Lecture 3
Euclidean Vector Spaces and Projections

The defining condition of a real vector space $V \subset R^n$ is closure under linear transformations, i.e., $\alpha y + \gamma z \in V$ whenever $y, z \in V$ and $\alpha, \beta \in R$. Often we use the Euclidean norm $||y|| = (y'y)^{1/2}$. Then, $(R^n, ||\cdot||)$ is $\mathcal{E}^n$ and $\mathcal{E}^n$ is a normed vector space. Also, $(R^n, \langle \cdot, \cdot \rangle)$ is a Hilbert space (complete inner product space), where $\langle a, b \rangle = a'b$. A subset of a vector space that is a vector space in its own right is called a *linear subspace.*

Recall that the required properties of a norm are: for $a, b \in V$,

(i) $||a|| = 0$ iff $a = 0$
(ii) $||a + b|| \leq ||a|| + ||b||$, and
(iii) $||\alpha a|| = |\alpha| ||a||$ for $\alpha \in R$.

Also, the required properties of an inner product are

(i) $\langle a, b \rangle = \langle b, a \rangle$,
(ii) $\langle \alpha_1 a + \alpha_2 b, c \rangle = \alpha_1 \langle a, c \rangle + \alpha_2 \langle b, c \rangle$ for $\alpha_1, \alpha_2 \in R$, and
(iii) $\langle a, a \rangle = ||a||^2$.

Note that a normed vector space is complete if every absolutely summable sequence is summable. That is, if $\sum_{i=1}^{\infty} ||v_j|| < \infty$, then there exists a $v \in V$ such that $||v - \sum_{i=1}^{m} v_j|| \to 0$ as $m \to \infty$.

Basis of a Vector Space

A set of vectors $v_1, ..., v_m$ is said to *span* a vector space $V \subset R^n$, if each $v \in V$ can be written as

$$v = \sum_{i=1}^{m} \alpha_i v_i \quad \text{for} \quad \alpha_i \in R.$$

The smallest number of vectors that spans V is called the *dimension* of V. We will only consider finite dimensional vector spaces. Vectors $v_1, ..., v_m$ are linearly independent if $\sum_{i=1}^{m} \alpha_i v_i = 0 \Rightarrow \alpha_i = 0 \ \forall i \leq m$, where $\alpha_i \in R$. The largest number of linearly independent vectors is the dimension of V. (This is an equivalent definition of dimension.) A *basis* of a vector space V is any set of linearly independent vectors that spans V. From any basis we can always construct an *orthonormal basis*, i.e., a basis $\{v_i : i \leq \dim(V)\}$ whose vectors are of length one and are orthogonal:

$$v_i' v_i = 1, \ v_i' v_j = 0, \ i \neq j.$$

This can be proved using the Gramm–Schmidt procedure. By definition, any $v \in V$ can be written in terms of an orthonormal basis $v_1, ..., v_p$ (where $p = \dim V$), because by definition $v_1, ..., v_p$ spans V, i.e., $v = \sum_{i=1}^{p} \alpha_i v_i$ for some $\alpha_i \in R \ \forall i \leq p$.

Furthermore, if $L \subset V$ is a linear subspace with orthonormal basis $\alpha_1, ..., \alpha_q$, then we can find additional vectors $\alpha_{q+1}, ..., \alpha_p$ such that $\alpha_1, ..., \alpha_p$ span V and $\alpha_1, ..., \alpha_p$ are orthonormal.

Proof: Consider different vectors until you obtain a collection that includes $\alpha_1, ..., \alpha_q$ and that spans V. The number of vectors in the collection can be reduced to p linearly independent vectors including $\alpha_1, ..., \alpha_q$. Start with $\alpha_1, ..., \alpha_q$ and use the Gramm–Schmidt procedure to obtain the remaining $p - q$ orthonormal vectors.

The *orthogonal complement* of a linear subspace $L \subset V$ is defined by

$$L^{\perp} = \{z \in V : z'\ell = 0, \ \forall \ell \in L\}.$$

It is straightforward to see that $L^{\perp}$ also is a linear subspace.

Theorem 1 (Vector Decomposition Theorem): Let V be a Hilbert space and let L be a linear subspace of V. Every element y of V can be written as the sum of two vectors y_1 and y_2 where $y_1 \in L$ and $y_2 \in L^{\perp}$. That is,

$$y = y_1 + y_2.$$

Proof: Let $v_1, ..., v_p$ be an orthonormal basis of L. Let $y_1 = \sum_{i=1}^{p}(y'v_i)v_i$. Let $y_2 = y - y_1$. Clearly, $y_1 \in L$. Also, for $j = 1, ..., p$,

$$v_j'y_2 = v_j'(y - y_1) = v_j'y - v_j'\sum_{i=1}^{p}(y'v_i)v_i = v_j'y - v_j'v_jv_j'y = 0.$$

Hence, y_2 is orthogonal to v_j, $\forall j = 1, ..., p$ and $y_2 \in L^{\perp}$. $\square$

An alternative way of expressing the above result is to say that $V = L \oplus L^{\perp}$. That is, V is the direct sum of L and $L^{\perp}$.

Let $y \in V$, a vector space. Let L be a linear subspace of V. An element of L is called a *projection* of y on L and is denoted $P_L y$, if

$$||y - P_L y|| = \inf_{\ell \in L} ||y - \ell||.$$

Theorem 2 (Projection Theorem): (a) *$P_L y$ exists, is unique, and is a linear function of y.* (b) *$P_L y \in L$ is the projection of y on L iff $y - P_L y \perp L$. (That is, if we write $y = y_1 + y_2$ for $y_1 \in L$, $y_2 \in L^{\perp}$, then $y_1 = P_L y$.)*

Proof: (a) By the proof that $V = L \oplus L^{\perp}$, we can write $y = y_1 + y_2$, where $y_1 \in L$, $y_2 \in L^{\perp}$, and $y_1 = \sum_{i=1}^{p} y'v_iv_i$. We want to show that $y_1 = P_L y$. If true, then $P_L y$ exists and is a linear function of y. Consider any $\ell \in L$. We have

$$\begin{aligned} ||y - \ell||^2 &= ||y - y_1 + y_1 - \ell||^2 \\ &= ||y - y_1||^2 + ||y_1 - \ell||^2 + 2(y - y_1)'(y_1 - \ell) \\ &= ||y - y_1||^2 + ||y_1 - \ell||^2 \\ &\geq ||y - y_1||^2 \end{aligned}$$

with strict inequality unless $||y_1 - \ell|| = 0$, i.e., unless $\ell = y_1$. (The inner product term is zero because $y - y_1 \in L^{\perp}$ and $y_1 - \ell \in L$.) Hence, y_1 is a projection of y on L and it is unique.
(b) If $P_L y$ is the projection of y on L, then by part (a) $P_L y = y_1$ and $y - P_L y = y_2 \in L^{\perp}$. Conversely, suppose $P_L y$ is some element of L for which $y - P_L y \in L^{\perp}$, then $y_1 + y_2 - P_L y \in L^{\perp}$, $y_1 - P_L y \in L^{\perp}$, $y_1 - P_L y \in L$ (since $y_1 \in L$ and $P_L y \in L$), and, hence, $y_1 - P_L y = 0$. That is, $P_L y$ is the projection of y on L. $\square$

We know that a projection is a linear map from V to L, so we may ask: How can such a linear map be characterized? Answer: It can be characterized in a simple way in terms of what it does to vectors in L and in $L^{\perp}$.

Theorem 3: *The linear map $G : V \to L$ is the* projection map *onto L iff $Gy = y$ $\forall y \in L$, and $Gy = 0$ $\forall y \in L^\perp$.*

Proof: (a) Suppose G is the projection map from V to L. Then, if we consider $y \in L$, we find that y minimizes $||y - \ell||$ over $\ell \in L$ and so $Gy = y$. If $y \in L^\perp$, then $||y - \ell||^2 = ||y||^2 + ||\ell||^2$. The latter is minimized over $\ell \in L$ by $\ell = 0$, so Gy must equal 0.
(b) Conversely, suppose G is a linear map for which $Gy = y$ $\forall y \in L$ and $Gy = 0$ $\forall y \in L^\perp$. Let y be any vector in V. By the vector decomposition theorem, we can write $y = y_1 + y_2$, where $y_1 \in L$ and $y_2 \in L^\perp$. In consequence, $Gy = Gy_1 + Gy_2 = y_1$. Note that $y - y_1 = y_2 \perp L$. So, by the projection theorem, $y_1 = P_L y$. That is, Gy is the projection of y on L. □

Matrices as Linear Maps

An $n \times k$ matrix is just a linear transformation from $\mathcal{E}^k$ to $\mathcal{E}^n$:

$$X(\alpha_1 y_1 + \alpha_2 y_2) = \alpha_1 X y_1 + \alpha_2 X y_2 \quad \forall y_1, y_2 \in \mathcal{E}^k,\ \forall \alpha_1, \alpha_2 \in R.$$

The *range* of the function (or linear map) X is $R(X) = \{x \in R^n : x = Xb \text{ for some } b \in R^k\}$. Alternatively, the range is called the *column space of* X, because $R(X)$ consists of all linear combinations of the columns of X.

Note that $R(X)$ is a vector space. It is a linear subspace of $\mathcal{E}^n$. This holds because

$$\begin{aligned} x_1, x_2 \in R(X) &\Rightarrow x_1 = Xb_1,\ x_2 = Xb_2 \\ &\Rightarrow \alpha x_1 + \beta x_2 = X(\alpha b_1 + \beta b_2) \in R(X) \text{ since } \alpha b_1 + \beta b_2 \in R^k. \end{aligned}$$

Another linear subspace associated with X is a linear subspace of R^k called the *null space of* X. By definition, the null space of X is

$$N(X) = \{b \in R^k : Xb = 0\}.$$

Note that $N(X)$ is a linear subspace, because

$$\begin{aligned} b_1, b_2 \in N(X) &\Rightarrow X(\alpha b_1 + \beta b_2) = \alpha X b_1 + \beta X b_2 = 0 \\ &\Rightarrow \alpha b_1 + \beta b_2 \in N(X). \end{aligned}$$

The null space of X contains all vectors that are orthogonal or perpendicular to all rows of X.

The range space and null space of X are related as follows: $N(X) = R(X')^\perp$, where $\perp$ denotes orthogonal complement. This holds because $b \perp R(X') \Leftrightarrow Xb = 0 \Leftrightarrow b \in N(X)$. This result implies that $R^k = R(X') \oplus N(X)$.

As an example, suppose

$$X = \begin{bmatrix} 1 & 1 \\ -2 & -2 \end{bmatrix}.$$

Then,

$$N(X) = \{y \in R^2 : Xy = 0\} = \{y : (1,1)y = 0\} = \{y : y_1 = -y_2\}.$$

In addition, we have

$$X' = \begin{bmatrix} 1 & -2 \\ 1 & -2 \end{bmatrix} \text{ and } \begin{aligned} R(X') &= \{y : y = X'b \text{ for some } b \in R^2\} \\ &= \{y : y_1 = y_2\}. \end{aligned}$$

Thus, in this example, $N(X) = R(X')^{\perp}$.

Projection Maps

A projection map (like any linear map) can be represented as a matrix. What does such a matrix look like?

Theorem 4: *For the Euclidean vector space $V \subset R^n$, a linear map $G : V \to V$ is a projection map (onto its range $R(G)$) iff $G' = G$ and $G^2 = G$. That is, iff G is symmetric and idempotent.*

Proof: (a) Suppose G is a projection map. Let x, y be any vectors in V. Write $x = x_1 + x_2$ and $y = y_1 + y_2$, where $x_1, y_1 \in L$, $x_2, y_2 \in L^{\perp}$, and $L = R(G)$. Note that $x_1 = Gx$ and $y_1 = Gy$. Now,

$$\begin{aligned} x'G'y = (Gx)'(y_1 + y_2) &= (Gx)'y_1 && \text{because } Gx \in L,\ y_2 \in L^{\perp} \\ &= (x - x_2)'y_1 \\ &= x'y_1 && \text{because } x_2'y_1 = 0 \\ &= x'Gy. \end{aligned}$$

So, $G = G'$.

Also, $G^2 y = G(Gy) = Gy$ because $Gy \in L$ and by the theorem above $Gx = x\ \forall x \in L$.

(b) Suppose $G^2 = G$ and $G = G'$. Let g be in $R(G)$. That is, $g = Gv$ for some $v \in V$. We need to show that $(y - Gy)'g = 0$. Then, the projection theorem implies that Gy is the projection of y on $R(G)$. But,

$$\begin{aligned} (y - Gy)'g = (y - Gy)'Gv &= [G(y - Gy)]'\nu && \text{by symmetry of } G \\ &= (Gy - Gy)'v && \text{by idempotency of } G \\ &= 0. \quad \square \end{aligned}$$

Thus, all projection maps in $\mathcal{E}^n$ are symmetric idempotent matrices. So, if P_1 and P_2 are symmetric idempotent matrices, then they are both projection matrices onto linear subspaces. If $P_1 P_2 = 0$, then the subspaces that P_1 and P_2 project onto are orthogonal. Now we see that the condition $P_1 P_2 = 0$, which guarantees statistical independence of $y'P_1 y$ and $y'P_2 y$, for P_1, P_2 symmetric and idempotent matrices, makes sense. The reason is that we can write: $y'P_1 y = (P_1 y)'P_1 y$ and $y'P_2 y = (P_2 y)'P_2 y$, and $P_1 y$ and $P_2 y$ are those parts of the vector y that lie in the orthogonal subspaces $R(P_1)$ and $R(P_2)$. Because a multivariate normal density with covariance matrix $\sigma^2 I$ is spherically symmetric, these two pieces are statistically independent.

Suppose we define a linear subspace using the linear map or matrix X via $R(X) = \{v \in R^n : v = Xb$, for some $b \in R^k\}$. What is the projection map onto $R(X)$? To answer this question, we need to define g-inverses, and in particular, the Moore–Penrose inverse.

A g-inverse of an $r \times s$ matrix H is any $s \times r$ matrix H^- such that

$$HH^-H = H.$$

Note that if the system of equations $Hx = b$ has a solution for given H and b, then $x = H^-b$ is a solution. This follows because if $Hx = b$ has a solution, then $b \in R(H)$, i.e., $b = Hw$ for some $w \in R^r$. In consequence, we obtain

$$\begin{aligned} Hx = HH^-b &= HH^-Hw, \\ &= Hw \\ &= b. \end{aligned}$$

For example, consider the normal equations for the least squares estimator $\widehat{\beta}$: $X'X\widehat{\beta} = X'y$. The solution is $\widehat{\beta} = (X'X)^{-}X'y$. If $X'X$ is nonsingular, then $(X'X)^{-}$ is unique and equals $(X'X)^{-1}$.

If H is diagonal with less than full rank, i.e.,

$$H = \begin{bmatrix} h_1 & & & 0 \\ & \ddots & & \\ & & h_p & \\ 0 & & & 0 \end{bmatrix},$$

then any matrix of the form

$$\begin{bmatrix} h_1^{-1} & & 0 & H_2 \\ & \ddots & & \\ 0 & & h_p^{-1} & \\ H_3 & & & H_4 \end{bmatrix},$$

where H_2, H_3, and H_4 are arbitrary, is a g-inverse of H.

For a symmetric matrix A, we can find the totality of g-inverses of A by writing $A = B\Lambda B'$, where Λ is diagonal and B is orthogonal, and defining $A^{-} = B\Lambda^{-}B'$ for some g-inverse of Λ. All g-inverses of a diagonal matrix are of the form given above.

Note that $rk(H^{-}) \geq rk(H)$, where $rk(\cdot)$ denotes the rank of a matrix. This follows from the definition $HH^{-}H = H$. Also, there always exist full rank g-inverses. By placing additional conditions on the g-inverse, we can find g-inverses that have the same rank as the original matrix.

A g-inverse H^{-} is *reflexive* if it satisfies $H^{-}HH^{-} = H^{-}$.

Theorem 5: *H^{-} is a reflexive g-inverse iff $rk(H^{-}) = rk(H)$.*

Proof: See Rao and Mitra (1971) or Pringle and Rayner (1971). (The proof is easy, actually, at least in one direction.)

Definition: *The* Moore–Penrose i*nverse of a matrix, denoted H^{+}, is a uniquely defined reflexive g-inverse that satisfies*

$$(HH^{+})' = HH^{+} \textit{and } (H^{+}H)' = H^{+}H.$$

For example, the Moore–Penrose inverse of a non-zero scalar is its reciprocal. If the scalar is zero, the Moore–Penrose inverse is zero. Any number is a g-inverse of zero.

The Moore–Penrose inverse of a diagonal matrix $\Lambda = \text{diag}(\lambda_1, ..., \lambda_p, 0, ..., 0)$ is just $\Lambda^{+} = \text{diag}(\lambda_1^{-1}, ..., \lambda_p^{-1}, 0, ..., 0)$. So, we can find the Moore–Penrose inverse of any matrix by diagonalization. (Use the singular value decomposition for non-square matrices.)

Theorem 6: *The projection map onto $R(X)$ is $P_X = X(X'X)^{-}X'$ (for any choice of g-inverse of $X'X$).*

Proof: We will show that $P_X y \in R(X)$ and $y - P_X y \perp R(X)$. Then, the projection theorem gives the result. Clearly, $P_X y \in R(X)$. To show $y - P_X y \perp R(X)$, write $y = y_1 + y_2$, where $y_1 \in R(X)$ and $y_2 \perp R(X)$. That is, $y = Xw + y_2$, for some $w \in R^k$. Consider any $b \in R(X)$. We can write $b = Xv$, for some $v \in R^k$, and we have

$$
\begin{aligned}
b'(y - P_X y) &= v'X'(I - P_X)(Xw + y_2) \\
&= v'X'(I - P_X)Xw + v'X'(I - P_X)y_2 \\
&= v'[X'X - X'X(X'X)^-(X'X)]w + v'X'y_2 - v'X'X(X'X)^-X'y_2 \\
&= 0,
\end{aligned}
$$

using the fact that $X'y_2 = 0$ because $y_2 \perp R(X)$. Thus, $y - P_X y \perp R(X)$. □

So, if $y \in R^n$ and we want to project y onto the column space of X, i.e., we want to find the closest point in $R(X)$ to y, we have the answer: $X(X'X)^-X'y$.

The matrix $X(X'X)^-X'$ is the same no matter what g-inverse is used. This follows from the uniqueness of projections.

The final results for projection matrices onto $R(X)$ are the following:

Theorem 7: (a) $rk(P_X) = rk(X)$ *and*
(b) $rk(I - P_X) = n - rk(X)$, where X is an $n \times k$ matrix.

Proof: Since $P_X = X(X'X)^-X$ does not depend on the choice of g-inverse, we can use the Moore–Penrose inverse. Then, we have

$$
\begin{aligned}
rk(P_X) &= \mathrm{tr}(X(X'X)^+X') && \text{because } P_X \text{ is idempotent,} \\
&= \mathrm{tr}(X'X(X'X)^+) && \text{by the properties of the trace,} \\
&= \mathrm{tr}(B\Lambda B'B\Lambda^+B') && \text{since } X'X = B\Lambda B', \\
&= \mathrm{tr}(\Lambda\Lambda^+) && \\
&= rk(\Lambda\Lambda^+) && \text{because } \Lambda\Lambda^+ \text{ is idempotent,} \\
&= rk(\Lambda) && \text{because } rk(\Lambda^+) \geq rk(\Lambda) \text{ and } \Lambda\Lambda^+\Lambda = \Lambda, \\
& && \text{i.e., } R(\Lambda^+) \supseteq R(\Lambda), \\
&= rk(X'X) && \text{because } X'X = B\Lambda B' \text{ and } B \text{ is full rank,} \\
&= rk(X) && \text{by known results.}
\end{aligned}
$$

Also,

$$
\begin{aligned}
rk(I - P_X) &= \mathrm{tr}(I - P_X) \quad \text{because } I - P_X \text{ is idempotent,} \\
&= n - \mathrm{tr}(P_X) \\
&= n - rk(X). \quad \square
\end{aligned}
$$