1 Recap

- Orthogonality, orthonormal bases.
- Gram-Schmidt procedure (example).

2 Proof of GS

The proof of properties $(a),(b)$ is by constructing the y_i inductively/recursively. Taking cue from the $\mathbb{R}^2$ example, let $y_1 = x_1$. (Caveat: x_1 and hence y_1 is allowed to be 0.) Define y_2 as $x_2 - ax_1$ for some, as of now, undetermined $a \in \mathbb{F}$. Note that $y_2 + ay_1 = x_2$ and hence $L(x_1, x_2) = L(y_1, y_2)$ (Why ?)
We want y_2 to be orthogonal to $L(y_1) = L(x_1) = \{cx_1 \mid c \in \mathbb{F}\}$. Thus $0 = \langle y_2, cy_1 \rangle = c\langle y_2, y_1 \rangle$. This happens if and only if $0 = \langle y_2, x_1 \rangle = \langle x_2, x_1 \rangle - a\langle x_1, x_1 \rangle$. If $x_1 = 0$, define $y_2 = x_2$. Otherwise, $a = \frac{\langle x_2, x_1 \rangle}{\langle x_1, x_1 \rangle}$. For y_3, as before, we define $y_3 = x_3 - a_1 y_1 - a_2 y_2$. Then $0 = \langle y_3, y_1 \rangle = \langle x_3, y_1 \rangle - a_1 \langle y_1, y_1 \rangle$. Likewise, $0 = \langle y_3, y_2 \rangle = \langle x_3, y_2 \rangle - a_2 \langle y_2, y_2 \rangle$. If $y_i = 0$, define $a_i = 0$. If not, $a_i = \frac{\langle x_3, y_i \rangle}{\langle y_i, y_i \rangle}$.

Proof of $(a),(b)$: We see a pattern. Assume that $y_1, \ldots, y_k$ have been defined satisfying the first two properties. Define $y_{k+1} = x_{k+1} - \sum_i a_i y_i$ where $a_i = 0$ if $y_i = 0$ and $a_i = \frac{\langle x_{k+1}, y_i \rangle}{\langle y_i, y_i \rangle}$. Therefore, y_{k+1} is orthogonal to each of the y_i and hence to $L(y_1, y_2, \ldots, y_k)$. Therefore, the first property is met by y_{k+1}. The second property is a HW exercise.
Proof of (c): Now we prove property (c) by induction again. The case $k = 1$ is easy because $L(y'_1) = L(x_1)$ implies that $y'_1 = cx_1 = cy_1$. Assume truth for k, i.e., $y'_i = c_i y_i$. We shall prove for $k+1$. Suppose we have an element y'_{k+1} satisfying both properties, i.e., y'_{k+1} is orthogonal to $L(y_1, \ldots, y_k) = L(y'_1, y'_2, \ldots)$ and $L(y'_1, y'_2, \ldots, y'_{k+1}) = L(y_1, y_2, \ldots, y_{k+1}) = L(x_1, x_2, \ldots, x_{k+1})$. By the second property, $y'_{k+1} = \sum_{i=1}^{k+1} a_i y_i = z + a_{k+1} y_{k+1}$ where $z \in L(y'_1, y'_2, \ldots, y'_k) = L(y_1, \ldots, y_k) = L(x_1, \ldots, x_k)$.
By the first property, $0 = \langle y'_{k+1}, z \rangle = \langle z, z \rangle + 0$. Hence $z = 0$. We are done. $\square$

3 More on Gram-Schmidt

Suppose in the above procedure, $y_{i+1} = 0$ for some i. Then $x_{i+1} \in L(x_1, \ldots, x_i)$ and therefore $x_1, \ldots, x_{i+1}$ are linearly dependent. As a consequence, if $x_1, \ldots, x_n$ are linearly independent, then none of the y_i are 0 and since they are mutually orthogonal, they are linearly independent too. Thus, every finite-dimensional inner product space has an orthogonal basis. By dividing each element by its norm, we can convert an orthogonal basis to an orthonormal basis.

4 Orthogonal complement

Let $S \subseteq V$ be a subset of an inner product space. An element $v \in V$ is said to be orthogonal to S if it is so to every element of S. The set of all v orthogonal to S is

denoted as $S^\perp$.
$S^\perp$ is *always* a subspace regardless of whether S is or not (HW). When S is a subspace, $S^\perp$ is called the *orthogonal complement* of S.
Examples :

- The perpendicular subspace to the set $\{(1,1),(1,2)\}$ in $\mathbb{R}^2$ with the usual inner product is $\{(0,0)\}$. Indeed, $(a,b).(1,1) = a+b = 0$ and $(a,b).(1,2) = a+2b = 0$ imply that $a = b = 0$.
- Given a line $t(1,2,3)$ in $\mathbb{R}^3$, its orthogonal complement is a plane: $0 = (x,y,z).(1,2,3) = x+2y+3z$.
- The continuous functions orthogonal to 1 with the integration inner product on $[0,1]$ are the ones with zero average.

5 Orthogonal decomposition

Recall how we used $(1,1)$ and $(1,2)$ to create orthogonal vectors. Motivated by this construction, we have a theorem:
Let $(V, \langle,\rangle)$ be an inner product space and $S \subseteq V$ be a f.d. subspace. Then every element $x \in V$ can be represented uniquely as a sum $x = s + s^\perp$ where $s \in S$ and $s^\perp \in S^\perp$. Moreover, $\|x\|^2 = \|s\|^2 + \|s^\perp\|^2$.
Caveat: If S is not f.d., the above result is NOT true in general!
Proof: Let $e_1, \ldots, e_n$ be an orthonormal basis of S. Define $s = \sum_i \langle x, e_i \rangle e_i$. Clearly, $s \in S$. Let $s_1 \in S$ be an arbitrary element. Then $s_1 = \sum_j c_j e_j$. $\langle x - s, s_1 \rangle = \sum_j \bar{c_j} \langle x, e_j \rangle - \sum_{j,k} \bar{c_j} \langle x, e_i \rangle \langle e_i, e_j \rangle$. By orthonormality, the latter is $\sum_j \bar{c_j} \langle x, e_j \rangle - \sum_j \bar{c_j} \langle x, e_j \rangle = 0$. Hence, $s^\perp = x - s \in S^\perp$. If $x = t + t^\perp = s + s^\perp$, then $t - s = s^\perp - t^\perp \in S \cap S^\perp = \{0\}$. $\|x\|^2 = \|s\|^2 + \|s^\perp\|^2 + \langle s, s^\perp \rangle + \langle s^\perp, s \rangle$ but $\langle s, s^\perp \rangle = 0$. $\square$The element $s = \sum_i \langle x, e_i \rangle e_i$ is called the *orthogonal projection* of x on the (f.d.) subspace S. It is basically the "shadow" of x on S.

6 The approximation problem

Consider the following questions:

- What is the best way to approximate continuous functions using sines and cosines ?
- What is the best way to approximate continuous functions using polynomials ?
- If we plot the price of houses vs their area (in a particular locality) what is the "best" estimate of price per square foot ?