

## Orthogonal bases

The text emphasizes *orthonormal lists*, in keeping with tradition. Remember that a list  $(e_1, \dots, e_m)$  is called *orthonormal* if

$$\begin{aligned}\langle e_i, e_j \rangle &= 0, & (1 \leq i \neq j \leq m) \\ \langle e_i, e_i \rangle &= 1, & (1 \leq i \leq m).\end{aligned}\tag{ONL}$$

These lists are indeed convenient, and lead to lots of simple formulas. In most applications, the first condition (pairwise orthogonality) has a fairly clear and simple meaning. In a few applications (like quantum mechanics) the second condition (length one) also has a clear and natural meaning: the length of a vector can sometimes be related to a probability, and “probability one” is a very natural condition.

But it is more typical that the length in an application refers to some chosen measurement units, so that “length one” changes when the units are changed. In these cases the length one condition is just an organizational convenience.

There are also costs associated to working with orthonormal lists. A central strength of linear algebra is that almost all the computations required are arithmetic: addition, subtraction, multiplication, and division. In the world of computers, a consequence is that when the problems involve rational numbers, all computations can be performed *exactly*, without roundoff errors. The Gram-Schmidt process as described in the text involves extraction of  $m$  square roots. (Can you see where? There are no square root signs on page 183 of the text.) This is computationally expensive. Computer algebra systems can manipulate square roots formally without error, but this is a (sometimes unnecessary) layer of complexity.

The purpose of these notes is therefore to rewrite pages 180– of the text in terms of *orthogonal* lists rather than orthonormal ones. A list  $(f_1, \dots, f_m)$  is called *orthogonal* if

$$\langle f_i, f_j \rangle = 0, \quad (1 \leq i \neq j \leq m).\tag{OL}$$

The rest of these notes is a collection of substitutes for things in the text, using orthogonal lists rather than orthonormal ones. The proofs are almost identical to those in the text, and I’ll in general omit them.

**6.25’ Proposition.** *If  $(f_1, \dots, f_m)$  is an orthogonal list of vectors in  $V$ , then*

$$\|b_1 f_1 + \dots + b_m f_m\|^2 = |b_1|^2 \|f_1\|^2 + \dots + |b_m|^2 \|f_m\|^2$$

for all  $b_1, \dots, b_m \in F$ .

**6.26’ Corollary.** *The list of nonzero vectors in an orthogonal list is linearly independent.*

An *orthogonal basis* of  $V$  is an orthogonal list of vectors in  $V$  that is also a basis of  $V$ . Every orthogonal list of nonzero vectors in  $V$  with length  $\dim V$  is automatically an orthogonal basis of  $V$ . To illustrate this principle, consider the following list of three vectors in  $\mathbb{R}^3$ :

$$((1, 1, 1), (2, -1, -1), (0, 1, -1))$$

The verification that this list is orthogonal is easy (do it!); because we have an orthogonal list of length three in a three-dimensional vector space, it must be an orthogonal basis.

**6.30’ Theorem.** *Suppose  $(f_1, \dots, f_n)$  is an orthogonal basis of  $V$ . Then*

$$v = \frac{\langle v, f_1 \rangle}{\langle f_1, f_1 \rangle} f_1 + \dots + \frac{\langle v, f_n \rangle}{\langle f_n, f_n \rangle} f_n$$

and

$$\|v\|^2 = \frac{|\langle v, f_1 \rangle|^2}{\langle f_1, f_1 \rangle} + \dots + \frac{|\langle v, f_n \rangle|^2}{\langle f_n, f_n \rangle}$$

for every  $v \in V$ .

6.31' **Gram-Schmidt.** If  $(v_1, \dots, v_m)$  is a list of vectors in  $V$ , then there is an orthogonal list  $(f_1, \dots, f_m)$  of vectors in  $V$  such that

$$6.21' \quad \text{span}(v_1, \dots, v_j) = \text{span}(f_1, \dots, f_j)$$

for  $j = 1, \dots, m$ .

The proof and the algorithm changes only in the inductive step, constructing  $f_j$  after  $(f_1, \dots, f_{j-1})$  are constructed. The new definition is

$$6.23' \quad f_j = v_j - \sum_{1 \leq i \leq j-1, f_i \neq 0} \frac{\langle v_j, f_i \rangle}{\langle f_i, f_i \rangle} f_i.$$

What has changed is that we *do not* divide by the length to get a vector of length one.

The new algorithm is always replacing  $v_j$  by  $v_j$  plus a linear combination of *strictly earlier*  $v_i$ . So we end up with formulas

$$f_j = v_j + \sum_{i < j} a_{ij} v_i. \quad (LT)$$

In the language of Gaussian elimination, thinking of the  $v_j$  as the rows of a matrix, the operations we are performing are the ones called  $L(j, i; \mu)$ , corresponding to lower triangular matrices with ones on the diagonal. The fact that

$$\text{span}(v_1, \dots, v_j) = \text{span}(f_1, \dots, f_j)$$

follows immediately from (LT); we have used things like this often, for example in the proof of Proposition 5.1 in the notes on Gaussian elimination.

The rest of the section about Gram-Schmidt requires little change. In order for orthogonal bases to be useful, we should also generalize a little some formulas later in the chapter. On page 194, proof of Theorem 6.47, we can say instead

Let  $(f_1, \dots, f_m)$  be an orthogonal basis of  $U$ . Obviously

$$v = \underbrace{\frac{\langle v, f_1 \rangle}{\langle f_1, f_1 \rangle} f_1 + \dots + \frac{\langle v, f_m \rangle}{\langle f_m, f_m \rangle} f_m}_u + v - \underbrace{\frac{\langle v, f_1 \rangle}{\langle f_1, f_1 \rangle} f_1 - \dots - \frac{\langle v, f_m \rangle}{\langle f_m, f_m \rangle} f_m}_w.$$

Clearly  $u \in U$ . Because  $(f_1, \dots, f_m)$  is an orthogonal list, for each  $j$  we have

$$\langle w, f_j \rangle = \langle v, f_j \rangle - \frac{\langle v, f_j \rangle}{\langle f_j, f_j \rangle} \langle f_j, f_j \rangle = 0.$$

Thus  $w$  is orthogonal to every vector in  $\text{span}(f_1, \dots, f_m)$ .

on page 196: if  $(f_1, \dots, f_m)$  is an orthogonal basis of  $U$ , then the formula for  $P_U$  becomes

$$6.35' \quad P_U(v) = \frac{\langle v, f_1 \rangle}{\langle f_1, f_1 \rangle} f_1 + \dots + \frac{\langle v, f_m \rangle}{\langle f_m, f_m \rangle} f_m$$

for every  $v \in V$ .