Math S-21b – Lecture #7-8 Notes

We now take up in greater detail the ideas of **inner products** and **orthogonality** beyond the more basic constructions introduced earlier in the course. It should be noted that most of what we did in $\mathbf{R}^n$ and everything we've done so far with general vector spaces was based only on the ability to add and scale elements. The structure necessary to measure lengths and angles is a very useful additional structure. We'll look at the basic properties of inner products, derive some algebraic facts, and define and focus on **orthonormal bases** and their advantages. Among these advantages are the simplicity of determining coordinates and producing the matrix for **orthogonal projection**. We'll also define the **orthogonal complement of a subspace** and give a very simple method for finding it.

Inner products and inner product spaces

Definition: An **inner product** in a linear space V is a rule that assigns a scalar (denoted by $\langle f, g \rangle$) to any pair f, g of elements of V, such that the following properties hold for all $f, g, h \in V$, and all scalars c:

a. $\langle f, g \rangle = \langle g, f \rangle$ (symmetry)

b. $\langle f, g+h \rangle = \langle f, g \rangle + \langle f, h \rangle$ and
$\langle f+g, h \rangle = \langle f, h \rangle + \langle g, h \rangle$
(left and right distributive laws)

c. $\langle cf, g \rangle = c\langle f, g \rangle$

d. $\langle f, f \rangle = \| f \|^2 \geq 0$ for all $f \in V$,
and $\langle f, f \rangle = 0$ only if f is the zero element. (positive definiteness)

A linear space endowed with an inner product is called an **inner product space**. Because the last property enables us to define the norm (or magnitude) $\|f\|$ of any element, this is also sometimes referred to as a **normed linear space**.

Examples

I. We will primarily focus on the **dot product** in $\mathbf{R}^n$, i.e. $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$.

This makes $\mathbf{R}^n$ not only a vector space but also an inner product space because the dot product enables us to define the length or norm of any vector as $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle}$). We have previous shown (using the Law of Cosines) that $\boxed{\mathbf{x} \cdot \mathbf{y} = \|\mathbf{x}\| \|\mathbf{y}\| \cos\theta}$ for any two vectors. [This is also consistent with the **Cauchy-Schwarz inequality** which states that $\boxed{|\mathbf{x} \cdot \mathbf{y}| \le \|\mathbf{x}\| \|\mathbf{y}\|}$. (See text for proof.)] It should be noted that this is used to <u>define</u> angles in $\mathbf{R}^n$ and thus enables us to talk about orthogonality of vectors even in spaces that we cannot fully visualize. We can define acute angles and obtuse angles, but perhaps most valuable is the ability to say that <u>two nonzero vector are orthogonal (or perpendicular) if and only if their dot product is zero</u>. The inner product properties are easy to verify:

a. $\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$

b. $\mathbf{x} \cdot (\mathbf{y} + \mathbf{z}) = \mathbf{x} \cdot \mathbf{y} + \mathbf{x} \cdot \mathbf{z}$ and
$(\mathbf{x} + \mathbf{y}) \cdot \mathbf{z} = \mathbf{x} \cdot \mathbf{z} + \mathbf{y} \cdot \mathbf{z}$

c. $(c\mathbf{x}) \cdot \mathbf{y} = c(\mathbf{x} \cdot \mathbf{y}) = \mathbf{x} \cdot (c\mathbf{y})$

d. $\langle \mathbf{x} \cdot \mathbf{x} \rangle = \|\mathbf{x}\|^2 \ge 0$ for all $\mathbf{x} \in \mathbf{R}^n$, and
$\mathbf{x} \cdot \mathbf{x} = \|\mathbf{x}\|^2 = 0$ only if $\mathbf{x} = 0$.

It's an interesting fact that not only can we define length in terms of the dot product, we can also express the dot product in terms of length. Specifically:

$$\begin{aligned}\|\mathbf{x}+\mathbf{y}\|^2 &= (\mathbf{x}+\mathbf{y})\cdot(\mathbf{x}+\mathbf{y}) = \|\mathbf{x}\|^2+\|\mathbf{y}\|^2+2\mathbf{x}\cdot\mathbf{y} \\ \|\mathbf{x}-\mathbf{y}\|^2 &= (\mathbf{x}-\mathbf{y})\cdot(\mathbf{x}-\mathbf{y}) = \|\mathbf{x}\|^2+\|\mathbf{y}\|^2-2\mathbf{x}\cdot\mathbf{y}\end{aligned} \Rightarrow \|\mathbf{x}+\mathbf{y}\|^2-\|\mathbf{x}-\mathbf{y}\|^2 = 4\mathbf{x}\cdot\mathbf{y}$$

$$\Rightarrow \boxed{\mathbf{x}\cdot\mathbf{y} = \tfrac{1}{4}\left(\|\mathbf{x}+\mathbf{y}\|^2-\|\mathbf{x}-\mathbf{y}\|^2\right)}$$

II. In the linear space $\mathbf{R}^{m\times n}$ consisting of all $m\times n$ matrices with real entries, we can define $\langle \mathbf{A},\mathbf{B}\rangle = \mathrm{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{B})$ where the trace is the sum of the diagonal entries of the $n\times n$ matrix $\mathbf{A}^{\mathrm{T}}\mathbf{B}$. It's not difficult to verify the four axioms for an inner product. One interesting aspect of this inner product is that it enables us to define a norm on the space of $m\times n$ matrices, i.e. $\|\mathbf{A}\|^2 = \langle \mathbf{A},\mathbf{A}\rangle = \mathrm{trace}(\mathbf{A}^{\mathrm{T}}\mathbf{A})$. If we express $\mathbf{A} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_n \\ \downarrow & & \downarrow \end{bmatrix}$, a quick calculation shows that $\|\mathbf{A}\|^2 = \|\mathbf{v}_1\|^2+\cdots+\|\mathbf{v}_n\|^2$.

III. Arguably the most useful inner products are those defined for various spaces of functions. If you think of the components of a vector as values of a function, i.e. $\mathbf{x} = \langle x_1, \cdots, x_n \rangle = \langle x(1), \cdots, x(n) \rangle$, then the dot product is just the (finite, discrete) sum of the product of the respective values. If we have a real-valued function defined at all points in some interval $[a,b]$, then we might use integration as the analogous summation and define $\langle f, g \rangle = \int_a^b f(x)g(x)dx$. This integral may not be defined for all functions, so we may have to restrict the class of function for which this inner product is defined. Common choices are to restrict to continuous functions or piecewise-continuous functions. We can easily establish the first three axioms for an inner product space. The fourth axiom follows in the case of continuous functions but requires some additional interpretation in the case of more general functions. An inner product in a linear space of functions enables us to define **orthogonal functions** and the **norm of a function**. We can also think of the "distance between two functions" as $\|f-g\|$ where $\|f-g\|^2 = \int_a^b [f(x)-g(x)]^2 dx$. We can also modify the inner product by scaling by a factor matched to the width of the interval $[a,b]$ and still satisfy all the necessary axioms. For example, when considering piecewise-continuous functions defined on the interval $[-\pi, \pi]$, one good choice is to define $\langle f, g \rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x)g(x)dx$. This choice is the foundation for understanding **Fourier Series** based on the orthogonality of trigonometric functions with respect to this inner product.

Back to $\mathbf{R}^n$:

Definition: A collection of vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ is called **orthonormal** if it consists of mutually orthogonal unit vectors. That is, $\mathbf{u}_i \cdot \mathbf{u}_j = \begin{Bmatrix} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{Bmatrix}$.

Example: The standard basis $\mathcal{E} = \{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n\}$ for $\mathbf{R}^n$ is orthonormal. Any subset of this is also orthonormal.

Proposition: Orthonormal vectors are always linearly independent.

Proof: Let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be orthonormal and suppose that $c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_m\mathbf{u}_m = \mathbf{0}$. Then for any k we have:

$$\mathbf{u}_k \cdot (c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \ldots + c_m\mathbf{u}_m) = \mathbf{u}_k \cdot \mathbf{0} = 0 \quad \Rightarrow \quad c_k(\mathbf{u}_k \cdot \mathbf{u}_k) = c_k = 0 \text{ for all } k.$$

So the vectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ are linearly independent.

Corollary: If $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ are orthonormal in $\mathbf{R}^n$, then they form an **orthonormal basis** for $\mathbf{R}^n$.

An **orthonormal basis** (for either a subspace or all of $\mathbf{R}^n$) is advantageous in several ways. In particular, such a basis makes the calculation of coordinates relative to an orthonormal basis simple, and it provides a simple way to define and calculate the orthogonal projection of a vector onto a subspace. This begins with the definition of the orthogonal complement of a subspace.

Definition: If $V \subseteq \mathbf{R}^n$ is a subspace, its **orthogonal complement** is $V^\perp = \{\mathbf{x} \in \mathbf{R}^n : \mathbf{x} \cdot \mathbf{v} = 0 \text{ for all } \mathbf{v} \in V\}$. This is also a subspace (with complementary dimension). In words, the orthogonal complement of a subspace consists of all vectors that are orthogonal to (every vector in) this subspace.

Finding $V^\perp$: Suppose $V = \text{Span}\{\mathbf{v}_1, \cdots, \mathbf{v}_k\}$. Ideally this should be a basis for *V*, but this is not essential. If we let $\mathbf{A} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{v}_1 & \cdots & \mathbf{v}_k \\ \downarrow & & \downarrow \end{bmatrix}$, then A will be an $n \times k$ matrix with $V = \text{im}(\mathbf{A})$ and $\text{rank}(\mathbf{A}) = \dim(V)$. The transpose of **A** is defined to be the $k \times n$ matrix whose rows are the columns of **A**, i.e. $\mathbf{A}^\text{T} = \begin{bmatrix} \leftarrow & \mathbf{v}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{v}_k & \rightarrow \end{bmatrix}$. Note that:

$$\mathbf{x} \in V^\perp \Leftrightarrow \left\{ \begin{matrix} \mathbf{v}_1 \cdot \mathbf{x} = 0 \\ \vdots \\ \mathbf{v}_k \cdot \mathbf{x} = 0 \end{matrix} \right\} \Leftrightarrow \begin{bmatrix} \leftarrow & \mathbf{v}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{v}_k & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \mathbf{x} \\ \downarrow \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0} \Leftrightarrow \mathbf{A}^\text{T}\mathbf{x} = \mathbf{0} \Leftrightarrow \mathbf{x} \in \ker \mathbf{A}^\text{T}$$

So $\boxed{\mathbf{x} \in V^\perp \Leftrightarrow \mathbf{x} \in \ker(\mathbf{A}^\text{T})}$. This argument also establishes the fact that $\boxed{(\text{im}\,\mathbf{A})^\perp = \ker(\mathbf{A}^\text{T})}$ for any matrix **A**.

Definition: Given a subspace $V \subseteq \mathbf{R}^n$, the **orthogonal projection** of a vector **x** onto this subspace is the unique vector $\text{Proj}_V \mathbf{x}$ such that $\mathbf{x} - \text{Proj}_V \mathbf{x} \in V^\perp$. (Drawing a picture helps here.)

Why should such a vector be unique? How do we know that this is well-defined? If we have an orthonormal basis for the subspace *V*, this follows from the next proposition.

Proposition: Suppose $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_k\}$ is an orthonormal (ON) basis for a subspace $V \subseteq \mathbf{R}^n$. Then for any $\mathbf{x} \in \mathbf{R}^n$, $\text{Proj}_V \mathbf{x} = (\mathbf{x} \cdot \mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x} \cdot \mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{x} \cdot \mathbf{u}_k)\mathbf{u}_k$. That is, the coordinates of the projection are just the scalar projections of **x** in the direction of the respective unit vectors of the ON basis.

Proof: Suppose $\text{Proj}_V \mathbf{x} = c_1\mathbf{u}_1 + c_2\mathbf{u}_2 + \cdots + c_k\mathbf{u}_k$. By definition, $\mathbf{x} - \text{Proj}_V \mathbf{x} = \mathbf{x} - c_1\mathbf{u}_1 - c_2\mathbf{u}_2 - \cdots - c_k\mathbf{u}_k \in V^\perp$. Therefore, for all *i*, $\mathbf{u}_i \cdot (\mathbf{x} - c_1\mathbf{u}_1 - c_2\mathbf{u}_2 - \cdots - c_k\mathbf{u}_k) = \mathbf{u}_i \cdot \mathbf{x} - c_i = 0$. So $c_i = \mathbf{u}_i \cdot \mathbf{x} = \mathbf{x} \cdot \mathbf{u}_i$ for all *i*. This not only gives us the formula as stated in the proposition, the fact that coordinates relative to a basis are unique establishes the uniqueness of the projection.

It is important, however, to note that this demonstration was based on the existence of an ON basis for any subspace. We'll soon see a method for constructing such a basis out of any given basis.

Formula for the matrix of orthogonal projection

It's not obvious, but the formula $\text{Proj}_V\mathbf{x} = (\mathbf{x}\cdot\mathbf{u}_1)\mathbf{u}_1 + (\mathbf{x}\cdot\mathbf{u}_2)\mathbf{u}_2 + \cdots + (\mathbf{x}\cdot\mathbf{u}_k)\mathbf{u}_k$ enables us to find the matrix for orthogonal projection onto any subspace $V \subseteq \mathbf{R}^n$ with ON basis $\mathcal{B} = \{\mathbf{u}_1,\cdots,\mathbf{u}_k\}$. We write:

$$\text{Proj}_V\mathbf{x} = (\mathbf{x}\cdot\mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{x}\cdot\mathbf{u}_k)\mathbf{u}_k = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \mathbf{u}_1\cdot\mathbf{x} \\ \vdots \\ \mathbf{u}_k\cdot\mathbf{x} \end{bmatrix}$$

$$= \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k \\ \downarrow & & \downarrow \end{bmatrix} \begin{bmatrix} \leftarrow & \mathbf{u}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{u}_k & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow \\ \mathbf{x} \\ \downarrow \end{bmatrix} = \mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{x}$$

where $\mathbf{B}$ is the $n\times k$ matrix with ON columns given by the ON basis. So $\mathbf{A} = \mathbf{B}\mathbf{B}^{\mathrm{T}}$ is the matrix for Proj_V.

Corollary: The matrix for **reflection** through the subspace V is given by $2\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{I}$.

Proof: We have already seen (and a picture makes clear) that if $\mathbf{p} = \text{Proj}_V\mathbf{x}$, then $\text{Ref}_V\mathbf{x} = \mathbf{x} + 2(\mathbf{p}-\mathbf{x}) = 2\mathbf{p}-\mathbf{x} = 2\mathbf{B}\mathbf{B}^{\mathrm{T}}\mathbf{x} - \mathbf{x} = (2\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{I})\mathbf{x}$, so $\text{Ref}_V = 2\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{I}$.

Example: Let L be the line in $\mathbf{R}^3$ spanned by the vector $\mathbf{v}=\begin{bmatrix}1\\2\\-2\end{bmatrix}$. If we normalize this, then

$\mathbf{u}=\frac{1}{3}\begin{bmatrix}1\\2\\-2\end{bmatrix}$ provides an ON basis for this line (subspace). We have

$$\text{Proj}_V=\mathbf{B}\mathbf{B}^{\mathrm{T}}=\mathbf{u}\mathbf{u}^{\mathrm{T}}=\frac{1}{9}\begin{bmatrix}1\\2\\-2\end{bmatrix}\begin{bmatrix}1 & 2 & -2\end{bmatrix}=\frac{1}{9}\begin{bmatrix}1 & 2 & -2\\2 & 4 & -4\\-2 & -4 & 4\end{bmatrix} \text{ and}$$

$$\text{Ref}_V=2\mathbf{B}\mathbf{B}^{\mathrm{T}}-\mathbf{I}=\frac{2}{9}\begin{bmatrix}1 & 2 & -2\\2 & 4 & -4\\-2 & -4 & 4\end{bmatrix}-\begin{bmatrix}1 & 0 & 0\\0 & 1 & 0\\0 & 0 & 1\end{bmatrix}=\frac{1}{9}\begin{bmatrix}-7 & 4 & -4\\4 & -1 & -8\\-4 & -8 & -1\end{bmatrix}.$$

We can easily construct an orthonormal basis for the plane $S=L^{\perp}$, namely

$$\mathcal{B}=\left\{\frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\1\end{bmatrix},\frac{1}{\sqrt{18}}\begin{bmatrix}4\\-1\\1\end{bmatrix}\right\}.$$

If we write $\mathbf{B} = \begin{bmatrix} 0 & 4/\sqrt{18} \\ 1/\sqrt{2} & -1/\sqrt{18} \\ 1/\sqrt{2} & 1/\sqrt{18} \end{bmatrix}$, then:

$$\mathrm{Proj}_S = \mathbf{B}\mathbf{B}^{\mathrm{T}} = \begin{bmatrix} 0 & 4/\sqrt{18} \\ 1/\sqrt{2} & -1/\sqrt{18} \\ 1/\sqrt{2} & 1/\sqrt{18} \end{bmatrix} \begin{bmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 4/\sqrt{18} & -1/\sqrt{18} & 1/\sqrt{18} \end{bmatrix}.$$

$$= \begin{bmatrix} 8/9 & -2/9 & 2/9 \\ -2/9 & 5/9 & 4/9 \\ 2/9 & 4/9 & 5/9 \end{bmatrix} = \tfrac{1}{9}\begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$$

$$\mathrm{Ref}_S = 2\mathbf{B}\mathbf{B}^{\mathrm{T}} - \mathbf{I} = \tfrac{2}{9}\begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix} - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \tfrac{1}{9}\begin{bmatrix} 7 & -4 & 4 \\ -4 & 1 & 8 \\ 4 & 8 & 1 \end{bmatrix}.$$

Special Case: If $V = \mathbf{R}^n$ and $\mathcal{B} = \{\mathbf{u}_1, \cdots, \mathbf{u}_n\}$ is an orthonormal basis for all of $\mathbf{R}^n$, then

$\text{Proj}_V = \text{Identity}$ and $\mathbf{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_n \\ \downarrow & & \downarrow \end{bmatrix}$ will be an $n \times n$ matrix with ON columns (hence invertible), and $\text{Proj}_V = \mathbf{B}\mathbf{B}^{\mathrm{T}} = \mathbf{I}$. Therefore in this special case we'll have $\mathbf{B}^{-1} = \mathbf{B}^{\mathrm{T}}$. Such a matrix is called an **orthogonal matrix**. We'll take a different approach to this in the next lecture when we discuss isometries and orthogonal transformations.

Note: If $\mathbf{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k \\ \downarrow & & \downarrow \end{bmatrix}$ is any $n \times k$ matrix with orthonormal columns, it's easy to calculate that

$$\mathbf{B}^{\mathrm{T}}\mathbf{B} = \begin{bmatrix} \leftarrow & \mathbf{u}_1 & \rightarrow \\ & \vdots & \\ \leftarrow & \mathbf{u}_k & \rightarrow \end{bmatrix} \begin{bmatrix} \uparrow & & \uparrow \\ \mathbf{u}_1 & \cdots & \mathbf{u}_k \\ \downarrow & & \downarrow \end{bmatrix} = \begin{bmatrix} \mathbf{u}_1 \cdot \mathbf{u}_1 & \cdots & \mathbf{u}_1 \cdot \mathbf{u}_k \\ \vdots & \ddots & \vdots \\ \mathbf{u}_k \cdot \mathbf{u}_1 & \cdots & \mathbf{u}_k \cdot \mathbf{u}_k \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{bmatrix} = \mathbf{I}_k .$$

In the special case where **B** is an $n \times n$ matrix with orthonormal columns, this gives $\mathbf{B}^{\mathrm{T}}\mathbf{B} = \mathbf{I}_n$.

Notes by Robert Winters