LINEAR ALGEBRA REVIEW

When we define a term, we put it in **boldface**. This is a very compressed review; please read it very carefully and be sure to ask questions on parts you aren't sure of.

We denote the set of real numbers by $\mathbb{R}$. If $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ is a column vector with ith entry x_i and $y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$ is another column vector, their **dot product** is $x \cdot y = \sum_{i=1}^n x_i y_i$, which can also be written as the transpose, x', times y, i.e., $x'y$. The two vectors are called **orthogonal** if $x \cdot y = 0$; in this case, we write $x \perp y$. The **norm** of x is $\|x\| = \sqrt{x \cdot x}$; this is the length of the vector x.

If $A = [\, a_1 \quad a_2 \quad \cdots \quad a_n \,]$ is a matrix with n columns, where a_i is the ith column, and $x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{R}^n$ is a column vector with ith entry x_i, then $Ax = \sum_{i=1}^n x_i a_i$. In particular, $a_i = A \begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$, where the 1 is in the ith place.

Suppose the matrix product AB is defined. If $B = [\, b_1 \quad b_2 \quad \cdots \quad b_m \,]$ with columns b_j, then $AB = [\, Ab_1 \quad Ab_2 \quad \cdots \quad Ab_m \,]$, i.e., the columns of AB are obtained by multiplying the columns of B by A. The matrix product is also $B'A'$ defined, and we have $B'A' = (AB)'$.

The **span** of a list $(v_1, \ldots, v_k)$ of vectors is the set of vectors that can be written as linear combinations $\sum_{i=1}^k x_i v_i$ of vectors in the list for some real numbers $x_1, \ldots, x_k$.

A list $(v_1, \ldots, v_k)$ of vectors is **linearly independent** if the only linear combination $\sum_{i=1}^k x_i v_i$ that equals the 0 vector is the combination in which all the coefficients x_i are themselves 0. Equivalently, the list is linearly independent when the only column vector solution x of the matrix equation $[\, v_1 \quad \cdots \quad v_k \,]\, x = \mathbf{0}_n$ is $x = \mathbf{0}_n$. In this case, the matrix $[\, v_1 \quad \cdots \quad v_k \,]$ is called **one-to-one** or **injective**. The list is called **orthonormal** if each vector in the list has norm 1 and each pair from the list is orthogonal.

A list $(v_1, \ldots, v_k)$ in a vector space V is a **basis** for V if the list is linearly independent and the span of the list is V. Equivalently, every vector in V can be written in exactly one way as a linear combination of the list. In this case, k is uniquely determined by V (this is proved in a linear algebra course) and is called the **dimension** of V, written $\dim V$. Suppose that there is an inner (dot) product in V, like there is in $\mathbb{R}^n$. Then a list of vectors $(v_1, \ldots, v_k)$ is **orthonormal** if

$$v_i \cdot v_j = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

In linear algebra, it is proved that every (finite-dimensional) vector space with an inner product has an orthonormal basis. In fact, it is proved that every orthonormal list may be extended to an orthonormal basis.

A **linear map** $T\colon \mathbb{R}^n \to \mathbb{R}^m$ is a map that satisfies $T(ax + by) = aT(x) + bT(y)$ for all vectors x, y and all reals a, b. Given a basis $(v_1, \ldots, v_n)$ of $\mathbb{R}^n$ and a basis $(w_1, \ldots, w_m)$ of $\mathbb{R}^m$, such a map has an $m \times n$ matrix with respect to these bases; the (i, j)-entry $t_{i,j}$ of the matrix is given by the coefficient of w_i in $T(v_j)$; i.e., $T(v_j) = \sum_{i=1}^m t_{i,j} w_i$. If the basis $(w_1, \ldots, w_m)$ happens to be orthonormal, then this coefficient can be easily calculated by the equation $t_{i,j} = w_i \cdot T(v_j)$. Unless otherwise specified, we use the standard basis of $\mathbb{R}^n$ for all n, which is orthonormal. Often, this standard basis is written $(\mathbf{e}_1, \ldots, \mathbf{e}_n)$, where $\mathbf{e}_i = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ with the 1 in the ith place, so that a vector $(a_1, \ldots, a_n)$ is written as $\sum_{i=1}^n a_i \mathbf{e}_i$.

A subset W of $\mathbb{R}^n$ is a **subspace** if $\mathbf{0}_n \in W$ and W is closed under vector addition and scalar multiplication. A theorem says that if W is a subspace of $\mathbb{R}^n$ and $\dim W = n$, then $W = \mathbb{R}^n$.

The span of $(v_1, \ldots, v_k)$ is the same as the image of the linear map whose matrix (in the standard bases) is $A = [\, v_1 \quad \cdots \quad v_k \,]$; see Linear Algebra Homework problem 2. This is a subspace, called the **column space** of A and written $\operatorname{col} A$. Its dimension is called the **rank** of A. Another subspace associated to A is its **null space**, the set of vectors x that satisfy $Ax = \mathbf{0}_n$. Thus, A is injective iff ["if and only if"] its null space is just $\{\mathbf{0}_n\}$. A theorem says that if A is $n \times n$, then A is injective iff its rank is n (so it is surjective) iff A is invertible.

In order to illustrate some new concepts, we'll use the following examples: In $\mathbb{R}^2$, denote one axis by $\mathbb{R}_1 := \left\{ \begin{bmatrix} a \\ 0 \end{bmatrix} \; ; \; a \in \mathbb{R} \right\}$ and the other axis by $\mathbb{R}_2 := \left\{ \begin{bmatrix} 0 \\ a \end{bmatrix} \; ; \; a \in \mathbb{R} \right\}$. (The semi-colon stands for "such that". The letter a is a dummy variable. Any other variable could have been used and it would not change the sets.) Then $\mathbb{R}_1$ and $\mathbb{R}_2$ are subspaces of $\mathbb{R}^2$. In fact, $\mathbb{R}_i$ is the span of $(\mathbf{e}_i)$ for $i = 1, 2$. Note that every vector in $\mathbb{R}_1$ is orthogonal to every vector in $\mathbb{R}_2$. We write this as $\mathbb{R}_1 \perp \mathbb{R}_2$. Also note that every vector in $\mathbb{R}^2$ can be written (in only one way) as a sum of one vector from $\mathbb{R}_1$ and one vector from $\mathbb{R}_2$, namely, $\begin{bmatrix} a \\ b \end{bmatrix} = a\mathbf{e}_1 + b\mathbf{e}_2$. We write this as $\mathbb{R}^2 = \mathbb{R}_1 \oplus \mathbb{R}_2$.

Note too that the only vectors that are orthogonal to all vectors in $\mathbb{R}_1$ are the vectors in $\mathbb{R}_2$, and vice versa. We write this as $\mathbb{R}_1^\perp = \mathbb{R}_2$ and $\mathbb{R}_2^\perp = \mathbb{R}_1$. More generally, for any subset W of $\mathbb{R}^n$, we write $v \perp W$ for a vector v if $v \perp w$ for every $w \in W$. We write $W^\perp$ for the set of vectors v that satisfy $v \perp W$. This set is a subspace (check!). In the preceding example, besides what we wrote, we also have $\mathbb{R}_2 = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}^\perp$ and $\mathbb{R}^2 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}^\perp$.

Assume now that W is not merely a subset of $\mathbb{R}^n$, but also a subspace. Call d the dimension of W. Choose an orthonormal basis $(w_1, \dots, w_d)$ of W and extend it to an orthonormal basis $(w_1, \dots, w_d, w_{d+1}, \dots, w_n)$ of $\mathbb{R}^n$. We can write any vector $v \in \mathbb{R}^n$ in this basis: $v = \sum_{i=1}^n a_i w_i$. When is $v \in W$? Because a vector can be written as a linear combination of basis vectors in only one way, $v \in W$ iff $a_i = 0$ for all $i \geq d+1$. When is $v \perp W$? We claim this holds iff $a_j = 0$ for all $j \leq d$. First, if $v \perp W$, then $v \perp w_j$ for all $j \leq d$ (because all such vectors w_j lie in W), i.e., $v \cdot w_j = 0$ for all $j \leq d$, i.e., $a_j = 0$ for all $j \leq d$. Second, if $a_j = 0$ for all $j \leq d$, then $v \cdot w = 0$ for all $w \in W$ since for $w = \sum_{j=1}^d b_j w_j$, all dot products in the expansion of $v \cdot w = \left(\sum_{i=d+1}^n a_i w_i \right) \cdot \left(\sum_{j=1}^d b_j w_j \right)$ are 0. In other words, we may write

$$v = \sum_{j=1}^d a_j w_j + \sum_{i=d+1}^n a_i w_i \,,$$

where the first sum is a vector in W and the second sum is a vector in $W^\perp$. Clearly this was the only way to write v as a sum of a vector in W plus a vector in $W^\perp$. We have proved that $\mathbb{R}^n = W \oplus W^\perp$. Also, we proved that $(w_{d+1}, \dots, w_n)$ form a basis for $W^\perp$. Finally, similar reasoning shows that the only vectors orthogonal to $W^\perp$ are the vectors in W, i.e., $\left(W^\perp\right)^\perp = W$.

To recap, $\mathbb{R}^n = W \oplus W^\perp$, which means, by definition, that for every $v \in \mathbb{R}^n$, there are unique $w \in W$ and $x \in W^\perp$ such that $v = w + x$. Since $w \perp x$, we have the Pythagorean Theorem: $\|v\|^2 = \|w\|^2 + \|x\|^2$ (check!). Since to each v there corresponds a unique w in this way, this defines the **orthogonal projection** $P_W \colon \mathbb{R}^n \to \mathbb{R}^n$ **onto** W as the map $P_W(v) = w$. In the example above, $P_{\mathbb{R}_1}(\begin{bmatrix} a \\ b \end{bmatrix}) = \begin{bmatrix} a \\ 0 \end{bmatrix}$. In general, with the orthonormal basis convenient for W, we have

$$P_W \Big(\sum_{i=1}^n a_i w_i \Big) = \sum_{i=1}^d a_i w_i \,.$$

In other words, we just set to 0 all the coefficients of the basis that lie outside (and thus perpendicular to) W. Note in particular that for $w \in W$, we have $P_W(w) = w$ and for $x \perp W$, we have $P_W(x) = \mathbf{0}_n$.

Orthogonal projection is always a linear map (check!) and a theorem says that it finds the closest point in W to v, i.e., $\|v - P_W(v)\| < \|v - u\|$ for all $u \in W$ except for $u = P_W(v)$. (Proof: Write $v = w + x$ with $w \in W$ and $x \perp W$. Thus, $w = P_W(v)$. For all $u \in W$, we have $w - u \in W$, whence $w - u \perp x$, which implies that

$$\|v - u\|^2 = \|(w - u) + x\|^2 = \|w - u\|^2 + \|x\|^2$$

by the Pythagorean Theorem. This is a minimum exactly when $\|w - u\| = 0$, i.e., when $u = w$, as desired.) If A is a matrix of P_W, then $A^2 = A$ (check!). In the example above, the matrix of $P_{\mathbb{R}_1}$ with respect to the standard basis of $\mathbb{R}^2$ is $\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$. If the basis is orthonormal, then a theorem says that $A = A'$. The column space of A is W and its null space is $W^\perp$ (check!). In particular, the rank of A equals $\dim W$. If I denotes the identity map, then $I - P_W$ is the orthogonal projection onto $W^\perp$ (check!). If W_1 and W_2 are subspaces of $\mathbb{R}^n$ with $W_1 \subseteq W_2$, then $P_{W_1} P_{W_2} = P_{W_1}$ (check!).

Here are some more examples of the above concepts and definitions. You should verify that what is stated is true.

Let $U := \left\{ \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} ;\ a \in \mathbb{R} \right\}$, $W := \left\{ \begin{bmatrix} 0 \\ a \\ 0 \end{bmatrix} ;\ a \in \mathbb{R} \right\}$, and $V := \left\{ \begin{bmatrix} a \\ a \\ 0 \end{bmatrix} ;\ a \in \mathbb{R} \right\}$. Then

$$\left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} ;\ a, b \in \mathbb{R} \right\} = U \oplus W = U \oplus V\,,$$

meaning that there is exactly one way to write each vector $\begin{bmatrix} a \\ b \\ 0 \end{bmatrix}$ as a sum of a vector in U plus a vector in W; and likewise as a sum of a vector in U plus a vector in V.

A list of orthonormal vectors is automatically linearly independent. The list

$$\left(\begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} 1/2 \\ 1/2 \\ -1/2 \\ -1/2 \end{bmatrix}, \begin{bmatrix} 1/2 \\ -1/2 \\ -1/2 \\ 1/2 \end{bmatrix}, \begin{bmatrix} -1/2 \\ 1/2 \\ -1/2 \\ 1/2 \end{bmatrix} \right)$$

in $\mathbb{R}^4$ is orthonormal. Since the list has 4 vectors, it is also a basis for $\mathbb{R}^4$.

Let $u_1 := \begin{bmatrix} 2 \\ 5 \\ -1 \end{bmatrix}$, $u_2 := \begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$, $v := \begin{bmatrix} 5 \\ 10 \\ 15 \end{bmatrix}$, $w := \begin{bmatrix} -2 \\ 10 \\ 1 \end{bmatrix}$, and $x := \begin{bmatrix} 7 \\ 0 \\ 14 \end{bmatrix}$. Let W be the span of (u_1, u_2). Then $v = w + x$, $w = (3/2)u_1 + (5/2)u_2 \in W$, $x \in W^\perp$, and $w = P_W(v)$.

In case you had difficulty checking the things above that were marked "check!", here are the solutions. (But make sure you try to do them yourself first. The proofs below are not the only ways to prove these things. You may prefer proofs that use orthonormal bases.)

That $W^\perp$ is a subspace: We have to check the definition of subspace. To show $\mathbf{0}_n \in W^\perp$, we need to show that $\mathbf{0}_n \perp w$ for all $w \in W$. Since $\mathbf{0}_n \cdot w = 0$, this is obvious. Next, we need to show that for all $x, y \in W^\perp$, we have $x + y \in W^\perp$. So take $w \in W$. We need to show $(x+y) \perp w$. But $(x+y)\cdot w = x\cdot w + y\cdot w = 0+0 = 0$, so this is true. Finally, we need to show that for every real a and every $x \in W^\perp$, we have $ax \in W^\perp$. Again, take $w \in W$. We have $(ax)\cdot w = (ax)'w = ax'w = a0 = 0$, so this holds. This completes the proof.

That $w \perp x$ implies $\|v\|^2 = \|w\|^2 + \|x\|^2$, where $v = w + x$: We have

$$\|v\|^2 = v'v = (w+x)'(w+x) = w'w + w'x + x'w + x'x = w'w + 0 + 0 + x'x = \|w\|^2 + \|x\|^2 .$$

That P_W is a linear map: We have to show that for every $x, y \in \mathbb{R}^n$ and every $a, b \in \mathbb{R}$, we have $P_W(ax+by) = aP_W(x) + bP_W(y)$. Since $\mathbb{R}^n = W + W^\perp$, we may write $x = w_1 + u_1$ and $y = w_2 + u_2$, where $w_1, w_2 \in W$ and $u_1, u_2 \in W^\perp$. Then $P_W x = w_1$ and $P_W y = w_2$ by definition of P_W. Now

$$ax + by = (aw_1 + bw_2) + (au_1 + bu_2) .$$

Furthermore, the first summand is in W since W is a subspace and the second summand is in $W^\perp$ since $W^\perp$ is also a subspace. This means that $P_W(ax + by) = aw_1 + bw_2$ by definition of P_W. This is the same as $aP_W(x) + bP_W(y)$.

That if A is a matrix of P_W, then $A^2 = A$: This is the same as saying that $P_W(P_W v) = P_W(v)$ for all $v \in \mathbb{R}^n$. So take $v \in \mathbb{R}^n$. Write $v = w + x$ with $w \in W$ and $x \in W^\perp$. Then $w = P_W(v)$ by definition of P_W. Since $w \in W$, we have $P_W(w) = w$. Substituting $P_W(v)$ for w here, we get what we wanted to show.

That the column space of A is W and its null space is $W^\perp$, where A is a matrix of P_W: The column space of A is the range of P_W. By definition of P_W, its range is included in W. Furthermore, for every $w \in W$, we have $P_W(w) = w$. This means that the range of P_W is actually equal to W. Thus, so is $\operatorname{col} A$. The null space of A is the set of vectors v with $P_W(v) = \mathbf{0}_n$, i.e., with $v = \mathbf{0}_n + x$ for some $x \in W^\perp$ (by definition of P_W). This is precisely the set of $v \in W^\perp$, so the null space is $W^\perp$.

That $I - P_W$ is the orthogonal projection onto $W^{\perp}$: For any $v \in \mathbb{R}^n$, write $v = w + x$ with $w \in W$ and $x \in W^{\perp}$. Since also $v = x + w$ and since $W = \left(W^{\perp}\right)^{\perp}$, this shows that $x = P_{W^{\perp}}(v)$. Since $x = v - w = I(v) - P_W(v) = (I - P_W)(v)$, this shows that $P_{W^{\perp}} = I - P_W$.

That if W_1 and W_2 are subspaces of $\mathbb{R}^n$ with $W_1 \subseteq W_2$, then $P_{W_1} P_{W_2} = P_{W_1}$: Let $v \in \mathbb{R}^n$ and write $v = w + x$ with $w \in W_2$ and $x \perp W_2$. Then also $x \perp W_1$. We have $P_{W_1}(v) = P_{W_1}(w) + P_{W_1}(x) = P_{W_1}(w) = P_{W_1} P_{W_2}(v)$, which proves the identity since v was arbitrary.

Test your understanding by doing the following quiz.

Quiz: Say whether each is true or false and why. You will be given this same quiz in class.

1. The orthogonal projection of v onto the span of a single vector w is always a scalar multiple of v.

2. If a vector v coincides with its orthogonal projection onto a subspace W, then $v \in W$.

3. If W is a subspace of $\mathbb{R}^n$, then W and $W^{\perp}$ have no vectors in common.

4. If W is a subspace of $\mathbb{R}^n$, then $\|P_W(v)\|^2 + \|v - P_W(v)\|^2 = \|v\|^2$ for all $v \in \mathbb{R}^n$.