Math 113: Linear Algebra
Perpendicular Spaces

Ilya Sherman

November 10, 2008

1 Recap

Last time, we discussed orthonormal lists and bases. In particular, we showed that if \( u_1, \ldots, u_k \) is orthonormal in \( V \) and \( v \in V \), then \( v_0 = \sum \langle v, u_i \rangle u_i \) is the closest vector in \( \text{span}(u_i) \) to \( v_i \), i.e. it minimizes \( \|v - v_0\| \). We also showed that any maximal orthonormal list is a basis, and hence that orthonormal bases exist.

2 Perpendicular Spaces

Let \( V \) be a finite dimensional vector space over \( F \), which is \( \mathbb{R} \) or \( \mathbb{C} \), with an inner product. If \( U \subseteq V \), we define \( U^\perp = \{ v \in V \mid \langle v, u \rangle = 0 \text{ for all } u \in U \} \).

**Proposition 1.** \( V \) is always the internal direct sum of \( U \) and \( U^\perp \) (therefore, \( \dim V = \dim U + \dim U^\perp \)). For any \( v \in V \), if we write \( v = u + w \) with \( u \in U \) and \( w \in U^\perp \) (since we are working with a direct sum, this can be done uniquely), then \( u \) is the closest vector on \( U \) to \( v \).

**Proof.** Pick an orthonormal basis \( u_1, \ldots, u_m \) for \( U \). Extend this to an orthonormal basis for \( V \): \( u_1, \ldots, u_m, v_1, \ldots, v_r \) (we can do this since a maximal orthonormal list is a basis).

It remains to show that \( U^\perp = \text{span}(v_1, \ldots, v_r) \); this will imply \( V = U \oplus U^\perp \). Indeed, given any \( v \in V \), we can write \( v = \sum \alpha_i u_i + \sum \beta_i v_i \text{ for } \alpha_i, \beta_i \in F \). If \( v \in U^\perp \), then \( \langle v, u_i \rangle = 0 \), and hence \( \alpha_i = 0 \text{ for all } i \). Hence, \( v \in \text{span}(v_1, \ldots, v_r) \), and so indeed \( V = U \oplus U^\perp \).

Given any \( v \in V \), we already saw that

\[
v_0 = \sum_{i=1}^{m} \langle v, u_i \rangle u_i
\]

has the properties
1. \( v - v_0 \) is perpendicular to \( U \)

2. \( v_0 \) is the closest vector on \( U \) to \( v \).

By the first property, we can write

\[
v = v_0 + v - v_0. \quad \text{\( \in U \)}
\]

By the second property, this proves the other part of the proposition.

Exercise: Use proposition 1 to show that \( (U^\perp)^\perp = U \).

### 2.1 Linear regressions

Suppose we are given some \((x_i, y_i) \in \mathbb{R}^2\) for \(1 \leq i \leq N\) and we want to fit a line \(y = ax + b\). Let

\[
X = (x_1, \ldots, x_N) \in \mathbb{R}^N \\
Y = (y_1, \ldots, y_N) \in \mathbb{R}^N \\
C = (1, \ldots, 1) \in \mathbb{R}^N
\]

We want to find numbers \(a, b \in \mathbb{R}\) so that for all \(i\), \(y_i\) is close to \(ax_i + b\), i.e. \(Y\) is “close to” \(aX + bC\).

One way to do this, if we give \(\mathbb{R}^n\) the “dot product” inner product, we can find the vector on \(\text{span}(X, C)\) closest to \(Y\) by using the previous formulas: find an orthonormal basis \(E_1, E_2\) for \(\text{span}(X, C)\); then, \(\langle Y, E_1 \rangle E_1 + \langle Y, E_2 \rangle E_2\) gives the best linear approximation to \(Y\). Here, the best linear approximation is the one minimizing \(\sum_{i=1}^{N} |y_i - ax_i - b|^2\).

Exercise: Show this gives the usual formula for the least-squares regression.

### 2.2 Constructing orthonormal bases (Gram-Schmidt process)

Given a basis \(e_1, \ldots, e_n\) for \(V\), the Gram-Schmidt process will yield an orthonormal basis \((f_1, \ldots, f_n)\) so that \(\text{span}(e_1) = \text{span}(f_1)\), \(\text{span}(e_1, e_2) = \text{span}(f_1, f_2)\), and so on.

**Construction:** Pick

\[
f_1 = \frac{e_1}{\|e_1\|} \quad \text{\( e_1 \) scaled to length 1}
\]

\[
f_2' = e_2 - \langle e_2, f_1 \rangle f_1 \quad \text{perpendicular to } f_1
\]

\[
f_2 = \frac{f_2'}{\|f_2'\|} \quad \text{scaled to length 1}
\]

\[
f_3' = e_3 - \langle e_3, f_1 \rangle f_1 - \langle e_3, f_2 \rangle f_2 \quad \text{perpendicular to } f_1, f_2
\]

\[
f_3 = \frac{f_3'}{\|f_3'\|} \quad \text{scaled to length 1}
\]

\vdots
Exercises:

1. Prove that the process works

2. Is \( f_1, \ldots, f_n \) the unique orthonormal basis with the property that \( \text{span}(f_1, \ldots, f_k) = \text{span}(e_1, \ldots, e_k) \)? (The answer is no—why not?)

2.2.1 Example (similar to Axler 6.40)

How can we approximate \( \sin(x) \) by a polynomial on the range \( 0 \leq x \leq 1 \)? We could try the Taylor series, \( x - \frac{x^3}{6} + \cdots \). But, linear algebra will give us a better answer. In particular, we’ll find the quadratic polynomial \( Q \in \text{span}(1, x, x^2) \) which minimizes \( \int_0^1 |\sin x - Q(x)|^2 \, dx \). We’ll work in the space \( V \) of continuous functions \( [0, 1] \to \mathbb{R} \) with inner product \( \langle f, g \rangle = \int_0^1 f(x)g(x) \, dx \). We’ll find an orthonormal basis for \( \text{span}(1, x, x^2) \), and then project \( \sin x \) onto it.

Let \( e_1 = 1 \), \( e_2 = x \), \( e_3 = x^2 \). By the Gram-Schmidt process, we have

\[
\begin{align*}
f_1 &= \frac{e_1}{\|e_1\|} \\
&= 1 \\
f'_2 &= e_2 - \langle e_2, f_1 \rangle f_1 \\
&= e_2 - \left( \int_0^1 x \, dx \right) f_1 \\
&= x - \frac{1}{2} \\
f_2 &= \frac{f'_2}{\|f'_2\|} \\
&= \sqrt{12} \left( x - \frac{1}{2} \right) \\
f'_3 &= e_3 - \langle e_3, f_1 \rangle f_1 - \langle e_3, f_2 \rangle f_2 \\
&= x^2 - x + \frac{1}{6} \\
f_3 &= \frac{f'_3}{\|f'_3\|} \\
&= \sqrt{180} \left( x^2 - x + \frac{1}{6} \right)
\end{align*}
\]

So, our approximation \( Q \) is given by

\[
\langle \sin(x), f_1 \rangle f_1 + \langle \sin(x), f_2 \rangle f_2 + \langle \sin(x), f_3 \rangle f_3 \approx -0.0074 + 1.0913x - 0.2355x^2
\]

Note that this is not the start of the Taylor series; and indeed, the error \( |\sin(x) - Q(x)| \) is maximal when \( x = 0 \).
Note that there are many orthonormal bases for span$(1, x, x^2)$, but $Q(x)$ comes out to the same thing for each.

We could have chosen a different inner product, like $\langle f, g \rangle = \int f(x)g(x)(1 + x^4)$, i.e. $f(x) = \int f(x)^2(1 + x^4)$. This will give more weight to $x = 1$, and hence optimize the approximation in this direction.