Linear Transformations

Carlo Tomasi

August 24, 2017

When not given in the main text, proofs are in Appendix A.

1 Matrices and Vectors

A (real) *matrix* of *size* $m \times n$ is an array of mn real numbers arranged in m rows and n columns:

$$A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} .$$

The $n \times m$ matrix A^T obtained by exchanging rows and columns of A is called the *transpose* of A. A matrix A is said to be *symmetric* if $A = A^T$.

The *sum* of two matrices of equal size is the matrix of the entry-by-entry sums, and the *scalar product* of a real number a and an $m \times n$ matrix A is the $m \times n$ matrix of all the entries of A, each multiplied by a. The *difference* of two matrices of equal size A and B is

$$A - B = A + (-1)B .$$

The *product* of an $m \times p$ matrix A and a $p \times n$ matrix B is an $m \times n$ matrix C with entries

$$c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj} .$$

The matrix

$$\begin{bmatrix} v_1 \\ \vdots \\ v_n \end{bmatrix}$$

is called a *column vector*, and the matrix

$$\begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$$

is called a *row vector*. Column vectors are denoted by lowercase bold symbols, say $\mathbf{a}$. The corresponding row vector (that is, the row vector with the same entries in the same order) is $\mathbf{a}^T$.

All algebraic operations on vectors are inherited from the corresponding matrix operations, when defined. In addition, the *inner product* of two n-dimensional vectors

$$\mathbf{a} = (a_1, \ldots, a_n) \quad \text{and} \quad \mathbf{b} = (b_1, \ldots, b_n)$$

is the real number equal to the matrix product $\mathbf{a}^T\mathbf{b}$. It is easy to verify that this is also equal to $\mathbf{b}^T\mathbf{a}$. Two vectors that have a zero inner product are said to be *orthogonal*.

The *norm* of a vector $\mathbf{a}$ is

$$\|\mathbf{a}\| = \sqrt{\mathbf{a}^T\mathbf{a}}\ .$$

A *unit vector* is a vector with norm one.

The *outer product* of an m dimensional vector $\mathbf{a}$ with an n-dimensional vector $\mathbf{b}$ is the $m \times n$ matrix $\mathbf{a}\mathbf{b}^T$.

2 Vector Spaces

Given n vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and n real numbers $x_1, \ldots, x_n$, the vector

$$\mathbf{b} = \sum_{j=1}^{n} x_j \mathbf{a}_j \tag{1}$$

is said to be a *linear combination* of $\mathbf{a}_1, \ldots, \mathbf{a}_n$ with coefficients $x_1, \ldots, x_n$.

The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are *linearly dependent* if they admit the null vector as a nonzero linear combination. In other words, they are linearly dependent if there is a set of coefficients $x_1, \ldots, x_n$, not all of which are zero, such that

$$\sum_{j=1}^{n} x_j \mathbf{a}_j = \mathbf{0}\ . \tag{2}$$

For later reference, it is useful to rewrite the last two equalities in a different form. Equation (1) is the same as

$$A\mathbf{x} = \mathbf{b} \tag{3}$$

and equation (2) is the same as

$$A\mathbf{x} = \mathbf{0} \tag{4}$$

where

$$A = \begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_n \end{bmatrix}\ , \qquad \mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\ , \qquad \mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}\ .$$

If you are not convinced of these equivalences, take the time to write out the components of each expression for a small example. This is important. Make sure that you are comfortable with this.

Thus, the columns of a matrix A are dependent if there is a nonzero solution to the homogeneous system (4). Vectors that are not dependent are *independent*.

Theorem 2.1. *The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent iff*[1] *at least one of them is a linear combination of the others.*

Even more specifically:

Corollary 2.2. *If n nonzero vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent then at least one of them is a linear combination of the ones that* precede *it.*

[1] "iff" means "if and only if."

A set $\mathbf{a}_1, \ldots, \mathbf{a}_n$ is said to be a *basis* for a set B of vectors if the $\mathbf{a}_j$ are linearly independent and every vector in B can be written as a linear combination of them. B is said to be a *vector space* if it contains *all* the linear combinations of its basis vectors. In particular, this implies that every linear space contains the zero vector. The basis vectors are said to *span* the vector space.

Theorem 2.3. *Given a vector* $\mathbf{b}$ *in the vector space* B *and a basis* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *for* B*, the coefficients* $x_1, \ldots, x_n$ *such that*

$$\mathbf{b} = \sum_{j=1}^{n} x_j \mathbf{a}_j$$

are uniquely determined.

The previous theorem is a very important result. An equivalent formulation is the following:

> *If the columns* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *of* A *are linearly independent and the system* $A\mathbf{x} = \mathbf{b}$ *admits a solution, then the solution is unique.*

Pause for a minute to verify that this formulation is equivalent.

Theorem 2.4. *Two different bases for the same vector space* B *have the same number of vectors.*

A consequence of this theorem is that any basis for $\mathbf{R}^m$ has m vectors. In fact, the basis of *elementary vectors*

$$\mathbf{e}_j = j\text{th column of the } m \times m \text{ identity matrix}$$

is clearly a basis for $\mathbf{R}^m$, since any vector

$$\mathbf{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_m \end{bmatrix}$$

can be written as

$$\mathbf{b} = \sum_{j=1}^{m} b_j \mathbf{e}_j$$

and the $\mathbf{e}_j$ are clearly independent. Since this elementary basis has m vectors, theorem 2.4 implies that any other basis for $\mathbf{R}^m$ has m vectors.

Another consequence of theorem 2.4 is that n vectors of dimension $m < n$ are bound to be dependent, since any basis for $\mathbf{R}^m$ can only have m vectors.

Since all bases for a space have the same number of vectors, it makes sense to define the *dimension* of a space as the number of vectors in any of its bases.

3 Linear Transformations

Linear transformations map spaces into spaces. It is important to understand exactly what is being mapped into what in order to determine whether a linear system has solutions, and if so how many. This section introduces the notion of orthogonality between spaces, defines the null space and range of a matrix, and its rank. In the process, we also introduce a useful procedure (Gram-Schmidt) for orthonormalizing a set of linearly independent vectors.

Two vector spaces A and B are said to be *orthogonal* to one another when every vector in A is orthogonal to every vector in B. If vector space A is a subspace of $\mathbf{R}^m$ for some m, then the *orthogonal complement* of A is the set of all vectors in $\mathbf{R}^m$ that are orthogonal to all the vectors in A.

Notice that complement and orthogonal complement are very different notions. For instance, the complement of the xy plane in $\mathbf{R}^3$ is all of $\mathbf{R}^3$ except the xy plane, while the orthogonal complement of the xy plane is the z axis.

Theorem 3.1. *Any basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ for a subspace A of $\mathbf{R}^m$ can be extended into a basis for $\mathbf{R}^m$ by adding $m - n$ vectors $\mathbf{a}_{n+1}, \ldots, \mathbf{a}_m$.*

The following is called the *Gram-Schmidt* theorem.

Theorem 3.2. *Given n vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, the following construction*

$$
\begin{array}{l}
r = 0 \\
\textit{for } j = 1 \textit{ to } n \\
\quad \mathbf{a}'_j = \mathbf{a}_j - \sum_{l=1}^{r} (\mathbf{q}_l^T \mathbf{a}_j) \mathbf{q}_l \\
\quad \textit{if } \|\mathbf{a}'_j\| \neq \mathbf{0} \\
\quad\quad r = r + 1 \\
\quad\quad \mathbf{q}_r = \frac{\mathbf{a}'_j}{\|\mathbf{a}'_j\|} \\
\quad \textit{end} \\
\textit{end}
\end{array}
$$

yields a set of orthonormal[2] *vectors $\mathbf{q}_1 \ldots, \mathbf{q}_r$ that span the same space as $\mathbf{a}_1, \ldots, \mathbf{a}_n$.*

The Gram-Schmidt theorem is a useful procedure in its own right. It also leads to a simple proof for the following result.

Theorem 3.3. *If A is a subspace of $\mathbf{R}^m$ and $A^\perp$ is the orthogonal complement of A in $\mathbf{R}^m$, then*

$$\dim(A) + \dim(A^\perp) = m \; .$$

We can now start to talk about matrices in terms of the subspaces associated with them. The *null space* null(A) of an $m \times n$ matrix A is the space of all n-dimensional vectors that are orthogonal to the rows of A. The *range* of A is the space of all m-dimensional vectors that are generated by the columns of A. Thus, $\mathbf{x} \in$ null(A) iff $A\mathbf{x} = 0$, and $\mathbf{b} \in$ range(A) iff $A\mathbf{x} = \mathbf{b}$ for some $\mathbf{x}$. This can be restated into the following immediate but very important statement:

Theorem 3.4. *The matrix A transforms a vector $\mathbf{x}$ in its null space into the zero vector, and an arbitrary vector $\mathbf{x}$ into a vector in* range(A).

The spaces orthogonal to null(A) and range(A) occur frequently enough to deserve names of their own. The space range$(A)^\perp$ is called the *left nullspace* of the matrix, and null$(A)^\perp$ is called the *rowspace* of A.

[2] Orthonormal means orthogonal and with unit norm.

A frequently used synonym for "range" is *column space*. It should be obvious from the meaning of these spaces that

$$\begin{aligned} \text{null}(A)^\perp &= \text{range}(A^T) \\ \text{range}(A)^\perp &= \text{null}(A^T) \end{aligned}$$

where A^T is the *transpose* of A, defined as the matrix obtained by exchanging the rows of A with its columns.

In summary, four spaces are associated with an $m \times n$ matrix A:

range(A);
null(A);
range$(A)^\perp =$ leftnull(A);
null$(A)^\perp =$ rowspace(A) .

In order to count solutions to a linear system, it is important to establish how the dimensions of these spaces relate to each other. From theorem 3.3, if null(A) has dimension h, then the space generated by the rows of A has dimension $r = n - h$, that is, A has $n - h$ linearly independent rows. It is not obvious that the space generated by the *columns* of A has also dimension $r = n - h$. Even more strongly, the following theorem holds:

Theorem 3.5. *The matrix A establishes a one-to-one mapping between* rowspace(A) *and* range(A).

Thus, the two linear vector spaces rowspace(A) and range(A) are isomorphic to each other, and therefore have equal dimension. In summary, if we define

$r = \dim(\text{range}(A))$
$h = \dim(\text{null}(A))$

then theorems 3.3 and 3.5 yield the following:

$\dim(\text{leftnull}(A)) = \dim(\text{range}(A)^\perp) = m - r$
$\dim(\text{rowspace}(A)) = \dim(\text{null}(A)^\perp) = n - h = r$.

This also implies the following result:

Corollary 3.6. *The number r of linearly independent columns of any $m \times n$ matrix A is equal to the number of its independent rows.*

As a result, we can define the *rank* of A to be equivalently the number of linearly independent columns or of linearly independent rows of A:

$$r = \text{rank}(A) = \dim(\text{range}(A)) = n - \dim(\text{null}(A)) = n - h \; .$$

Note that if $A\mathbf{x} = \mathbf{b}$, then for any vector $\mathbf{y} \in$ null(A) we also have $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y} = A\mathbf{x}$ because $A\mathbf{y} = 0$. Therefore, the matrix A maps vectors in R^n that differ only by a vector in null(A) to the same point. Since rowspace(A) is isomorphic to range(A), it is then convenient to take each point $\mathbf{x}_r$ of rowspace(A) as a representative of the *affine space*

$$\mathcal{A}(\mathbf{x}_r) = \mathbf{x}_r + \text{null}(A)$$

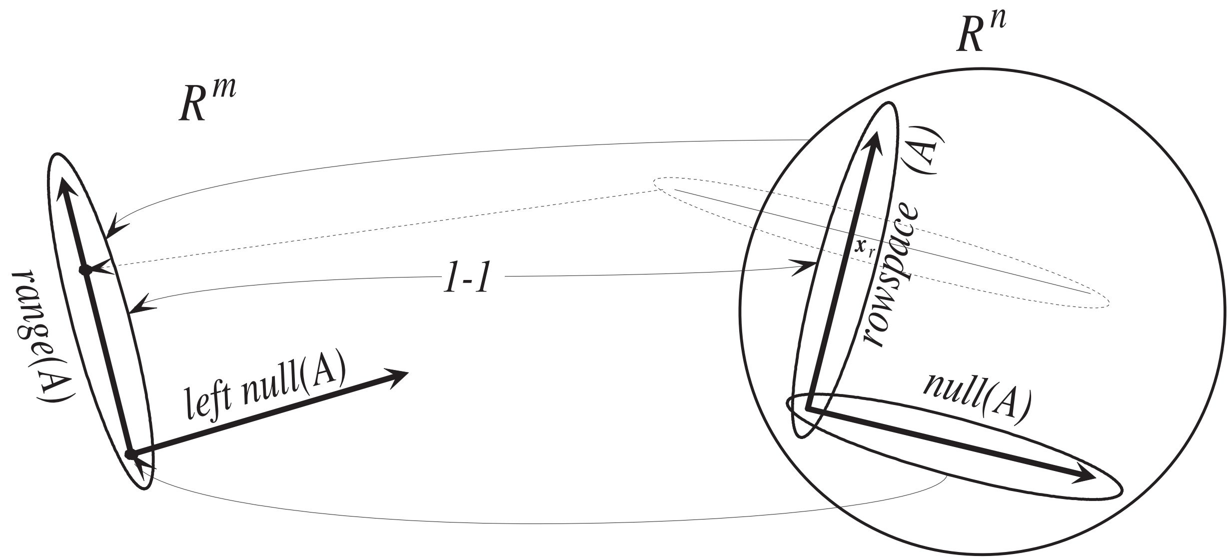

Figure 1: An $m \times n$ matrix A maps all of R^n to range(A) (top arrow), and null(A) to zero (bottom arrow). The row space and range of A are isomorphic to each other (*i.e.*, in 1-1 correspondence), and for each point $\mathbf{x}_r \in$ rowspace(A) there is an affine space $\mathbf{x}_r +$ null(A) of dimension $h = \dim(\text{null}(A)) = n -$ rank(A) that maps (dotted arrow) to the single point $A\mathbf{x}_r$.

of points that all map to the single point $A\mathbf{x}_r$. The sum in the expression above means that the single vector $\mathbf{x}_r$ is added to every vector of the linear space null(A) to produce the affine space $\mathcal{A}(\mathbf{x}_r)$.

The foregoing discussion allows forming the picture of a linear mapping shown in figure 1.

As a brief aside, the picture of the isomorphism between the two linear spaces rowspace(A) and range(A) can be made stronger by observing that A also transforms any basis for rowspace(A) into a basis for range(A). This is not immediately obvious, since if $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are a basis for rowspace(A) then $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ might conceivably be dependent, or fail to span all of range(A). However, this is not so:

Theorem 3.7. *If the vectors* $\mathbf{v}_1, \ldots, \mathbf{v}_r$ *are a basis for* rowspace(A)*, then the vectors* $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ *are a basis for* range(A).

A Proofs

Theorem 2.1

The vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent iff[3] *at least one of them is a linear combination of the others.*

Proof. In one direction, dependency means that there is a nonzero vector $\mathbf{x}$ such that

$$\sum_{j=1}^{n} x_j \mathbf{a}_j = \mathbf{0} \ .$$

Let x_k be nonzero for some k. We have

$$\sum_{j=1}^{n} x_j \mathbf{a}_j = x_k \mathbf{a}_k + \sum_{j=1,\, j\neq k}^{n} x_j \mathbf{a}_j = \mathbf{0}$$

so that

$$\mathbf{a}_k = - \sum_{j=1,\, j\neq k}^{n} \frac{x_j}{x_k} \mathbf{a}_j \tag{5}$$

as desired. The converse is proven similarly: if

$$\mathbf{a}_k = \sum_{j=1,\, j\neq k}^{n} x_j \mathbf{a}_j$$

for some k, then

$$\sum_{j=1}^{n} x_j \mathbf{a}_j = \mathbf{0}$$

by letting $x_k = -1$ (so that $\mathbf{x}$ is nonzero).

Corollary 2.2

If n nonzero vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$ are linearly dependent then at least one of them is a linear combination of the ones that precede *it.*

Proof. Let k be the *last* of the nonzero x_j in the proof of theorem 2.1. Then $x_j = 0$ for $j > k$ in (5), which then becomes

$$\mathbf{a}_k = \sum_{j<k}^{n} \frac{x_j}{x_k} \mathbf{a}_j$$

as desired.

[3] "iff" means "if and only if."

Theorem 2.3

Given a vector $\mathbf{b}$ *in the vector space* B *and a basis* $\mathbf{a}_1, \ldots, \mathbf{a}_n$ *for* B*, the coefficients* $x_1, \ldots, x_n$ *such that*

$$\mathbf{b} = \sum_{j=1}^{n} x_j \mathbf{a}_j$$

are uniquely determined.

Proof. Let

$$\mathbf{b} = \sum_{j=1}^{n} x'_j \mathbf{a}_j \ .$$

Then,

$$\mathbf{0} = \mathbf{b} - \mathbf{b} = \sum_{j=1}^{n} x_j \mathbf{a}_j - \sum_{j=1}^{n} x'_j \mathbf{a}_j = \sum_{j=1}^{n} (x_j - x'_j) \mathbf{a}_j$$

but because the $\mathbf{a}_j$ are linearly independent, this is possible only when $x_j - x'_j = 0$ for every j.

Theorem 2.4

Two different bases for the same vector space B *have the same number of vectors.*

Proof. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ and $\mathbf{a}'_1, \ldots, \mathbf{a}'_{n'}$ be two different bases for B. Then each $\mathbf{a}'_j$ is in B (why?), and can therefore be written as a linear combination of $\mathbf{a}_1, \ldots, \mathbf{a}_n$. Consequently, the vectors of the set

$$G = \mathbf{a}'_1, \mathbf{a}_1, \ldots, \mathbf{a}_n$$

must be linearly dependent. We call a set of vectors that contains a basis for B a *generating set* for B. Thus, G is a generating set for B.

The rest of the proof now proceeds as follows: we keep removing $\mathbf{a}$ vectors from G and replacing them with $\mathbf{a}'$ vectors in such a way as to keep G a generating set for B. Then we show that we cannot run out of $\mathbf{a}$ vectors before we run out of $\mathbf{a}'$ vectors, which proves that $n \geq n'$. We then switch the roles of $\mathbf{a}$ and $\mathbf{a}'$ vectors to conclude that $n' \geq n$. This proves that $n = n'$.

From corollary 2.2, one of the vectors in G is a linear combination of those preceding it. This vector cannot be $\mathbf{a}'_1$, since it has no other vectors preceding it. So it must be one of the $\mathbf{a}_j$ vectors. Removing the latter keeps G a generating set, since the removed vector depends on the others. Now we can add $\mathbf{a}'_2$ to G, writing it right after $\mathbf{a}'_1$:

$$G = \mathbf{a}'_1, \mathbf{a}'_2, \ldots \ .$$

G is still a generating set for B.

Let us continue this procedure until we run out of either $\mathbf{a}$ vectors to remove or $\mathbf{a}'$ vectors to add. The $\mathbf{a}$ vectors cannot run out first. Suppose in fact *per absurdum* that G is now made only of $\mathbf{a}'$ vectors, and that there are still left-over $\mathbf{a}'$ vectors that have not been put into G. Since the $\mathbf{a}'$s form a basis, they are mutually linearly independent. Since B is a vector space, all the $\mathbf{a}'$s are in B. But then G cannot be a generating set, since the vectors in it cannot generate the left-over $\mathbf{a}'$s, which are independent of those in G. This is absurd, because at every step we have made sure that G remains a generating set. Consequently, we must run out of $\mathbf{a}'$s first (or simultaneously with the last $\mathbf{a}$). That is, $n \geq n'$.

Now we can repeat the whole procedure with the roles of $\mathbf{a}$ vectors and $\mathbf{a}'$ vectors exchanged. This shows that $n' \geq n$, and the two results together imply that $n = n'$.

Theorem 3.1

Any basis $\mathbf{a}_1, \ldots, \mathbf{a}_n$ for a subspace A of $\mathbf{R}^m$ can be extended into a basis for $\mathbf{R}^m$ by adding $m-n$ vectors $\mathbf{a}_{n+1}, \ldots, \mathbf{a}_m$.

Proof. If $n = m$ we are done. If $n < m$, the given basis cannot generate all of $\mathbf{R}^m$, so there must be a vector, call it $\mathbf{a}_{n+1}$, that is linearly independent of $\mathbf{a}_1, \ldots, \mathbf{a}_n$. This argument can be repeated until the basis spans all of $\mathbf{R}^m$, that is, until $m = n$.

Theorem 3.2

Given n vectors $\mathbf{a}_1, \ldots, \mathbf{a}_n$, the following construction

$$
\begin{array}{l}
r = 0 \\
\textit{for } j = 1 \textit{ to } n \\
\quad \mathbf{a}'_j = \mathbf{a}_j - \sum_{l=1}^{r} (\mathbf{q}_l^T \mathbf{a}_j) \mathbf{q}_l \\
\quad \textit{if } \|\mathbf{a}'_j\| \neq \mathbf{0} \\
\quad\quad r = r + 1 \\
\quad\quad \mathbf{q}_r = \frac{\mathbf{a}'_j}{\|\mathbf{a}'_j\|} \\
\quad \textit{end} \\
\textit{end}
\end{array}
$$

yields a set of orthonormal[4] *vectors $\mathbf{q}_1 \ldots, \mathbf{q}_r$ that span the same space as $\mathbf{a}_1, \ldots, \mathbf{a}_n$.*

Proof. We first prove by induction on r that the vectors $\mathbf{q}_r$ are mutually orthonormal. If $r = 1$, there is little to prove. The normalization in the above procedure ensures that $\mathbf{q}_1$ has unit norm. Let us now assume that the procedure above has been performed a number $j-1$ of times sufficient to find $r-1$ vectors $\mathbf{q}_1, \ldots, \mathbf{q}_{r-1}$, and that these vectors are orthonormal (the inductive assumption). Then for any $i < r$ we have

$$\mathbf{q}_i^T \mathbf{a}'_j = \mathbf{q}_i^T \mathbf{a}_j - \sum_{l=1}^{r-1} (\mathbf{q}_l^T \mathbf{a}_j) \mathbf{q}_i^T \mathbf{q}_l = 0$$

because the term $\mathbf{q}_i^T \mathbf{a}_j$ cancels the i-th term $(\mathbf{q}_i^T \mathbf{a}_j)\mathbf{q}_i^T \mathbf{q}_i$ of the sum (remember that $\mathbf{q}_i^T \mathbf{q}_i = 1$), and the remaining inner products of the form $\mathbf{q}_i^T \mathbf{q}_l$ are zero by the inductive assumption. Because of the explicit normalization step $\mathbf{q}_r = \mathbf{a}'_j / \|\mathbf{a}'_j\|$, the vector $\mathbf{q}_r$, if computed, has unit norm, and because $\mathbf{q}_i^T \mathbf{a}'_j = 0$, it follows that $\mathbf{q}_r$ is orthogonal to all its predecessors, $\mathbf{q}_i^T \mathbf{q}_r = 0$ for $i = 1, \ldots, r-1$.

Finally, we notice that the vectors $\mathbf{q}_j$ span the same space as the $\mathbf{a}_j$s, because the former are linear combinations of the latter, are orthonormal (and therefore independent), and equal in number to the number of linearly independent vectors in $\mathbf{a}_1, \ldots, \mathbf{a}_n$.

Theorem 3.3

If A is a subspace of $\mathbf{R}^m$ and $A^\perp$ is the orthogonal complement of A in $\mathbf{R}^m$, then

$$\dim(A) + \dim(A^\perp) = m \; .$$

[4]Orthonormal means orthogonal and with unit norm.

Proof. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n$ be a basis for A. Extend this basis to a basis $\mathbf{a}_1, \ldots, \mathbf{a}_m$ for $\mathbf{R}^m$ (theorem 3.1). Orthonormalize this basis by the Gram-Schmidt procedure (theorem 3.2) to obtain $\mathbf{q}_1, \ldots, \mathbf{q}_m$. By construction, $\mathbf{q}_1, \ldots, \mathbf{q}_n$ span A. Because the new basis is orthonormal, all vectors generated by $\mathbf{q}_{n+1}, \ldots, \mathbf{q}_m$ are orthogonal to all vectors generated by $\mathbf{q}_1, \ldots, \mathbf{q}_n$, so there is a space of dimension at least $m-n$ that is orthogonal to A. On the other hand, the dimension of this orthogonal space cannot exceed $m-n$, because otherwise we would have more than m vectors in a basis for $\mathbf{R}^m$. Thus, the dimension of the orthogonal space $A^\perp$ is exactly $m-n$, as promised.

Theorem 3.5

The matrix A establishes a one-to-one mapping between rowspace(A) *and* range(A).

Proof. This statement means that A maps different elements $\mathbf{x} \in$ rowspace(A) into different elements $\mathbf{b} = A\mathbf{x} \in$ range(A). Let then $\mathbf{r}_1$ and $\mathbf{r}_2$ be two different vectors in rowspace(A). We need to show that $A\mathbf{r}_1$ and $A\mathbf{r}_2$ are different as well.

Since $\mathbf{r}_1$ and $\mathbf{r}_2$ are different linear combinations of the vectors in any given basis for the row space, their difference $\mathbf{d} = \mathbf{r}_1 - \mathbf{r}_2$ is a nonzero linear combination of the basis vectors of the row space. As a consequence, $\mathbf{d}$ is nonzero and orthogonal to all vectors in null(A), and therefore $A\mathbf{d}$ is nonzero. Then,

$$0 \neq A\mathbf{d} = A(\mathbf{r}_1 - \mathbf{r}_2) = A\mathbf{r}_1 - A\mathbf{r}_2 \ ,$$

so that

$$A\mathbf{r}_1 \neq A\mathbf{r}_2 \ .$$

Theorem 3.7

If the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_r$ are a basis for rowspace(A)*, then the vectors $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ are a basis for* range(A).

Proof. First, the r vectors $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ generate the range of A. In fact, given an arbitrary vector $\mathbf{b} \in$ range(A), there must be a linear combination of the columns of A that is equal to $\mathbf{b}$. In symbols, there is an n-tuple $\mathbf{x}$ such that $A\mathbf{x} = \mathbf{b}$. Let $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ be a basis for null(A), so that $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis for $\mathbf{R}^n$. Then,

$$\mathbf{x} = \sum_{j=1}^{n} c_j \mathbf{v}_j \ .$$

Thus,

$$\mathbf{b} = A\mathbf{x} = A\sum_{j=1}^{n} c_j \mathbf{v}_j = \sum_{j=1}^{n} c_j A\mathbf{v}_j = \sum_{j=1}^{r} c_j A\mathbf{v}_j$$

since $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ span null(A), so that $A\mathbf{v}_j = 0$ for $j = r+1, \ldots, n$. This proves that the r vectors $A\mathbf{v}_1, \ldots, A\mathbf{v}_r$ generate range(A).

Second, we prove that these vectors are linearly independent. Suppose, *per absurdum*, that they are not. Then there exist numbers $x_1, \ldots, x_r$, not all zero, such that

$$\sum_{j=1}^{r} x_j A\mathbf{v}_j = 0$$

so that

$$A \sum_{j=1}^{r} x_j \mathbf{v}_j = 0 \ .$$

But then the vector $\sum_{j=1}^{r} x_j \mathbf{v}_j$ is in the null space of A. Since the vectors $\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n$ are a basis for null(A), there must exist coefficients $x_{r+1}, \ldots, x_n$ such that

$$\sum_{j=1}^{r} x_j \mathbf{v}_j = \sum_{j=r+1}^{n} x_j \mathbf{v}_j \ ,$$

in conflict with the assumption that the vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$ are linearly independent.