This document has been substantially revised.

1. Vector subspaces

What does it mean to have a vector space inside another vector space, such as a tilted plane sitting inside three-dimensional space?

Definition. Let V be a vector space. Then a nonempty subset $W \subseteq V$ is a *vector subspace* of V if any linear combination of elements of W is again in W.[1]

Note that any vector subspace of V is a vector space in its own right. The addition and scalar multiplication operations on W are inherited from V. In particular, every vector subspace contains $0 \in V$.

If $W \subseteq \mathbb{R}^n$ is a vector subspace, how can we specify what W is? In general, there are two ways:

- We can write down some vectors $v_1, \ldots, v_s \in \mathbb{R}^n$ such that W is the set of all linear combinations of $v_1, \ldots, v_s$. In other words, $W = \mathrm{col}(A)$ where A is the $n \times s$ matrix whose columns are $v_1, \ldots, v_s$. When giving your answer in this form, you should try to use as few vectors as possible. (The number of vectors needed is equal to the dimension of W, denoted $\dim(W)$.)
- We can write down some equations that a vector $(x_1, \ldots, x_n) \in \mathbb{R}^n$ must satisfy in order to belong to W. I.e., we can try to find numbers a_{ij} such that
$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in W \quad \text{if and only if} \quad \begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\ &\vdots \\ a_{s1}x_1 + a_{s2}x_2 + \cdots + a_{sn}x_n &= 0 \end{aligned}$$
In other words, if $A = (a_{ij})$ is the $s \times n$ matrix consisting of the a_{ij}'s, we have expressed $W = \mathrm{null}(A)$. When giving your answer in this form (e.g. for Problem 5b in Homework 4), you should try to use as few equations as possible. (The number of equations needed is equal to $n - \dim(W)$.)

For example, let $W = \mathrm{col}\left(\begin{pmatrix} 1 & 1 & 2 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}\right)$, which is a vector subspace of $\mathbb{R}^3$.

- In the first format, we could say that
$$W = \mathrm{col}\left(\begin{pmatrix} 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{pmatrix}\right).$$
This is equivalent to saying that
$$W = \left\{ \begin{pmatrix} y \\ x \\ x \end{pmatrix} \text{ for all } x, y \in \mathbb{R} \right\}.$$
Since $\dim(W) = 2$, we need to take all linear combinations of *two* vectors.
- In the second format, we could say that
$$W = \mathrm{null}\left(\begin{pmatrix} 0 & 1 & -1 \end{pmatrix}\right).$$
This is equivalent to saying that
$$W = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \text{ such that } x_2 - x_3 = 0 \right\}.$$
Since $\dim(W) = 2$, we need *one* equation to cut out W.

[1] Equivalently, we could require that W is closed under addition and scalar multiplication.

2. Computing the column space and null space

We have the following facts:

(1) If A is an $n \times n$ invertible matrix, then $\mathrm{null}(A) = \{0\}$ and $\mathrm{col}(A) = \mathbb{R}^n$.

Proof. If $x \in \mathrm{null}(A)$, then $Ax = 0$, so $A^{-1}Ax = 0$, so $x = 0$. Therefore $\mathrm{null}(A)$ contains only the zero vector. Any vector b lies in $\mathrm{col}(A)$ because we can write $b = A(A^{-1}b)$. □

(2) Let A be an $n \times m$ matrix, and let B be an $m \times p$ matrix.

(i) We have $\mathrm{col}(AB) = \{Ax \text{ for } x \in \mathrm{col}(B)\} \subseteq \mathrm{col}(A)$.

(ii) If $\mathrm{col}(B) = \mathbb{R}^m$, then $\mathrm{col}(AB) = \mathrm{col}(A)$, so $\mathrm{rank}(AB) = \mathrm{rank}(A)$.

(iii) We have $\mathrm{null}(AB) = \{x \text{ such that } Bx \in \mathrm{null}(A)\} \supseteq \mathrm{null}(B)$.

(iv) If $\mathrm{null}(A) = \{0\}$, then $\mathrm{null}(AB) = \mathrm{null}(B)$, and $\mathrm{rank}(AB) = \mathrm{rank}(B)$.

(v) If B is invertible, then $\mathrm{null}(AB) = \{B^{-1}y \text{ for } y \in \mathrm{null}(A)\}$.

Proof. For (i), note that the set of vectors which AB can output is given by taking all the vectors that B can output and feeding them into A.

For (ii), use part (i), and then observe that $\{Ax \text{ for } x \in \mathbb{R}^m\}$ is $\mathrm{col}(A)$ by definition.

For (iii), note that the set of vectors killed by AB is given by taking all the vectors x such that Bx is killed by A.

Next we prove (iv). For the statement $\mathrm{null}(AB) = \mathrm{null}(B)$, use part (iii), and then observe that $\{x \text{ such that } Bx \in \{0\}\}$ is $\mathrm{null}(B)$ by definition.

For the statement $\mathrm{rank}(AB) = \mathrm{rank}(B)$, we claim that $A : \mathbb{R}^m \to \mathbb{R}^n$ (defined by $x \mapsto Ax$) is a one-to-one map. Indeed, if $Ax_1 = Ax_2$, then $A(x_1 - x_2) = 0$, so $(x_1 - x_2) \in \mathrm{null}(A) = \{0\}$, so $x_1 = x_2$. This one-to-one property implies that feeding an r-dimensional vector space through A yields another r-dimensional vector space. Hence, if $\mathrm{col}(B)$ has dimension r, then part (i) tells us that $\mathrm{col}(AB)$ has dimension r as well.

For (v), note that, for any subspace $W \subseteq \mathbb{R}^p$, we have

$$\{x \text{ such that } Bx \in W\} = \{B^{-1}y \text{ for } y \in W\}.$$

Now take $W = \mathrm{null}(A)$ and use part (iii). □

(3) Let Σ be a diagonal $n \times m$ matrix, whose first r entries are nonzero, and all other entries are zero.

$$\begin{aligned}\mathrm{col}(\Sigma) &= \{(y_1, \dots, y_r, 0, \dots, 0) \in \mathbb{R}^n \text{ for } y_1, \dots, y_r \in \mathbb{R}\}\\ \mathrm{null}(\Sigma) &= \{(0, \dots, 0, x_{r+1}, \dots, x_m) \in \mathbb{R}^m \text{ for } x_{r+1}, \dots, x_m \in \mathbb{R}\}\\ \mathrm{rank}(\Sigma) &= r.\end{aligned}$$

(4) Let A be an $n \times m$ matrix. Then $\mathrm{rank}(A) + \dim(\mathrm{null}(A)) = m$.

Proof. Let $A = U\Sigma V^\top$ be a full SVD.

Since $V^\top$ is invertible, (1) tells us that $\mathrm{col}(V^\top) = \mathbb{R}^m$. Then (2.ii) tells us that $\mathrm{rank}(U\Sigma V^\top) = \mathrm{rank}(U\Sigma)$. Next, since U is invertible, (1) tells us that $\mathrm{null}(U) = \{0\}$. Then (2.iv) tells us that $\mathrm{rank}(U\Sigma) = \mathrm{rank}(\Sigma)$. We conclude that $\mathrm{rank}(A) = \mathrm{rank}(\Sigma)$.

Similarly, since U is invertible, (1) tells us that $\mathrm{null}(U) = \{0\}$, so $\mathrm{null}(U\Sigma V^\top) = \mathrm{null}(\Sigma V^\top)$. Since $V^\top$ is invertible, (2.v) tells us that $\mathrm{null}(\Sigma V^\top)$ is related to $\mathrm{null}(\Sigma)$ by applying the invertible matrix $V = (V^\top)^{-1}$, which doesn't change the dimension. Therefore $\dim(\mathrm{null}(A)) = \dim(\mathrm{null}(\Sigma))$.

Now the desired result follows from (3) applied to Σ. □

(5) Let A be an invertible $n \times n$ matrix, and let A_1 be the $n \times r$ matrix built from the first r columns of A. Then

$$\begin{aligned} \operatorname{rank}(A_1) &= r \\ \operatorname{null}(A_1) &= \{0\} \\ \operatorname{col}(A_1^\top) &= \mathbb{R}^r \\ \operatorname{null}(A_1^\top) &= \operatorname{col}\begin{pmatrix} \text{The } n \times (n-r) \text{ matrix} \\ \text{built from the } \textit{last } n-r \\ \text{columns of } (A^\top)^{-1}. \end{pmatrix}. \end{aligned}$$

Proof. We have

$$A_1 = A \underbrace{\left(\frac{\mathrm{Id}_{r\times r}}{0_{(n-r)\times r}}\right)}_{B},$$

where A is multiplied by an $n \times r$ matrix B whose top $r \times r$ block is the identity matrix, and whose bottom $(n-r) \times r$ block is zero.

Since A is invertible, (1) tells us that $\operatorname{null}(A) = \{0\}$. Then (2.iv) implies that $\operatorname{rank}(AB) = \operatorname{rank}(B)$, which equals r by (3), since B is diagonal. Therefore $\operatorname{rank}(A_1) = r$.

Similarly, since A is invertible, (1) and (2.iv) imply that $\operatorname{null}(AB) = \operatorname{null}(B)$, which equals $\{0\}$ by (3). Therefore $\operatorname{null}(A_1) = \{0\}$.

Starting from $A_1 = AB$ and taking transposes, we find that $A_1^\top = B^\top A^\top$.

Since $A^\top$ is invertible, (1) tells us that $\operatorname{col}(A^\top) = \mathbb{R}^n$. Then (2.ii) implies that $\operatorname{col}(B^\top A^\top) = \operatorname{col}(B^\top)$, which equals $\mathbb{R}^r$ by (3). Therefore $\operatorname{col}(A_1^\top) = \mathbb{R}^r$.

Since $A^\top$ is invertible, (2.v) implies that

$$\operatorname{null}(B^\top A^\top) = \{(A^\top)^{-1} y \text{ for } y \in \operatorname{null}(B^\top)\}.$$

Point (3) tells us that

$$\operatorname{null}(B^\top) = \{(0, \ldots, 0, x_{r+1}, \ldots, x_n) \in \mathbb{R}^n \text{ for } x_{r+1}, \ldots, x_n \in \mathbb{R}^n\}.$$

Multiplying these vectors by $(A^\top)^{-1}$ gives all linear combinations of the *last* $n-r$ columns of $(A^\top)^{-1}$, and this shows the final statement of (5). □

(6) Suppose A is an $n \times m$ matrix, and $A = U_1 \Sigma_1 V_1^\top$ is a rank-r SVD. Then

$$\begin{aligned} \operatorname{rank}(A) &= r \\ \operatorname{col}(A) &= \operatorname{col}(U_1) \\ \operatorname{null}(A) &= \operatorname{col}(V_2) \\ \operatorname{row}(A) &= \operatorname{col}(V_1) \\ \operatorname{null}(A^\top) &= \operatorname{col}(U_2). \end{aligned}$$

Here U_2 and V_2 are matrices such that $U := \left(\begin{array}{c|c} U_1 & U_2 \end{array}\right)$ and $V := \left(\begin{array}{c|c} V_1 & V_2 \end{array}\right)$ are *square* orthogonal matrices.

Proof. By (5) applied to U, we know that $\operatorname{rank}(U_1) = r$ and $\operatorname{null}(U_1) = \{0\}$.

By (5) applied to V, we know that $\operatorname{col}(V_1^\top) = \mathbb{R}^r$ and $\operatorname{null}(V_1^\top) = \operatorname{col}(V_2)$. Indeed, V_2 is the $m \times (m-r)$ matrix built from the last $m-r$ columns of $(V^\top)^{-1} = V$, since V is orthogonal.

Since $\operatorname{col}(V_1^\top) = \mathbb{R}^r$, (2.ii) implies that $\operatorname{col}(U_1\Sigma_1V_1^\top) = \operatorname{col}(U_1\Sigma_1)$. Since Σ_1 is invertible, (1) implies that $\operatorname{col}(\Sigma_1) = \mathbb{R}^r$, so (2.ii) implies that $\operatorname{col}(U_1\Sigma_1) = \operatorname{col}(U_1)$. This shows that $\operatorname{col}(A) = \operatorname{col}(U_1)$. Since $\operatorname{rank}(U_1) = r$, we may also conclude $\operatorname{rank}(A) = r$.

Since $\mathrm{null}(U_1) = \{0\}$, (2.iv) implies that $\mathrm{null}(U_1\Sigma_1 V_1^\top) = \mathrm{null}(\Sigma_1 V_1^\top)$. Since Σ_1 is invertible, (1) implies that $\mathrm{null}(\Sigma_1) = \{0\}$, so (2.iv) implies that $\mathrm{null}(\Sigma_1 V_1^\top) = \mathrm{null}(V_1^\top)$. Now $\mathrm{null}(V_1^\top) = \mathrm{col}(V_2)$ implies that $\mathrm{null}(A) = \mathrm{col}(V_2)$.

The conclusions about $\mathrm{row}(A)$ and $\mathrm{null}(A^\top)$ follow from the already-proven statements by considering the rank-r SVD given by $A^\top = V_1\Sigma_1^\top U_1^\top$. Since Σ_1 is a square diagonal matrix, $\Sigma_1 = \Sigma_1^\top$. □

(7) Every rank r matrix can be expressed (nonuniquely) as the sum of r rank-one matrices.

Proof. See Lecture 9 slides. □

(8) Suppose A is an $n \times m$ matrix, and $A = QR$ where Q is orthogonal (and R is not necessarily invertible). Then

$$\mathrm{rank}(A) = \mathrm{rank}(R)$$
$$\mathrm{null}(A) = \mathrm{null}(R).$$

Proof. Since Q is orthogonal, $n \geq m$. We can find an $(n-m) \times n$ matrix Q_2 such that $\left(\, Q \mid Q_2 \, \right)$ is an $n \times n$ orthogonal matrix (see next section). Applying (5) to this square orthogonal matrix implies that $\mathrm{null}(Q) = \{0\}$. Then (2.iv) implies that $\mathrm{null}(QR) = \mathrm{null}(R)$ and $\mathrm{rank}(QR) = \mathrm{rank}(R)$, as desired. □

(9) Let U be an $n \times m$ matrix in *row echelon form*, [2] with r nonzero rows. Then

$$\mathrm{col}(U) = \{(x_1, \ldots, x_r, 0, \ldots, 0) \text{ for } x_1, \ldots, x_r \in \mathbb{R}\}$$
$$\mathrm{rank}(U) = r.$$

Proof. By doing column operations on U (including swaps), we can turn it into an $n \times m$ diagonal matrix Σ whose first r entries are nonzero, and all other entries of Σ are zero. Therefore we can write $U = \Sigma C$ where C is an $m \times m$ invertible matrix that keeps track of the column operations performed.

Since C is invertible, (1) implies that $\mathrm{col}(C) = \mathbb{R}^m$. Then (2.ii) implies that $\mathrm{col}(\Sigma C) = \mathrm{col}(\Sigma)$. The description of $\mathrm{col}(\Sigma)$ follows from (3). □

(10) Suppose A is an $n \times m$ matrix, and $A = LU$ where L is invertible and U is in row echelon form with r nonzero rows. Then

$$\mathrm{col}(A) = \mathrm{col}(L_1)$$
$$\mathrm{rank}(A) = r$$
$$\mathrm{null}(A) = \mathrm{null}(U).$$

Here L_1 is the $n \times r$ matrix built from the first r columns of L.

Proof. As observed in (9), we can write $U = \Sigma C$ where C is an $m \times m$ invertible matrix, and Σ is an $n \times m$ diagonal matrix whose first r entries are nonzero, and all other entries of Σ are zero. Then $A = L\Sigma C$.

Since C is invertible, (1) implies that $\mathrm{col}(C) = \mathbb{R}^m$. Then (2.ii) implies that $\mathrm{col}(L\Sigma C) = \mathrm{col}(L\Sigma)$. The matrix $L\Sigma$ is L_1 padded with some zeros on the right, so $\mathrm{col}(L\Sigma) = \mathrm{col}(L_1)$. We conclude that $\mathrm{col}(A) = \mathrm{col}(L_1)$.

Since L is invertible, (5) implies that $\mathrm{rank}(L_1) = r$. Therefore $\mathrm{rank}(A) = r$.

Since L is invertible, (1) implies that $\mathrm{null}(L) = \{0\}$. Then (2.iv) implies that $\mathrm{null}(LU) = \mathrm{null}(U)$. □

[2] See `https://en.wikipedia.org/wiki/Row_echelon_form`. The nonzero rows of U must come before the zero rows, and the number of zeros at the beginning of each nonzero row must be strictly increasing. This means that U looks like a staircase.

(11) Let A be an $n \times m$ matrix. Then

$$\text{null}(A^\top A) = \text{null}(A)$$
$$\text{rank}(A^\top A) = \text{rank}(A).$$

Proof. By (2.iii), we have $\text{null}(A^\top A) \supseteq \text{null}(A)$. To show the reverse inclusion, suppose that $x \in \text{null}(A^\top A)$. Then $A^\top Ax = 0$, so $(Ax)^\top Ax = 0$. This says that the vector Ax has length zero, so $Ax = 0$, so $x \in \text{null}(A)$, as desired.

By (4), we have

$$\text{rank}(A) = m - \dim(\text{null}(A))$$
$$\text{rank}(A^\top A) = m - \dim(\text{null}(A^\top A)).$$

Since we know the nullspaces of A and $A^\top A$ agree, so do the ranks. □

(12) Let P be an $n \times n$ matrix satisfying $P^2 = P$. Then

$$\text{col}(P) = \text{null}(P - \text{Id}_{n\times n}).$$

Proof. If $x \in \text{null}(P - \text{Id}_{n\times n})$, then $Px = x$, so $x \in \text{col}(P)$. Conversely, if $x \in \text{col}(P)$, we have $x = Py$ for some y. Applying P, we find that $Px = Py$, so we can write $x = Px$. Therefore $x \in \text{null}(P - \text{Id}_{n\times n})$, as desired. □

(13) Let Q be an $n \times m$ orthogonal matrix. Then

$$\text{col}(Q) = \text{col}(QQ^\top) = \text{null}(QQ^\top - \text{Id}_{n\times n}).$$

Proof. As in (8), we can find an $(n-m) \times n$ matrix Q_2 such that $\left(\begin{array}{c|c} Q & Q_2 \end{array} \right)$ is an $n \times n$ orthogonal matrix. Applying (5) to this square orthogonal matrix implies that $\text{col}(Q^\top) = \mathbb{R}^m$. Then (2.ii) implies that $\text{col}(Q) = \text{col}(QQ^\top)$.

The second equality follows from (2). Indeed, we may take $P = QQ^\top$, because $(QQ^\top)^2 = QQ^\top QQ^\top = QQ^\top$ since Q is orthogonal. □

(14) Let $A = QR$ where Q is an $n \times m$ orthogonal matrix, and $\text{col}(R) = \mathbb{R}^m$. Then

$$\text{col}(A) = \text{null}(QQ^\top - \text{Id}_{n\times n}).$$

Proof. Since $\text{col}(R) = \mathbb{R}^m$, (2.ii) implies that $\text{col}(QR) = \text{col}(Q)$. This equals the RHS by (13). □

(15) Let $A = U_1\Sigma_1V_1^\top$ be a rank-r SVD. Then

$$\text{col}(A) = \text{null}(U_1U_1^\top - \text{Id}_{n\times n}).$$

Proof. By (6), $\text{col}(A) = \text{col}(U_1)$. Now apply (13) to the orthogonal matrix U_1. □

3. Completing an orthogonal matrix to a square orthogonal matrix

In the proof of (8), we wanted to complete an orthogonal $n \times m$ matrix Q to a square orthogonal matrix $Q' := \left(\, Q \mid Q_2 \, \right)$ by finding some $(n-m) \times m$ matrix Q_2. Equivalently, if $v_1, \ldots, v_m$ are the columns of Q, which form an orthonormal collection, we want to find additional vectors $v_{m+1}, \ldots, v_n$ (which will be the columns of Q_2) such that $v_1, \ldots, v_m, v_{m+1}, \ldots, v_n$ is an orthonormal collection.

We explain how to do this inductively. Assume that $v_1, \ldots, v_s$ is an orthonormal collection of vectors, and let us attempt to find v_{s+1} such that $v_1, \ldots, v_s, v_{s+1}$ is still orthonormal. Take any vector $x \in \mathbb{R}^n$ which is not a linear combination of the $v_1, \ldots, v_s$. (Assuming that $s < n$, this is always possible.) Then, define a new vector

$$y = x - (x \cdot v_1)v_1 - (x \cdot v_2)v_2 \cdots - (x \cdot v_s)v_s.$$

By dotting with $v_1, \ldots, v_s$, we see that y is orthogonal to all of those vectors. Indeed,

$$\begin{aligned} y \cdot v_i &= x \cdot v_i - (x \cdot v_i)(v_i \cdot v_i) && (\text{since } v_j \cdot v_i = 0 \text{ when } j \neq i) \\ &= x \cdot v_i - x \cdot v_i && (\text{since } v_i \cdot v_i = 1) \\ &= 0. \end{aligned}$$

Since x is not a linear combination of the $v_1, \ldots, v_s$, we know that y is nonzero. Therefore, we may define a new vector

$$z = \frac{1}{\|y\|} y.$$

This vector is orthogonal to the $v_1, \ldots, v_s$, and it has length 1. We may now take $v_{s+1} := z$ and complete the inductive step.

This is a version of the Gram–Schmidt process.