Worksheet 17: Gram-Schmidt Orthogonalization and QR factorization (§5.2)

Theorem: An orthonormal set of vectors in $\mathbb{R}^n$ is linearly independent.

Corollary: Let W be a d-dimensional subspace of $\mathbb{R}^n$. Any orthonormal set of d vectors in W is an **orthonormal basis** for W.

The *Gram Schmidt Orthogonalization Process* is an algorithm to transform any given basis of a subspace W of $\mathbb{R}^n$ into an orthonormal basis.

Problem 1. Let W be the subspace of $\mathbb{R}^4$ consisting of the solutions of the system of equations

$$\begin{array}{ccccc} x_1 & -x_2 & -2x_3 & & = 0, \\ & x_2 & +x_3 & -2x_4 & = 0. \end{array}$$

(a) Find an orthonormal basis for W.
[Hint: First find some basis, then play around with it to find an orthonormal basis.]

(b) Compute the orthogonal projection of $\vec{x} = \begin{bmatrix} 1 & 2 & -1 & 2 \end{bmatrix}^\top \in \mathbb{R}^4$ onto W. [Hint: Use (a).]

Solution:

(a) There are many correct answers. The standard procedure for finding a basis for the solution set gives

$$(\vec{u}, \vec{v}) = \left(\begin{bmatrix} 1 \\ -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 0 \\ 1 \end{bmatrix} \right),$$

which is already orthogonal, so normalizing gives the orthonormal basis $\left(\frac{1}{\sqrt{3}}\vec{u}, \frac{1}{3}\vec{v}\right)$.

(b) Using the orthonormal basis $\vec{u}, \vec{v}$, we have

$$\mathrm{proj}_W(\vec{x}) = \left(\vec{x} \cdot \frac{1}{\sqrt{3}}\vec{u}\right) \frac{1}{\sqrt{3}}\vec{u} + \left(\vec{x} \cdot \frac{1}{3}\vec{v}\right) \frac{1}{3}\vec{v} = \frac{1}{9} \begin{bmatrix} 10 \\ 22 \\ -6 \\ 8 \end{bmatrix}.$$

Problem 2. Let $\mathcal{B} = (\vec{b}_1, \vec{b}_2)$ be a basis for $\mathbb{R}^2$, where $\vec{b}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \vec{b}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$.

(a) Draw a sketch showing the basis $\mathcal{B}$ for $\mathbb{R}^2$.

(b) Use the Gram-Schmidt process to construct an orthonormal basis $\mathcal{A} = (\vec{u}_1, \vec{u}_2)$ from $\mathcal{B}$. Show $\vec{u}_1, \vec{u}_2$ and other relevant vectors in your picture.

(c) Find the change of basis matrix $S_{\mathcal{B}\to\mathcal{A}}$. Why is it easier to find than $S_{\mathcal{A}\to\mathcal{B}}$? Why is it upper triangular?

Solution: Using Gram-Schmidt we get $\vec{u}_1 = \frac{1}{\sqrt{2}}\vec{b}_1$. To find $\vec{u}_2$, we decompose $\vec{b}_2$ into a component parallel to $\vec{u}_1$ and a component orthogonal to $\vec{u}_1$. The component parallel to $\vec{u}_1$ is the projection onto $\vec{u}_1$, which is

$$(\vec{b}_2 \cdot \vec{u}_1)\vec{u}_1, \quad \text{which is} \quad \frac{(\vec{b}_2 \cdot \vec{b}_1)}{(\vec{b}_1 \cdot \vec{b}_1)}\vec{b}_1 = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}.$$

Thus the component of $\vec{b}_2$ in the perpendicular direction is $\vec{b}_2^{\perp} = \vec{b}_2 - \vec{b}_2^{\parallel} = \begin{bmatrix} -1 \\ 2 \end{bmatrix} - \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} = \begin{bmatrix} -3/2 \\ 3/2 \end{bmatrix}$. Finally, we need to scale $\vec{b}_2^{\perp}$ so that it is a unit vector to get $\vec{u}_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix}$. So an orthonormal basis is $\left\{ \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}$.

The change of basis matrix is

$$S_{\mathcal{B}\to\mathcal{A}} = \begin{bmatrix} \vec{b}_1 \cdot \vec{u}_1 & \vec{b}_2\vec{u}_1 \\ \vec{b}_1 \cdot \vec{u}_2 & \vec{b}_2\vec{u}_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2} & 1/\sqrt{2} \\ 0 & 3/\sqrt{2} \end{bmatrix},$$

which is easy to find since we are converting *to* an orthonormal basis, so we can compute the coordinates using the dot product (Problem 3d on worksheet 16). It is upper triangular because the Gram-Schmidt process always picks $\vec{u}_2$ to be perpendicular to $\vec{b}_1$, so the entry in the $2-1$ spot of $S_{\mathcal{B}\to A}$ is $\vec{b}_1 \cdot \vec{u}_2 = 0$.

QR Factorization Theorem: Let M be an $n \times d$ matrix of rank d. Then there is a unique way to write M as

$$M = Q\ R.$$

where Q is an $n \times d$ matrix whose columns are orthonormal and R is a $d \times d$ upper triangular matrix with positive entries on the diagonal.

Technique to Find the QR factorization: View the columns of M as a basis $\mathcal{M}$ for a d-dimensional subspace of $\mathbb{R}^n$. Use Gram Schmidt to compute an orthonormal basis $\mathcal{Q}$: these are the columns of Q. The matrix R is the change of basis matrix $S_{\mathcal{M}\to\mathcal{Q}}$.

Problem 3. Consider the two bases for a subspace V of $\mathbb{R}^3$:

$$\mathcal{B} = \left\{ \begin{bmatrix} 3 \\ 4 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 11 \\ 12 \end{bmatrix} \right\}, \qquad \mathcal{A} = \left\{ \begin{bmatrix} 3/5 \\ 4/5 \\ 0 \end{bmatrix}, \begin{bmatrix} -4/13 \\ 3/13 \\ 12/13 \end{bmatrix} \right\}.$$

(a) Prove that $\mathcal{A}$ is an orthonormal basis.

(b) Find the change of basis matrix $S_{\mathcal{B}\to\mathcal{A}}$. Be smart! Use the fact that $\mathcal{A}$ is orthonormal!

(c) Use the Gram Schmidt process to orthogonalize the basis $\mathcal{B}$.

(d) Find the QR factorization of $\begin{bmatrix} 3 & 2 \\ 4 & 11 \\ 0 & 12 \end{bmatrix}$ using the technique above. Check your result by multiplying out the matrices.

Solution:

(a) Just check the dot product $\begin{bmatrix}3/5\\4/5\\0\end{bmatrix}\cdot\begin{bmatrix}-4/13\\3/13\\12/13\end{bmatrix}=0$ and also that $\begin{bmatrix}3/5\\4/5\\0\end{bmatrix}\cdot\begin{bmatrix}3/5\\4/5\\0\end{bmatrix}=1$ and $\begin{bmatrix}-4/13\\3/13\\12/13\end{bmatrix}\cdot\begin{bmatrix}-4/13\\3/13\\12/13\end{bmatrix}=1$, so both vectors are unit length.

(b) $S_{\mathcal{B}\to\mathcal{A}}$ is the 2×2 matrix $\begin{bmatrix}5&10\\0&13\end{bmatrix}$. We found the coordinates using dot product, not solving a system, which would be more complicated.

(c) The Gram Schmidt process turns $\mathcal{B}$ into $\mathcal{A}$.

(d) $\begin{bmatrix}3&2\\4&11\\0&12\end{bmatrix}=\begin{bmatrix}3/5&-4/13\\4/5&3/13\\0&12/13\end{bmatrix}\begin{bmatrix}5&10\\0&13\end{bmatrix}$.

Problem 5. Let $\Lambda\subseteq\mathbb{R}^3$ be the plane given by $x+y+z=0$. Let $\pi:\mathbb{R}^3\to\mathbb{R}^3$ be the orthogonal projection onto Λ.

(a) Find an orthonormal basis $\mathcal{U}$ for Λ, and an extension $\mathcal{U}'$ to an orthonormal basis for $\mathbb{R}^3$.

(b) Describe the kernel and image of π both geometrically and by giving bases.

(c) Find the $\mathcal{U}'$-matrix of π. Discuss how to find the standard matrix of π (there are at least three ways!). Is $[\pi]_{\mathcal{U}'}$ or $[\pi]_{\mathcal{E}}$ easier to find?

(d) Write a matrix equation expressing $[\pi]_{\mathcal{E}}$ in terms of $[\pi]_{\mathcal{U}'}$ and other well-chosen explicit matrices and their inverses.

Solution:

(a) There are many correct answers. To find one, we start with a basis for Λ, say $\left\{\begin{bmatrix}1\\-1\\0\end{bmatrix},\begin{bmatrix}0\\-1\\1\end{bmatrix}\right\}$.

Orthonormalize it using Gram-Schmidt:

$$\mathcal{U}=\left\{\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix},\frac{2}{\sqrt{6}}\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\\1\end{bmatrix}\right\}.$$

To get an orthonormal basis for $\mathbb{R}^3$ extending $\mathcal{U}$, we need a unit vector perpendicular to these. Since $\Lambda^\perp$ is spanned by $\begin{bmatrix}1\\1\\1\end{bmatrix}$, we can take

$$\mathcal{U}'=\left\{\frac{1}{\sqrt{2}}\begin{bmatrix}1\\-1\\0\end{bmatrix},\frac{2}{\sqrt{6}}\begin{bmatrix}-\frac{1}{2}\\-\frac{1}{2}\\1\end{bmatrix},\frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}.$$

(b) The kernel of π is the line normal to Λ through the origin, or Span $\left(\begin{bmatrix}1\\1\\1\end{bmatrix}\right)$. The image is W, or Span $\mathcal{U}$.

(c) The $\mathcal{U}'$ matrix is much easier to find! It is

$$[\pi]_{\mathcal{U}'} = \begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 0\end{bmatrix}$$

To find the standard matrix of π, we could use the key theorem to compute $\pi(\vec{e}_i)$, which is not so bad but involves three computations of the type $(\vec{e}_i \cdot \vec{u}_1)\vec{u}_1 + (\vec{e}_i \cdot \vec{u}_2)\vec{u}_2 + (\vec{e}_i \cdot \vec{u}_3)\vec{u}_3$. Or we could use Theorem 5.3.10 (also covered on the last worksheet) showing that the standard matrix is the product $Q\,Q^\top$ where $Q = \begin{bmatrix}\vec{u}_1 & \vec{u}_2\end{bmatrix}$. The third way is to use change of basis (see part d).

(d) We know that $[\pi]_{\mathcal{E}} = S_{\mathcal{U}'\to\mathcal{E}}[\pi]_{\mathcal{U}'}S_{\mathcal{E}\to\mathcal{U}'}$. The change of basis matrix $S_{\mathcal{U}'\to\mathcal{E}}$ is easy to find. It is $S_{\mathcal{U}'\to\mathcal{E}} = \begin{bmatrix}\vec{u}_1 & \vec{u}_2 & \vec{u}_3\end{bmatrix}$ where the $\vec{u}_i$ are the elements of $\mathcal{U}'$.

Problem 6. Another Use of the Change of Basis Matrix. Consider two ordered bases for a subspace W of $\mathbb{R}^3$, $\mathcal{A} = (\vec{a}_1, \vec{a}_2)$ and $\mathcal{B} = (\vec{b}_1, \vec{b}_2)$, and let A be the 3×2 matrix $[\vec{a}_1\ \vec{a}_2]$ and let B be the 3×2 matrix $[\vec{b}_1\ \vec{b}_2]$.

(a) In the special case where

$$\vec{b}_1 = \begin{bmatrix}6\\4\\-1\end{bmatrix}, \quad \vec{b}_2 = \begin{bmatrix}2\\-2\\1\end{bmatrix}, \quad \vec{a}_1 = \begin{bmatrix}4\\1\\0\end{bmatrix}, \quad \vec{a}_2 = \begin{bmatrix}2\\3\\-1\end{bmatrix},$$

find the change of basis matrix $S_{\mathcal{B}\to\mathcal{A}}$.

(b) For the special case given in (a), compute the product $A\ S_{\mathcal{B}\to A}$ explicitly and compare to B.

(c) Now show that for *any two bases* $\mathcal{A}$ and $\mathcal{B}$ for W, $B = A\ S_{\mathcal{B}\to A}$.
[Hint: Write A and B as a "row of columns." Recall our *definition* of matrix multiplication, multiplying A by each column of $S_{\mathcal{B}\to\mathcal{A}}$.]

(d) Discuss the meaning and the proof of the Theorem below.

Matrix Factorization Theorem: *Let $\mathcal{B} = (\vec{v}_1, \vec{v}_2, \dots \vec{v}_d)$ and $\mathcal{A} = (\vec{w}_1, \vec{w}_2, \dots \vec{w}_d)$ be two ordered bases for a d-dimensional subspace W of $\mathbb{R}^n$. Then we have a matrix product*

$$[\vec{v}_1\ \vec{v}_2\ \ \dots\ \ \vec{v}_d] = [\vec{w}_1\ \vec{w}_2\ \ \dots\ \ \vec{w}_d]\ S_{\mathcal{B}\to\mathcal{A}}.$$

Solution:

(a) $S_{\mathcal{B}\to\mathcal{A}} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$. You can find this by solving a linear system for each column—for example, to find the first column we need to solve $\begin{bmatrix} 6 \\ 4 \\ -1 \end{bmatrix} = x \begin{bmatrix} 4 \\ 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 2 \\ 3 \\ -1 \end{bmatrix}$. In this case, it is actually easy to just "eyeball" the situation.

(b) We get $[\vec{b}_1 \ \ \vec{b}_2] = [\vec{a}_1 \ \ \vec{a}_2] \ S_{\mathcal{B}\to A}$.

(c) The first column of $[\vec{a}_1 \ \ \vec{a}_2] \ S_{\mathcal{B}\to A}$ is $[\vec{a}_1 \ \ \vec{a}_2] \begin{bmatrix} \lambda \\ \mu \end{bmatrix}$ where $\begin{bmatrix} \lambda \\ \mu \end{bmatrix}$ is the first column of $S_{\mathcal{B}\to A}$. This is then $\lambda\vec{a}_1 + \mu\vec{a}_2$, which is a linear combination of the basis elements of $\mathcal{A}$. But remembering that the first column of $S_{\mathcal{B}\to\mathcal{A}}$ is supposed to express $\vec{b}_1$ as a combination of the basis elements of $\mathcal{A}$, we see that this linear combination must be $\vec{b}_1$! Similarly, the second column of $[\vec{a}_1 \ \ \vec{a}_2] = S_{\mathcal{B}\to A}$ is $\vec{b}_2$. QED.

(d) In the earlier part of this problem we had a special case of this theorem with $d = 2$ and $n = 3$. Now, we have the matrix $[\vec{v}_1 \ \vec{v}_2 \ \ \ldots \ \ \vec{v}_d]$ which is $n \times d$. We also have the $n \times d$ matrix $[\vec{w}_1 \ \vec{w}_2 \ \ \ldots \ \ \vec{w}_d]$ and the invertible $d \times d$ matrix $S_{\mathcal{B}\to\mathcal{A}}$. The first column of $S_{\mathcal{B}\to\mathcal{A}}$ is the column of $\mathcal{A}$-coordinates of $\vec{v}_1$. This means that $\vec{v}_1 = a_1\vec{w}_1 + a_2\vec{w}_2 + \cdots + a_d\vec{w}_d$. This is the product $\vec{v}_1 = [\vec{w}_1 \ \vec{w}_2 \ \ \ldots \ \ \vec{w}_d] \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_d \end{bmatrix}$. So since $S_{\mathcal{B}\to A} = \big[[\vec{v}_1]_{\mathcal{A}} \ \ [\vec{v}_2]_{\mathcal{A}} \ \ \ldots \ \ [\vec{v}_d]_{\mathcal{A}}\big]$, then using block multiplication, we have

$$[\vec{w}_1 \ \vec{w}_2 \ \ \ldots \ \ \vec{w}_d] \ \big[[\vec{v}_1]_{\mathcal{A}} \ \ [\vec{v}_2]_{\mathcal{A}} \ \ \ldots \ \ [\vec{v}_d]_{\mathcal{A}}\big] = [\vec{v}_1 \ \vec{v}_2 \ \ \ldots \ \ \vec{v}_d].$$

This proves the theorem.

Problem 7. Proof of QR factorization. Suppose $\mathcal{B} = (\vec{v}_1, \ldots, \vec{v}_r)$ is a basis of the subspace V of $\mathbb{R}^n$, and let $\mathcal{U} = (\vec{u}_1, \ldots, \vec{u}_r)$ be the orthonormal basis obtained by applying the Gram-Schmidt process to $\mathcal{B}$. That is, set $\vec{u}_1 = \frac{\vec{v}_1}{\|\vec{v}_1\|}$, and then for each $1 \le k < r$ let

$$\vec{u}_{k+1} = \frac{\vec{w}_{k+1}}{\|\vec{w}_{k+1}\|} \qquad \text{where} \qquad \vec{w}_{k+1} = \vec{v}_{k+1} - \sum_{i=1}^{k} (\vec{v}_{k+1} \cdot \vec{u}_i)\vec{u}_i. \tag{1}$$

(a) Find the change-of-coordinates matrix $S_{\mathcal{B}\to\mathcal{U}}$ in terms of the vectors $\vec{v}_i$, $\vec{w}_i$ and $\vec{u}_i$. Explain why it is upper triangular with positive numbers on the diagonal. [HINT: Use (1) and show that $\vec{v}_i \cdot \vec{u}_i = \|\vec{w}_i\|$.]

(b) Deduce the existence of the QR Factorization for an arbitrary $n \times d$ matrix M from the Matrix Factorization theorem above.

Solution: $S_{\mathcal{B}\to\mathcal{U}} =$

$$\begin{bmatrix} \|\vec{v}_1\| & \vec{u}_1 \cdot \vec{v}_2 & \vec{u}_1 \cdot \vec{v}_3 & \cdots & \cdots & \vec{u}_1 \cdot \vec{v}_r \\ 0 & \|\vec{w}_2\| & \vec{u}_2 \cdot \vec{v}_3 & \cdots & \cdots & \vec{u}_2 \cdot \vec{v}_r \\ \vdots & 0 & \|\vec{w}_3\| & \cdots & \cdots & \vec{u}_3 \cdot \vec{v}_r \\ \vdots & \vdots & 0 & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \|\vec{w}_r\| \end{bmatrix}, \text{ i.e., } S_{\mathcal{B}\to\mathcal{U}}(i,j) = \begin{cases} \vec{u}_i \cdot \vec{v}_j & \text{if } i < j; \\ \|\vec{w}_i\| & \text{if } i = j; \\ 0 & \text{otherwise.} \end{cases}$$

The QR factorization theorem follows immediately: we let Q be the matrix whose columns are $\vec{u}_1, \ldots, \vec{u}_r$ and note that $R = S_{\mathcal{B}\to\mathcal{U}}$, which is upper triangular by (a).