1 Recap

- Proof of Gram-Schmidt.
- Orthogonal complements.

2 The approximation problem

Consider the following questions (mentioned the last time):

- 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 ?

These questions fall under the purview of the approximation problem: Let V be an inner product space and $S \subseteq V$ be a f.d. subspace. Given an element $x \in V$, determine an element $s \in S$ whose distance from x is as small as possible.
Let $S \subseteq V$ be a f.d. subspace of an inner product space $(V, \langle , \rangle)$ and let $x \in V$. Then if s is the projection of x on S, $\|x - s\| \leq \|x - t\|$ for any $t \in S$. Equality holds if and only if $t = s$.
Proof: Note that $x = s + s^{\perp}$ and hence $x - t = (s - t) + s^{\perp}$. So $\|x - t\|^2 = \|s - t\|^2 + \|x - s\|^2 \geq \|x - s\|^2$ with equality holding if and only if $s = t$. □

Examples

- Let $V = C[0, 2\pi]$ and S be the space spanned by $\phi_0 = \frac{1}{\sqrt{2\pi}}, \phi_1 = \frac{\cos(x)}{\sqrt{\pi}}, \phi_2 = \frac{\sin(x)}{\sqrt{\pi}}, \ldots, \phi_{2n}$.
- The best approximation of $f \in V$ by S is given by the projection $f_n = \sum_k \langle f, \phi_k \rangle \phi_k$ where $\langle f, \phi_k \rangle = \int_0^{2\pi} f \phi_k dx$. These numbers are called the Fourier coefficients of f.
- Let $V = C[-1, 1]$ and S be the space spanned by $1, x, \ldots, x^n$. The normalised Legendre polynomials $\psi_0 = \frac{1}{\sqrt{2}}, \psi_1 = \frac{\sqrt{3}}{\sqrt{2}} x, \psi_2 = \frac{\sqrt{5}}{2\sqrt{2}}(3x^2 - 1), \ldots, \psi_n$ form an orthonormal basis for S.
- The best polynomial approximation of $f \in V$ by S is given by $\tilde{f}_n = \sum_k \langle f, \psi_k \rangle \psi_k$. For instance, if $f(x) = \sin(\pi x)$, then $\langle f, \psi_0 \rangle = 0$, $\langle f, \psi_1 \rangle = \frac{2\sqrt{3}}{\pi\sqrt{2}}$, $\langle f, \psi_2 \rangle = 0$.

3 Least squares fit

Suppose a dependent variable (like the price of a house) y depends linearly on an independent variable (like the area) x as $y = mx + c$. Unfortunately, in real life, if one collects data points $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$, they will not all lie on a line! We want to find those m, c such that the corresponding line is the "best fit", i.e., $\sum_i (y_i - mx_i - c)^2$ is the smallest possible. Note that if we consider the subspace in $\mathbb{R}^n$ spanned by the vectors $(x_1, \ldots, x_n)$ and $(1, 1, \ldots, 1)$, we essentially want the vector s lying in this space that is the best approximation of the vector $(y_1, y_2, \ldots, y_n)$. We can calculate m, c using the formulae directly, or using the equations $(Y - X\beta)^T X = 0$ where Y is the column vector of y_i, X is the $n \times 2$ matrix whose first column is x and the second column is $(1, 1, \ldots)$, and $\beta = \begin{bmatrix} m \\ 1 \end{bmatrix}$.

It is of course possible for the equations to have infinitely many solutions! (This phenomenon is called overfitting.) More generally, if $y = m_1 x_1 + m_2 x_2 + \ldots + m_k x_k + c$, then the "data matrix" X is an $n \times (k+1)$ matrix, and the "slopes vector" β is a $(k+1)$-vector. Even then, the principle is to project y onto the "column space", i.e., the subspace of $\mathbb{R}^n$ generated by the columns of X. Alternatively, $(Y - X\beta)^T X = 0$. These equations are called *normal equations*. This procedure is called *linear regression*. By the way, if you want to fit polynomials, you can do exactly the same thing by the trick of introducing new variables ! $(x_1 = x, x_2 = x^2, \ldots)$.

4 Inverses

Recall that our aim was to solve linear equations. In other words, we wanted to solve an *inverse* problem. We shall study some generalities about inverse functions, and then specialise to inverses of linear maps.

Given two *sets* V, W and a *onto function* $T : V \to W$ A *left inverse* $L : W \to V$ is one that satisfies $L(T(x)) = x$, i.e., $LT = I_V$. A *right inverse* $R : W \to V$ satisfies $T(R(x)) = x$, i.e., $TR = I_W$. These are not mindless definitions. For instance, consider $V = \{1, 2\}$ and $W = \{0\}$. Define $T : V \to W$ as $T(1) = T(2) = 0$. So define $R_1, R_2 : W \to V$ as $R_1(0) = 1$ and $R_2(0) = 2$. These are right inverses. However, if $L(T(1)) = 1$, then $L(0) = 1$. That means $L(T(2)) = 1$. Hence there is *no* left inverse in this example.

Every onto function $f : V \to W$ has at least one right inverse. Indeed, for every $w \in W$ there is some $v \in V$ so that $f(v) = w$. Define $R(w) = v$. Clearly $f(R(w)) = f(v) = w$. In general, right inverses are not unique.

Here is an interesting result about left inverses: An onto function $T : V \to W$ can have *at most* one left inverse. If L is a left inverse of T then it is *also* a right inverse!

Proof: Suppose $L_1, L_2 : W \to V$ are left inverses. If $w = T(v)$, then $L_1(w) = v = L_2(w)$. Hence $L_1 = L_2$ (the onto assumption plays a role). $T(L(w)) = T(L(T(v))) = T(v) = w$. Hence L is also a right inverse. $\square$

Moreover, if a left inverse exists, the right inverse is THE left inverse, i.e., the right inverse is unique. (Indeed, $TR_1 = TR_2 = I$ and hence $LTR_1 = LTR_2 \Rightarrow R_1 = R_2$.)

An onto function $T : V \to W$ has a left inverse if and only if it is $1 - 1$ (HW).

A one-onto onto function has a unique left inverse (which we know is also a right inverse). Such a T is called invertible and its (left or right) inverse is denoted as T^{-1}.

5 Inverses of linear maps

Let V, W be vector spaces over the same field. Let $T : V \to W$ be an *onto* linear map. Then TFAE.

- T is $1-1$.
- T is invertible and the inverse T^{-1} is *linear*.
- $\forall x \in V$, $T(x) = 0$ if and only if $x = 0$.

item Proof: We prove that $a \Rightarrow b \Rightarrow c \Rightarrow a$.

- $a \Rightarrow b$: We already know that T^{-1} exists. Let $T^{-1}(av+bw) = c$. So $T(c) = av+bw$ which is $T(c) = aT(T^{-1}v) + bT(T^{-1}w) = T(aT^{-1}v + bT^{-1}w)$. Since T is $1-1$, $c = aT^{-1}v + bT^{-1}w$.
- $b \Rightarrow c$: If $T(x) = 0$, then $x = T^{-1}T(x) = 0$.
- $c \Rightarrow a$: If $T(v) = T(w)$, $T(v-w) = 0$ and hence $v = w$.

Let V be f.d with $\dim(V) = n$ and $T : V \to W$ is an onto linear map. Then TFAE (HW):

- T is $1-1$.
- If $e_1, \ldots, e_p$ are linearly independent in V, then $T(e_1), \ldots, T(e_p)$ are so in W.
- $\dim(W) = n$.
- If $e_1, \ldots, e_n$ is a basis for V, then $T(e_1), \ldots, T(e_n)$ is so for W.