Math 260 Section 6.4

Most important ideas:

- Gram-Schmidt Process: given a (non-orthogonal, but linearly independent) set of vectors, create a new, orthogonal set of vectors which span the same vector space.
- This process also leads to a useful factorization of a matrix, the QR factorization.
- Recall for a matrix times a matrix: $\boldsymbol{A}\,[\,\vec{\boldsymbol{b}}_1\ \vec{\boldsymbol{b}}_2 \cdots \vec{\boldsymbol{b}}_n\,] = [\,\boldsymbol{A}\vec{\boldsymbol{b}}_1\ \boldsymbol{A}\vec{\boldsymbol{b}}_2 \cdots \boldsymbol{A}\vec{\boldsymbol{b}}_n\,]$, and

  recall for a matrix times a vector: $[\,\vec{\boldsymbol{a}}_1\ \vec{\boldsymbol{a}}_2 \cdots \vec{\boldsymbol{a}}_n\,]\begin{bmatrix}\boldsymbol{x}_1\\ \boldsymbol{x}_2\\ \vdots\\ \boldsymbol{x}_n\end{bmatrix} = \boldsymbol{x}_1\vec{\boldsymbol{a}}_1 + \boldsymbol{x}_2\vec{\boldsymbol{a}}_2 + \cdots + \boldsymbol{x}_n\vec{\boldsymbol{a}}_n$.

First, an idea that will be useful to us a little later in this section: Suppose that we have a set of linearly independent vectors $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, and that we have three other vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$, each of which is a linear combination of $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$. **Then any vector that we can build using $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$ we can also build using $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$; that is, $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$.**

Example 1 (keep in mind $\vec{x}_1, \vec{x}_2, \vec{x}_3$ and $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are all vectors):

Suppose:

$$\begin{aligned}\vec{v}_1 &= -11\vec{x}_1 + 10\vec{x}_2 + 4\vec{x}_3\\ \vec{v}_2 &= 9\vec{x}_1 + 7\vec{x}_2\\ \vec{v}_3 &= 4\vec{x}_1 - 2\vec{x}_2 + \vec{x}_3\end{aligned}$$

That is:

$$\underbrace{[\,\vec{\boldsymbol{v}}_1\ \vec{\boldsymbol{v}}_2\ \vec{\boldsymbol{v}}_3\,]}_{\boldsymbol{V}} = \underbrace{[\,\vec{\boldsymbol{x}}_1\ \vec{\boldsymbol{x}}_2\ \vec{\boldsymbol{x}}_3\,]}_{\boldsymbol{X}}\underbrace{\begin{bmatrix}-11 & 9 & 4\\ 10 & 7 & -2\\ 4 & 0 & 1\end{bmatrix}}_{\boldsymbol{A}}.$$

That is, $V = XA$, so matrix A tells us how to build the $\vec{v}$ vectors out of the $\vec{x}$ vectors. And since $X = VA^{-1}$ (assuming A^{-1} exists) then A^{-1} tells us how to build the $\vec{x}$ vectors from the $\vec{v}$ vectors. We can go in either direction, from the $\vec{x}$ vectors to the $\vec{v}$ vectors or from the $\vec{v}$ vectors to the $\vec{x}$ vectors.

Where $\vec{w} \in span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$. Then

$$\vec{\boldsymbol{w}} = \boldsymbol{c}_1\vec{\boldsymbol{x}}_1 + \boldsymbol{c}_2\vec{\boldsymbol{x}}_2 + \boldsymbol{c}_3\vec{\boldsymbol{x}}_3 = [\,\vec{\boldsymbol{x}}_1\ \vec{\boldsymbol{x}}_2\ \vec{\boldsymbol{x}}_3\,]\begin{bmatrix}\boldsymbol{c}_1\\ \boldsymbol{c}_2\\ \boldsymbol{c}_3\end{bmatrix} = \boldsymbol{X}\vec{\boldsymbol{c}} = \boldsymbol{V}\boldsymbol{A}^{-1}\vec{\boldsymbol{c}} = \boldsymbol{V}\vec{\boldsymbol{d}} = [\,\vec{\boldsymbol{v}}_1\ \vec{\boldsymbol{v}}_2\ \vec{\boldsymbol{v}}_3\,]\begin{bmatrix}\boldsymbol{d}_1\\ \boldsymbol{d}_2\\ \boldsymbol{d}_3\end{bmatrix} = \boldsymbol{d}_1\vec{\boldsymbol{v}}_1 + \boldsymbol{d}_2\vec{\boldsymbol{v}}_2 + \boldsymbol{d}_3\vec{\boldsymbol{v}}_3$$

where $\begin{bmatrix}d_1\\ d_2\\ d_3\end{bmatrix} = \vec{d} = A^{-1}\vec{c}$.

Vector $\vec{c}$: how to build $\vec{w}$ out of $\vec{x}_1, \vec{x}_2, \vec{x}_3$. Vector $\vec{d}$: how to build $\vec{w}$ out of $\vec{v}_1, \vec{v}_2, \vec{v}_3$.

So we see that if $\vec{w} \in span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, then $\vec{w} \in span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$.

Conclusion: assuming A^{-1} exists (that is, assuming we build the new vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ out of the old vectors $\vec{x}_1, \vec{x}_2, \vec{x}_3$ in an invertible way), then every vector that can be built using the old vectors $\vec{x}_1, \vec{x}_2, \vec{x}_3$ can also be built using the new vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$, which themselves can be built out of the old vectors. That is, $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$.

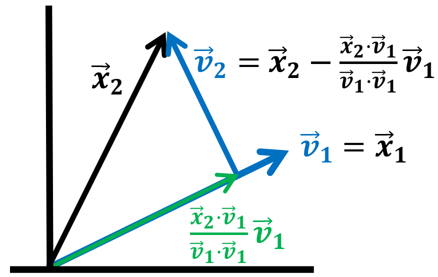
The above idea will come up a little later today. For now, we move on to another idea.

Main goal for this section: given a set of linearly independent vectors $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, build a new set of mutually <u>orthogonal</u> vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ so that $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$.

<u>Step 1</u> Given $\vec{x}_1$ and $\vec{x}_2$, find $\vec{v}_1$ and $\vec{v}_2$ that are mutually orthogonal (orthogonal to each other) and so that $span\{\vec{v}_1, \vec{v}_2\} = span\{\vec{x}_1, \vec{x}_2\}$.

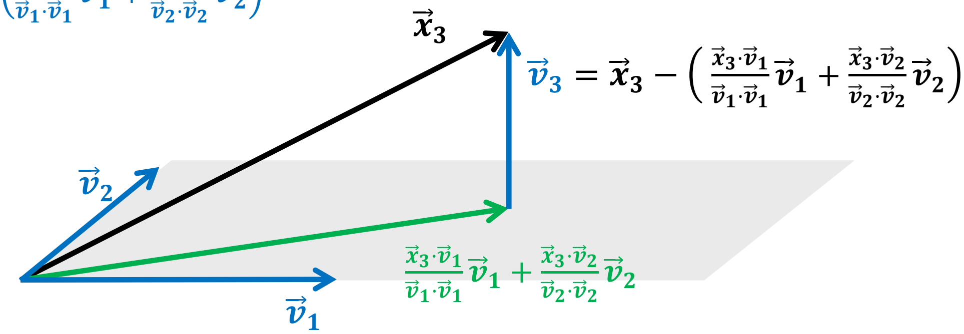
$$\vec{v}_1 = \vec{x}_1$$

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



<u>Step 2</u> Now given a third vector $\vec{x}_3$, find $\vec{v}_3$ so that $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ are mutually orthogonal and so that $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$.

$$\vec{v}_3 = \vec{x}_3 - \left(\frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2\right)$$



So:

$$\vec{v}_1 = \vec{x}_1$$

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

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

That is:

$$\begin{aligned} \vec{v}_1 \qquad\qquad\qquad &= \vec{x}_1 \\ \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \qquad \vec{v}_2 \qquad &= \vec{x}_2 \\ \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1}\vec{v}_1 + \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2}\vec{v}_2 + \vec{v}_3 &= \vec{x}_3 \end{aligned}$$

That is:

$$[\ \vec{v}_1\ \ \vec{v}_2\ \ \vec{v}_3\ ]\begin{bmatrix} 1 & \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} & \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \\ 0 & 1 & \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \\ 0 & 0 & 1 \end{bmatrix} = [\ \vec{x}_1\ \ \vec{x}_2\ \ \vec{x}_3\ ]$$

That is:

$$[\ \vec{v}_1\ \ \vec{v}_2\ \ \vec{v}_3\ ]\ R = [\ \vec{x}_1\ \ \vec{x}_2\ \ \vec{x}_3\ ]$$

So that:

$$[\ \vec{v}_1\ \ \vec{v}_2\ \ \vec{v}_3\ ] = [\ \vec{x}_1\ \ \vec{x}_2\ \ \vec{x}_3\ ]R^{-1}$$

So we see the new vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ are built—in an invertible way (how can we easily see that R is invertible?)—from the original vectors $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, and consequently that $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$, as described on the previous page.

See Theorem 11 on page 355 for the more general version of this process.

Example: Suppose we are given $\{\vec{x}_1, \vec{x}_2, \vec{x}_3\} = \{\begin{bmatrix} -1 \\ 3 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ -9 \\ -2 \\ -2 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 6 \\ 2 \end{bmatrix}\}$.

Find a set of orthogonal vectors $\{\vec{v}_1, \vec{v}_2, \vec{v}_3\}$ so that $span\{\vec{v}_1, \vec{v}_2, \vec{v}_3\} = span\{\vec{x}_1, \vec{x}_2, \vec{x}_3\}$.

Using the Gram-Schmidt process:

$$\vec{v}_1 = \vec{x}_1 = \begin{bmatrix} -1 \\ 3 \\ 1 \\ 1 \end{bmatrix}$$

$$\vec{v}_2 = \vec{x}_2 - \frac{\vec{x}_2 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 = \begin{bmatrix} 5 \\ -9 \\ -2 \\ -2 \end{bmatrix} - \frac{-36}{12} \begin{bmatrix} -1 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 5 \\ -9 \\ -2 \\ -2 \end{bmatrix} - (-3) \begin{bmatrix} -1 \\ 3 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix}.$$

$$\vec{v}_3 = \vec{x}_3 - \frac{\vec{x}_3 \cdot \vec{v}_1}{\vec{v}_1 \cdot \vec{v}_1} \vec{v}_1 - \frac{\vec{x}_3 \cdot \vec{v}_2}{\vec{v}_2 \cdot \vec{v}_2} \vec{v}_2 = \begin{bmatrix} 2 \\ 2 \\ 6 \\ 2 \end{bmatrix} - \frac{12}{12} \begin{bmatrix} -1 \\ 3 \\ 1 \\ 1 \end{bmatrix} - \frac{12}{6} \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 6 \\ 2 \end{bmatrix} - 1 \begin{bmatrix} -1 \\ 3 \\ 1 \\ 1 \end{bmatrix} - 2 \begin{bmatrix} 2 \\ 0 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \\ 3 \\ -1 \end{bmatrix}.$$

Notice that vectors $\vec{v}_1, \vec{v}_2, \vec{v}_3$ are mutually orthogonal: $\vec{v}_1 \cdot \vec{v}_2 = 0$, $\vec{v}_1 \cdot \vec{v}_3 = 0$, $\vec{v}_2 \cdot \vec{v}_3 = 0$. (I chose vectors so that all of the numbers would work out nicely. Usually that would not be the case, but computers don't care how nice or messy the numbers are.)

Notice that
$$\begin{aligned} \vec{v}_1 &= \vec{x}_1 \\ \vec{v}_2 &= \vec{x}_2 - (-3)\,\vec{v}_1 \\ \vec{v}_3 &= \vec{x}_3 - 1\,\vec{v}_1 - 2\,\vec{v}_2 \end{aligned}$$
, that is,
$$\begin{aligned} \vec{v}_1 &= \vec{x}_1 \\ (-3)\vec{v}_1 + \vec{v}_2 &= \vec{x}_2 \\ 1\,\vec{v}_1 + 2\,\vec{v}_2 + \vec{v}_3 &= \vec{x}_3 \end{aligned}$$
, that is,

$$\begin{bmatrix} -1 & 2 & -1 \\ 3 & 0 & -1 \\ 1 & 1 & 3 \\ 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} 1 & -3 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 5 & 2 \\ 3 & -9 & 2 \\ 1 & -2 & 6 \\ 1 & -2 & 2 \end{bmatrix}$$

Notation: $[\ \vec{v}_1\ \ \vec{v}_2\ \ \vec{v}_3\ ]\quad R \quad = \quad [\ \vec{x}_1\ \ \vec{x}_2\ \ \vec{x}_3]$

That is: $Q \quad R \quad = \quad A$

We could make the orthogonal vectors orthonormal by dividing each one by its own length. (You'd also have to modify matrix R.) This makes the matrix factorization a bit more useful:

$$\begin{bmatrix} -1/\sqrt{12} & 2/\sqrt{6} & -1/\sqrt{12} \\ 3/\sqrt{12} & 0/\sqrt{6} & -1/\sqrt{12} \\ 1/\sqrt{12} & 1/\sqrt{6} & 3/\sqrt{12} \\ 1/\sqrt{12} & 1/\sqrt{6} & -1/\sqrt{12} \end{bmatrix} \begin{bmatrix} 1\sqrt{12} & -3\sqrt{12} & 1\sqrt{12} \\ 0 & 1\sqrt{6} & 2\sqrt{6} \\ 0 & 0 & 1\sqrt{12} \end{bmatrix} = \begin{bmatrix} -1 & 5 & 2 \\ 3 & -9 & 2 \\ 1 & -2 & 6 \\ 1 & -2 & 2 \end{bmatrix}$$

The above is the QR factorization of the matrix A (the columns of A are the three vectors $\vec{x}_1, \vec{x}_2, \vec{x}_3$) where Q is an orthogonal matrix (so $Q^T Q = I$) and R is upper triangular. This is pretty useful: given $A\vec{x} = \vec{b}$ we can write $QR\vec{x} = \vec{b}$, which multiplying both sides by Q^T (which is the left inverse of Q) leads to $R\vec{x} = Q^T\vec{b}$, and solving for $\vec{x}$ is relatively easy since R is upper triangular.

Given matrix A, it's not too difficult to find Q, whose columns are the orthogonalized (using Gram-Schmidt) columns of A, but sometimes one can get a little confused in finding R. Here is another approach, as described in the book.

Once you find Q (the version of Q with ortho<u>normal</u> columns), then

$$QR = A \Rightarrow Q^T QR = Q^T A, \text{ i.e. } R = Q^T A$$

since $Q^T Q = I$ (since the columns of Q are orthonormal). So for the above example we have

$$R = Q^T A = \begin{bmatrix} -1/\sqrt{12} & 3/\sqrt{12} & 1/\sqrt{12} & 1/\sqrt{12} \\ 2/\sqrt{6} & 0/\sqrt{6} & 1/\sqrt{6} & 1/\sqrt{6} \\ -1/\sqrt{12} & -1/\sqrt{12} & 3/\sqrt{12} & -1/\sqrt{12} \end{bmatrix} \begin{bmatrix} -1 & 5 & 2 \\ 3 & -9 & 2 \\ 1 & -2 & 6 \\ 1 & -2 & 2 \end{bmatrix}$$

$$= \begin{bmatrix} 12/\sqrt{12} & -36/\sqrt{12} & 12/\sqrt{12} \\ 0 & 6/\sqrt{6} & 12/\sqrt{6} \\ 0 & 0 & 12/\sqrt{12} \end{bmatrix}$$

$$= \begin{bmatrix} \sqrt{12} & -3\sqrt{12} & \sqrt{12} \\ 0 & \sqrt{6} & 2\sqrt{6} \\ 0 & 0 & \sqrt{12} \end{bmatrix}$$

as we found on the previous page.