only vector in common to both $W$ and $W^\perp$ is $\mathbf{0}$ and so

$$\mathbf{w} - \mathbf{w}_1 = \mathbf{w}^\perp - \mathbf{w}_1^\perp = \mathbf{0}.$$

Therefore $\mathbf{w} = \mathbf{w}_1$ and $\mathbf{w}^\perp = \mathbf{w}_1^\perp$ and so we have uniqueness. □



An immediate consequence of this result is the following.

**THEOREM 11.16**

*Let $W$ be a subspace of $\mathbf{R}^n$. Then $W \oplus W^\perp = \mathbf{R}^n$ and*

$$\dim(W) + \dim\left(W^\perp\right) = n.$$

*Furthermore*

$$\text{perp}_W(\mathbf{v}) = \text{proj}_{W^\perp}(\mathbf{v})$$

*for all $\mathbf{v} \in \mathbf{R}^n$.*

Returning to the issue of the subspaces associated with a matrix $A$, we have

**THEOREM 11.17**

*Let $A$ be a $m \times n$ matrix. Then*

$$(\text{col}(A))^\perp = \text{null}\left(A^T\right)$$
$$(\text{row}(A))^\perp = \text{null}(A).$$

## Orthogonal Complements

Proof: If $\mathbf{x} \in (\text{col}(A))^{\perp}$ then $\mathbf{x}$ is orthogonal to every column of $A = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$. This occurs if and only if $\mathbf{x}^T \mathbf{a}_i = 0$ , that is

$$\mathbf{x}^T \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix} = \mathbf{x}^T A = \mathbf{0}^T.$$

Therefore

$$A^T \mathbf{x} = \mathbf{0}$$

and so $\mathbf{x} \in \text{null}\left(A^T\right)$ as required. For the second identity, we replace $A$ by $A^T$ and note $\text{col}\left(A^T\right) = \text{row}(A)$. □

For an arbitrary $m \times n$ matrix $A$, we see that $\text{row}(A)$ and $\text{null}(A)$ are orthogonal complements in $\mathbf{R}^n$ whereas $\text{col}(A)$ and $\text{null}\left(A^T\right)$ are orthogonal complements in $\mathbf{R}^m$. These four subspaces are called the *fundamental subspaces* of $A$. The particular case of a symmetric $n \times n$ matrix, we have $\text{row}(A) = \text{col}(A)$ and $\text{null}(A)$ are orthogonal complements in $\mathbf{R}^n$.

## Orthogonal Complements

Combining the above two theorems, we obtain

**Theorem 11.18**

*Let $A$ be a $m \times n$ matrix. Then*

$$\text{rk}(A) + \text{nullity}(A) = n$$
$$\text{rk}(A) + \text{nullity}\left(A^T\right) = m$$

The first of these equations is the rank equation.

# Orthogonal Complements

The linear transformation associated with $A$, $T_A(\mathbf{v}) = A\,\mathbf{v}$, is a mapping from $\mathbf{R}^n$ into $\mathbf{R}^m$. For $\mathbf{v} \in \mathbf{R}^n$, we have, by the orthogonal decomposition theorem,

$$\mathbf{v} = \mathbf{v}_1 + \mathbf{v}_2$$

where

$$\mathbf{v}_1 = \text{proj}_{\text{row}\,(A)}\,(\mathbf{v}) \in \text{row}\,(A) \quad \text{and} \quad \mathbf{v}_2 = \text{proj}_{\text{null}\,(A)}\,(\mathbf{v}) \in \text{null}\,(A).$$

Thus

$$A\,\mathbf{v} = A\,(\mathbf{v}_1 + \mathbf{v}_2) = A\,\mathbf{v}_1.$$

Therefore the projection of $\mathbf{v}$ onto null $(A)$ represents the information lost when applying the transformation $T_A$. If we restrict the domain of $T_A$ to row $(A)$ then we will get an invertible map from row $(A)$ onto col $(A)$.

# Least Squares

Reference: §7.3 in Poole.

## Problem

We have data from an experiment in the form of ordered pairs $(x_i, y_i)$. We want to find a relationship $y = f(x)$ that fits the data "well".

Suppose we wish to "fit" a straight line to a set of data $(x_i, y_i)$; that is we wish to find $c_0$ and $c_1$ such that

$$y_i = c_0 + c_1\,x_i$$

If the relationship is exact and there is no error in the measurements then only two data points would be needed to determine $c_0$ and $c_1$. In practice the linear relationship may only be an approximation and there will be error in the measurements. So more than two measurements are made and we try to find $c_0$ and $c_1$ that "best" fit the data

$$y_i \approx c_0 + c_1\,x_i;$$