

# ECE 586: Vector Space Methods

## Lecture 17: Frequently Asked Questions

Henry D. Pfister  
Duke University

October 31, 2022

Here we give a list of questions, and their answers, that were submitted by students after watching the flip video or reviewing the lecture slides.

### 1 Best Approximation

**How is best approximation in Banach spaces related with orthogonal projection in Hilbert spaces?** In a Hilbert space, The orthogonal projection onto a closed subspace is equivalent to the best approximation in the induced Banach space defined by the induced norm. See Theorem on slide 3, which establishes a link between the best approximation (defined as the vector  $w \in W$  closest to  $v \in V$ , meaning  $w$  achieves the smallest  $\|v - w\|$ ) and orthogonal projection in a Hilbert space. It provides a way to compute the best approximation of  $v$  if given an orthogonal basis  $w_1, \dots, w_n$  of  $W$ , based on the inner products between  $v$  and the basis vectors (you are guaranteed to have an inner product, by definition of a Hilbert space).

**Is a best approximation always unique, or only in Hilbert spaces?** Uniqueness of the best approximation holds for Hilbert spaces, but not for general Banach spaces. In finite dimensions, there is a weaker condition you can impose, called strict convexity, on the Banach space to get uniqueness of the best approximation. Hilbert spaces satisfy this condition. For more on this, see Theorem 4.6 in [this link](#).

**On slide 3, the orthogonal basis is countable. Shouldn't the superscript of sum be some number  $M > 0$ ?** Good question. Indeed, this assumes the basis is countably infinite when it could also be finite. Thus, it would be more accurate to replace  $\infty$  by  $\dim(W)$ .

**On Slide 3, can different  $v_1, v_2 \in V$  have the same best approximation  $w$ ?** Yes, here's a simple 2D example – think of projecting the vectors (1,1) and (1,3) onto the x-axis. In both cases, you get the vector (1,0).

**In Definition 2 on Slide 6, does a projection imply  $Tv = v$  and  $T^2 = T$ ?** For a projection,  $Tv = v$  holds if  $v$  is already in the range of  $T$  but not in general. Also,  $T^2 \triangleq T \circ T = T$  for any linear transform that is a projection (i.e., idempotent).

**To check that a matrix  $P$  is an orthogonal projection, can we take the range of  $P$  and the nullspace of  $P$  and verify orthogonality between the sets of basis vectors?** Yes, but an easier way to do this would be to check that  $P$  is both Hermitian and idempotent (see Theorem on slide 8).

**I can understand the definition of best approximation, but it's difficult to grasp intuitively. Given a vector, you project it onto an orthogonal set and then add the projections together. Then, you get the best approximation of this vector. It's difficult to picture this and it seems like magic.** Intuitively, if you want to project a vector  $x$  onto an orthogonal subspace (with orthogonal basis  $w_1, \dots, w_n$ ), you do it by computing “similarities” (i.e., inner products) between your vector  $x$  and the basis vectors  $w_i$  (i.e.,  $\langle w_i, x \rangle$ ). You then weigh the basis vectors  $w_1, \dots, w_n$  by their “similarities” to  $x$ . e.g., if  $x$  happens to be most “similar” to  $w_2$ , it makes sense to assign a higher weight (quantified by  $\langle w_2, x \rangle$ ) to  $w_2$ .

**What are the most important applications of vector approximation?** Principal Component Analysis (a machine-learning technique often used for dimension reduction), Linear Regression, and function approximation.

**Is the definition of a projection in Slides 6 and 7 a more general definition which encompasses the projection we have been working with?** Yes, the term *projection* can be defined in vector spaces without a norm or inner product. First, we give this definition (Slides 6-7) and then specialize it to the case of *orthogonal projection* (Slide 8) where an inner product is required for the definition.

**In the previous lecture we learned about the second-order bound for functions with a Lipschitz gradient. Is there a first-order bound for it?** When approximating a function, people often refer to the number of terms in the Taylor series as the order of the approximation. For example, consider various approximations of  $\ln(1+x)$ :

- Zeroth-order approximation:  $\ln(1+x) = 0 + O(x)$
- First-order approximation:  $\ln(1+x) = 0 + x + O(x^2)$
- Second-order approximation:  $\ln(1+x) = 0 + x - \frac{1}{2}x^2 + O(x^3)$

When the error in this approximation is precisely bounded (rather than left as  $O(x^k)$ ), people sometimes refer to this as a  $k$ -th order bound). For  $f: [0, a] \rightarrow \mathbb{R}$  defined by  $x \mapsto \ln(1+x)$ , we have

- First-order bound:  $|\ln(1+x) - 0| \leq Lx$  where  $L = \sup_{x \in [0, a]} f'(x)$  is the Lipschitz constant of  $f$  on  $[0, a]$
- Second-order bound:  $|\ln(1+x) - 0 - x| \leq \frac{1}{2}L'x^2$  where  $L' = \sup_{x \in [0, a]} f''(x)$  is the Lipschitz constant of  $f'$  on  $[0, a]$