Review. matrix approximation and compression

Function spaces

Recall the following:

- We call objects **vectors** if they can be added and scaled (subject to the usual laws).
- A set of vectors is a **vector space** if it is closed under addition and scaling.

  In other words, vector spaces are spans.

We will now discuss spaces of vectors, where the vectors are functions.

Why? Just one example why it is super useful to apply our linear algebra machinery to functions: we discussed the **distance** between vectors and how to find vectors closest to interesting subspaces (i.e. orthogonal projections). These notions are important for functions, too. For instance, given a (complicated) function, we want to find the closest function in a subspace of (simple) functions. In other words, we want to approximate functions using other (typically, simpler) functions.

Comment. Functions $f(x)$ and $g(x)$ can also be multiplied. This is an extra structure (it makes appropriate sets of functions an **algebra**, which is something more special than a **vector space**), which we ignore during our discussion of vector spaces.

An inner product on function spaces

On the space of, say, (piecewise) continuous functions $f\colon [a,b] \to \mathbb{R}$, it is natural to consider the dot product

$$\langle f, g \rangle = \int_a^b f(t) g(t) \mathrm{d}t.$$

Why? A (sensible) dot product provides a (sensible) notion of distance between functions. The dot product above is the continuous analog of the usual dot product $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{t=1}^n x_t y_t$ for vectors in $\mathbb{R}^n$. Do you see it?!

As a consequence, once we have the dot product, we can orthogonally project functions onto spaces of simple functions. In other words, we can compute best approximations of functions by simple functions (for instance, best quadratic approximations).

Why continuous? We need that any product $f(x)g(x)$ is integrable. That means we cannot work with all functions. Continuity is certainly sufficient. In fact, the right condition is that $f(x)^2$ should be integrable on $[a,b]$ (i.e. $f(x)$ is square-integrable). Such a function is said to be in $\mathcal{L}^2[a,b]$.

Example 183. What is the orthogonal projection of $f\colon [a,b] \to \mathbb{R}$ onto the space of constant functions (that is, $\operatorname{span}\{1\}$)?

Solution. The orthogonal projection of $f\colon [a,b] \to \mathbb{R}$ onto $\operatorname{span}\{1\}$ is

$$\frac{\langle f, 1 \rangle}{\langle 1, 1 \rangle} 1 = \frac{\int_a^b f(t) 1 \mathrm{d}t}{\int_a^b 1^2 \mathrm{d}t} = \frac{1}{b-a} \int_a^b f(t) \mathrm{d}t.$$

This is the average of $f(x)$ on $[a,b]$.

Comment. Makes perfect sense, doesn't it? Intuitively, the best approximation of a function by a constant should indeed be the one where the constant is the average.

Armin Straub
straub@southalabama.edu

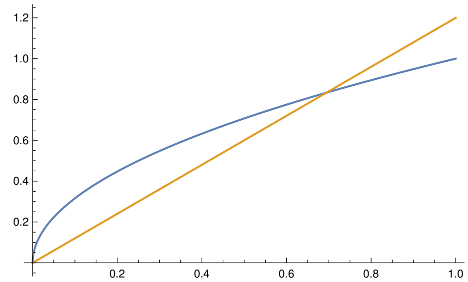
Example 184. Find the best approximation (in the L^2 sense) of $f(x) = \sqrt{x}$ on the interval $[0, 1]$ using a function of the form $y = ax$.

Solution. The orthogonal projection of $f: [0, 1] \to \mathbb{R}$ onto $\operatorname{span}\{x\}$ is

$$\frac{\langle f, x\rangle}{\langle x, x\rangle} x = \frac{\int_0^1 f(t) t \mathrm{d}t}{\int_0^1 t^2 \mathrm{d}t} x = 3x \int_0^1 t f(t) \mathrm{d}t.$$

In our case, the best approximation is

$$3x \int_0^1 t\sqrt{t} \mathrm{d}t = 3x \int_0^1 t^{3/2} \mathrm{d}t = 3x \left[\frac{1}{5/2} t^{5/2}\right]_0^1 = \frac{6}{5} x.$$



What does "in the L^2 sense" mean? There are various ways to make precise what "best approximation" should mean. Here, we mean that the norm of the error gets minimized: since the error is $\operatorname{error}(t) = \sqrt{t} - at$, this means that

$$\|\operatorname{error}\|^2 = \langle \operatorname{error}, \operatorname{error}\rangle = \int_0^1 \operatorname{error}(t)^2 \mathrm{d}t = \int_0^1 (\sqrt{t} - at)^2 \mathrm{d}t$$

is minimized over all choices for a (note that it does not make a difference whether we are minimizing the norm or the square of the norm). Since the above norm is known as the L^2 norm, we are finding the best approximation "in the L^2 sense". For instance, an alternative meaningful norm is the L^∞ norm, for which $\|\operatorname{error}\|$ simply measures the maximum absolute value of the error (this may sound simpler but there is no corresponding notion of orthogonality so that we cannot apply tools like orthogonal projections).

For comparison. Let's minimize $r(a) := \|\operatorname{error}\|^2$ directly! First, we compute that

$$r(a) = \int_0^1 (\sqrt{t} - at)^2 \mathrm{d}t = \int_0^1 (t - 2at^{3/2} + a^2 t^2) \mathrm{d}t = \left[\frac{1}{2} t^2 - \frac{4}{5} a t^{5/2} + \frac{1}{3} a^2 t^3\right]_0^1 = \frac{1}{2} - \frac{4}{5} a + \frac{1}{3} a^2.$$

We now minimize $r(a)$ over all choices of a using what we learned in Calculus I: if the minimum occurs at a, then we necessarily have $r'(a) = 0$. We compute that $r'(a) = -\frac{4}{5} + \frac{2}{3} a$. Solving $-\frac{4}{5} + \frac{2}{3} a = 0$, we find $a = \frac{6}{5}$ as the only candidate. Which matches exactly what we have found above!

Clearly, this approach becomes more challenging if our approximations have more than one degree of freedom. However, our linear algebra approach continues to work fine, as we will see in the next example.

Armin Straub
straub@southalabama.edu