LINEAR ALGEBRA REVIEW

You might benefit from the lecture notes available at `http://www.laylinalgebra.com/free_site/tm_index.html`, which are designed for a course using the excellent book *Linear Algebra* by David Lay. In particular, see Sections 1.7, 1.8, 1.9, 2.1, 4.1, 4.2, 4.3, 4.5, 4.6, 6.1, 6.2, 6.3, and 6.5 of the notes. There are also videotaped lectures by a master teacher at `http://ocw.mit.edu/OcwWeb/Mathematics/18-06Spring-2010/CourseHome/index.htm`.

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.

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 $[\, v_1 \quad \cdots \quad v_k \,] x = 0$ is $x = 0$. 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 $\dim V$. The basis is **orthonormal** if

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

A **linear map** $T : \mathbb{R}^n \to \mathbb{R}^m$ 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 Tv_j; i.e., $Tv_j = \sum_{i=1}^m t_{i,j} w_i$. If the basis $(w_1, \ldots, w_m)$ happens to be othonormal, then this coefficient can be easily calculated by the equation $t_{i,j} = w_i \cdot Tv_j$. Unless otherwise specified, we use the standard basis of $\mathbb{R}^n$ for all n, which is orthonormal.

A subset W of $\mathbb{R}^n$ is a **subspace** if $0 \in W$ 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 $A = [\, v_1 \quad \cdots \quad v_k \,]$. 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 $Ax = 0$. Thus, A is injective iff its null space is just $\{0\}$. 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.

If $W_1, \ldots, W_k$ are each subspaces of a vector space V, we write $W_1 + W_2 + \cdots + W_k$ for the set of vectors that can be written as a sum $\sum_{i=1}^k w_i$. If every vector $v \in V$ can be written in exactly one way as a sum $v = \sum_{i=1}^k w_i$ with each $w_i \in W_i$, then we write $V = W_1 \oplus \cdots \oplus W_k$. This is called a **direct sum**.

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!). Assume now that W is also a subspace. Then a theorem says that $\left(W^\perp\right)^\perp = W$ and $\dim W + \dim W^\perp = n$; moreover, we then have $\mathbb{R}^n = W \oplus W^\perp$. In other words, 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** $P_W : \mathbb{R}^n \to \mathbb{R}^n$ **onto** W as the map $P_W(v) = w$. It is 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!). 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 $(v_1, \ldots, v_n)$ is an orthonormal basis of $\mathbb{R}^n$ such that $(v_1, \ldots, v_p)$ is a basis for W, where $p = \dim W$, then P_W has a very simple form: we can write any $v \in \mathbb{R}^n$ uniquely as $v = \sum_{i=1}^n a_i v_i$, and then $P_W(v) = \sum_{i=1}^p a_i v_i$. In other words, we just set to 0 all the coefficients of the basis that lie outside (and thus perpendicular to) W. Also, $(v_{p+1}, \ldots, v_n)$ form a basis for $W^\perp$ (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 examples of the above concepts and definitions. You should verify that what is stated is true.

Let $U := \left\{ \begin{bmatrix} a \\ 0 \\ 0 \end{bmatrix} \; ; \; a \in \mathbb{R} \right\}$ and $W := \left\{ \begin{bmatrix} 0 \\ a \\ 0 \end{bmatrix} \; ; \; a \in \mathbb{R} \right\}$. (The semi-colon stands for "such that". Thus, U is the set of all column vectors of length 3 whose first entry is any real number and whose other 2 entries are 0. The letter a is a dummy variable. Any other variable could have been used and it would not change the set U.) Then

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

If $V := \left\{ \begin{bmatrix} a \\ a \\ 0 \end{bmatrix} \; ; \; a \in \mathbb{R} \right\}$, then $U + V = U + W = U \oplus V$.

Let $U_1 := \left\{ \begin{bmatrix} a \\ b \\ 0 \end{bmatrix} \; ; \; a, b \in \mathbb{R} \right\}$, $U_2 := \left\{ \begin{bmatrix} 0 \\ 0 \\ a \end{bmatrix} \; ; \; a \in \mathbb{R} \right\}$, and $U_3 := \left\{ \begin{bmatrix} 0 \\ b \\ b \end{bmatrix} \; ; \; b \in \mathbb{R} \right\}$.

Then $\mathbb{R}^3 = U_1 + U_2 + U_3$, but this is *not* a direct sum. A basis for U_1 is $\left(\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \right)$.

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

$$\left(\begin{bmatrix}0/2\\1/2\\1/2\\1/2\end{bmatrix}, \begin{bmatrix}1/2\\1/2\\-1/2\\-1/2\end{bmatrix}, \begin{bmatrix}1/2\\-1/2\\-1/2\\1/2\end{bmatrix}, \begin{bmatrix}-1/2\\1/2\\-1/2\\1/2\end{bmatrix}\right)$$

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.)

That $W^\perp$ is a subspace: We have to check the definition of subspace. To show $0 \in W^\perp$, we need to show that $0 \perp w$ for all $w \in W$. Since $0 \cdot w = 0$, 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 $(x + y) \cdot w = (x + y)'w = x'w + y'w = 0 + 0 = 0$, 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. Now $w = w + 0$ with $0 \in W^\perp$, so $P_W(w) = w$. This is 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 $w = w + 0$ and $0 \in W^\perp$, so $P_W(w) = w$. 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 $P_W(v) = 0$, i.e., with $v = 0 + x$ 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.

That if $(v_1, \ldots, v_n)$ is an orthonormal basis of $\mathbb{R}^n$ such that $(v_1, \ldots, v_p)$ is a basis for W, where $p = \dim W$, then $(v_{p+1}, \ldots, v_n)$ form a basis for $W^\perp$: Clearly the latter vectors are orthogonal to each v_i for $1 \le i \le p$, whence to all linear combinations of them, i.e., to W. Since $(v_{p+1}, \ldots, v_n)$ is linearly independent, this list is a basis for its span, U. We have just verified that U is a subspace of $W^\perp$. Since $\dim U = n - p = n - \dim W = \dim W^\perp$, it follows that $U = W^\perp$. This proves the claim.

Test your understanding by doing the following **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 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 $v \in \mathbb{R}^n$.

Homework:

1. Show that if $w \neq 0$ and W is the set of scalar multiples of w, then $P_W(v) = (v \cdot w)w/\|w\|^2$ and the matrix for P_W is $ww'/\|w\|^2$.

2. Let X be an $n \times p$ matrix. Let $W = \operatorname{col} X \subseteq \mathbb{R}^n$.

(a) Show that W is the same as the set of all vectors of the form Xc, where $c \in \mathbb{R}^p$.

(b) Show that $W^\perp$ is the same as the set of all vectors $u \in \mathbb{R}^n$ such that $X'u = 0$.

3. Let X be an $n \times p$ matrix. Suppose that the columns of X are linearly independent. Let $W = \operatorname{col} X \subseteq \mathbb{R}^n$.

(a) Use Problem 1 to show that in the special case that $p = 1$, the matrix of P_W is $X(X'X)^{-1}X'$. The rest of this problem shows that this holds in general (i.e., for all p).

(b) Show that if c is a $p \times 1$ vector with $Xc = 0$, then $c = 0$.

(c) Show that if $X'Xc = 0$, then $c = 0$. (Hint: Look at $c'X'Xc$.)

(d) Show that the null space of $X'X$ is $\{0\}$.

(e) Show that $X'X$ is invertible.

(f) Let $H = X(X'X)^{-1}X'$. Show that $HX = X$.

(g) Show that $Hw = w$ for all $w \in W$ and $Hu = 0$ for all $u \in W^\perp$. (Hint: Use (f) for the first part and think about $X'u$ for the second part.)

(h) Show that H is the matrix of P_W.

(i) Let $Q = (X'X)^{-1}X'$, so that $H = XQ$. If $Y \in \mathbb{R}^n$, then $P_W Y \in W$, so that $P_W Y$ is a linear combination of the columns of X. Show that the coefficients of this linear combination form the vector $\hat{\beta} = QY$.

(j) Let $e = Y - X\hat{\beta}$. Show that $e \in W^\perp$.

(k) Show that $\|Y - X\gamma\|$ is minimum exactly when $\gamma = \hat{\beta}$.

(l) Show that $Y - X\gamma \in W^\perp$ exactly when $\gamma = \hat{\beta}$.

4. Let ξ and ζ denote two random vectors in $\mathbb{R}^2$. We say that ξ and ζ are orthogonal if $P(\xi \perp \zeta) = 1$. In this problem, you are to find examples where the following occur, and prove that your examples have these properties. Note that "independent" means "stochastically independent", not "linearly independent".

(a) ξ and ζ are neither independent nor orthogonal;

(b) ξ and ζ are both independent and orthogonal;

(c) ξ and ζ are independent but not orthogonal;

(d) ξ and ζ are not independent but are orthogonal.