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 it is closed under addition and scalar multiplication.

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.

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 $\mathrm{rank}(AB) = \mathrm{rank}(B)$.

(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 $\mathrm{rank}(AB) = \mathrm{rank}(A)$.

(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), note 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.

For (iv), note that $\{x \text{ such that } Bx \in \{0\}\}$ is $\mathrm{null}(B)$ by definition. For the statement about rank, notice 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 $x_1 = x_2$. Therefore, feeding an r-dimensional vector space through A yields another r-dimensional vector space. In particular, if $\mathrm{col}(B)$ has dimension r, then (i) tells 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 $W = \mathrm{null}(A)$. □

(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. By (2.ii) and (2.iv), we know that $\mathrm{rank}(A) = \mathrm{rank}(\Sigma)$. By (2.iv) and (2.v), we also know that $\dim(\mathrm{null}(A)) = \dim(\mathrm{null}(\Sigma))$. Now use (3). □

(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

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

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. By (2.iv) and (3), we have

$$\mathrm{rank}(A_1) = \mathrm{rank}(B) = r$$

as well as

$$\mathrm{null}(A_1) = \mathrm{null}(B) = \{0\}.$$

We also have $A_1^\top = B^\top A^\top$, and $A^\top$ is invertible. Therefore, by (2.ii) and (3), we have

$$\mathrm{col}(A_1^\top) = \mathrm{col}(B^\top) = \mathbb{R}^r.$$

By (2.v) and (3), we have

$$\mathrm{null}(A_1^\top) = \{(A^\top)^{-1} \text{ for } y \in \mathrm{null}(B^\top)\},$$

and $\mathrm{null}(B^\top) = \{(0, \ldots, 0, x_{r+1}, \ldots, x_n) \in \mathbb{R}^n \text{ for } x_{r+1}, \ldots, x_n \in \mathbb{R}^n\}$. Multiplying such a vector by the matrix $(A^\top)^{-1}$ 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

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

Here U_2 and V_2 are matrices such that $U := \left(U_1 \mid U_2 \right)$ and $V := \left(V_1 \mid V_2 \right)$ are *square* orthogonal matrices.

Proof. By (2.iv) and (3), we have

$$\mathrm{rank}(A) = \mathrm{rank}(\Sigma_1) = r.$$

By (5), we have $\mathrm{col}(V_1^\top) = \mathbb{R}^r$. Therefore, by (2.ii), we have

$$\mathrm{col}(A) = \mathrm{col}(U_1).$$

By (5), we have $\mathrm{null}(U_1) = \{0\}$. Therefore, by (2.iv), we have

$$\mathrm{null}(A) = \mathrm{null}(V_1^\top).$$

Using (5) on the matrix V, and noting that $(V^\top)^{-1} = V$ since V is orthogonal, we conclude that

$$\mathrm{null}(V_1^\top) = \mathrm{col}(V_2),$$

since V_2 is the matrix built from the last $n-r$ columns of V.

The remaining conclusions in (5) follow by applying the preceding results to the SVD given by $A^\top = V_1 \Sigma_1^\top U_1^\top$. □

(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

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

Proof. Since Q is orthogonal, $n \geq m$. 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 (see next section). Then (5) implies that $\operatorname{null}(Q) = \{0\}$. Now the desired statements follow from (2.iv). □

(9) Let U be an $n \times m$ matrix in *row echelon form*, [1] with r nonzero rows. Then

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

Proof. By doing column operations on U (including swaps), we can turn it into an $n \times m$ diagonal matrix Σ with r nonzero entries. Therefore we can write $U = \Sigma C$ where C is an $m \times m$ invertible matrix that keeps track of the column operations performed. Now (2.ii) yields the desired statement. □

(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

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

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

Proof. This follows from (9), (2.i), and (2.iv). □

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

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

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

The statement about rank now follows from (4). □

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

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

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

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

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

[1] 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.

Proof. The first equality follows from (5) and (2.ii), using the same reasoning as in (8), of completing Q to a square orthogonal matrix. The second equality follows from (12). Indeed, we may take $P = QQ^\top$, because $(QQ^\top)^2 = QQ^\top QQ^\top = QQ^\top$ since Q is orthogonal. $\square$

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

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

Proof. From (2.ii), we know that $\mathrm{col}(A) = \mathrm{col}(Q)$. This equals the RHS by (13). $\square$

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

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

Proof. By (5), $\mathrm{col}(V_1^\top) = \mathbb{R}^r$, so $\mathrm{col}(\Sigma_1V_1^\top) = \mathbb{R}^r$ since Σ_1 is invertible. Now (14) yields the desired statement (taking Q to be U_1 and taking R to be $\Sigma_1V_1^\top$). $\square$

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 $Q' := \left(\begin{array}{c|c} Q & Q_2 \end{array} \right)$ 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.