CISC 271 Class 5

Spanning Sets and Basis Vectors

Text Correspondence: §3.5

Main Concepts:

- *Span: possibly redundant vectors that describe a vector space*
- *Basis: linearly independent span, a minimal description*
- *Orthogonal basis: a special and preferred basis set*

Sample Problem, Data Analytics: What is a minimal description of a data space?

Previously, we looked at matrix subspaces to help us find complete solutions to a linear equation. We will now examine subspaces in greater detail, seeking a fuller description of the vectors in the subspaces.

A span, or linear span, of a vector space is a set of vectors for which there is at least one linear combination that equals each vector in the space. It is possible that a span has redundance, or more vectors that are strictly needed; for a k-D vector space, this happens when a span has more than k vectors.

A mathematical way to describe the redundance is as the opposite of linear independence. A set of vectors $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$ is linearly independent can be defined as:

> *For real numbers $w_1, w_2, \ldots, w_n$, the vectors $\vec{u}_j$ are **linearly independent** if and only if the sum*
>
> $$w_1\vec{u}_1 + w_2\vec{u}_2 + \cdots + w_n\vec{u}_n$$
>
> *is the zero vector $\vec{0}$ if and only if every number $w_j = 0$*

A dependent span is simply a span that is not linearly independent. This has two implications:

- For some set of numbers w_j, at least one of which is non-zero, the sum $w_1\vec{u}_1 + w_2\vec{u}_2 + \cdots + w_n\vec{u}_n = \vec{0}$; and
- Any vector $\vec{u}_j$ in the span can be expressed as the linear combination of the other vectors in the span.

Here are 5 examples of size-2 vectors that are variously linearly independent or not. The student should understand why each of the examples is, or is not, linearly independent:

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}$$

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ 0.0001 \end{bmatrix} \right\}$$

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} -2 \\ -2 \end{bmatrix} \right\}$$

$$\left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}$$

$$\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \right\} \text{ for } \textit{any} \text{ values of } a \text{ and } b$$

A *basis* is another term for a linearly independent span. In this course we are most interested in a basis as it relates to a matrix, especially the column basis and a basis for the null space.

We now have enough terminology to be able to relate concepts about vector spaces to concepts about matrices. Consider a matrix that is composed as a set of vectors, that is, each column of the matrix is a distinct vector in the set. Each vector is in $\mathbb{R}^m$ and there are n such vectors, so the matrix is some $A \in \mathbb{R}^{m\times n}$. In such a case:

- The matrix is full rank if and only if $\mathrm{rank}(A) = \min(m, n)$.
- The set of vectors are linearly independent if and only if $\mathrm{rank}(A) = n$.
- If $n > m$ then the set of vectors is redundant.
- If the null space is non-trivial, containing one or more non-zero vectors, then the set of vectors is redundant.
- If the null space is trivial, containing only the zero vector, then the set of vectors is linearly independent.
- If the matrix is full rank and invertible, then the vectors are a basis for $\mathbb{R}^m$.

The final point is an important one. If a matrix is full rank, then the columns are linearly independent. If the matrix is invertible, then the number of columns is the same as the size of each column ($m = n$) and the columns are linearly independent. An invertible matrix provides a basis for $\mathbb{R}^m$ and a basis for $\mathbb{R}^m$ provides an invertible matrix.

This leads to a remarkable, fundamental result in linear algebra:

> For a n-D vector subspace $\mathbb{V} \subseteq \mathbb{R}^m$ with a basis set $\{\vec{u}_1, \vec{u}_2, \ldots, \vec{u}_n\}$, any non-zero vector $\vec{w} \in \mathbb{V}$ can be written as a unique linear combination of the basis vectors.

The proof of this assertion is usually by contradiction, that is, by supposing that there are 2 distinct ways of writing $\vec{w}$ and showing that these ways are identical.

The typical application of the assertion is for a basis of the complete vector space $\mathbb{R}^m$, in which case $m = n$ and the basis vectors can be composed into an invertible matrix A. Solving the linear equation $A\vec{w} = \vec{c}$ for the coefficients w_i gives a unique construction of $\vec{c}$ in the basis.

The student should work through examples in the text, both in the relevant section and in its problem set, to ensure that these concepts are thoroughly understood.

5.1 Orthogonal Basis and Orthonormal Basis

A pair of vector is *orthogonal* when their dot product, or inner product, is zero. The extra notes for Class #1 describes dot products and vector norms in more detail, so the student should be already familiar with these concepts.

An *orthogonal basis* is a basis with mutually orthogonal vectors, that is, $i \neq j \Rightarrow \vec{u}_i \cdot \vec{u}_j = 0$. Numerically, an orthogonal basis is usually preferable to "just" a basis, which is a linearly independent span; but the span or non-orthogonal basis may have an intuitive or other appeal[1].

When each vector in the orthogonal basis is of unit length, it is an *orthonormal basis*. (A confusing but nearly universal usage is that a basis is orthonormal but, when the vectors are augmented into a matrix, that matrix is called "orthogonal" but is not called orthonormal.)

With this new terminology, we can now make better assertions about what Matlab returns when operating on a matrix. Some assertions, for a matrix A in Matlab object `A`, are:

`orth(A)` Orthonormal basis for the column space; returned as a matrix, the columns of which are orthogonal vectors of unit length

`null(A)` Orthonormal basis for the null space

[1] The astute student will have noticed that this is the first time we have used the dot product or vector norm. There is a specialization in mathematics and mathematical physics – called differential geometry – that is devoted to the study of non-linear spaces in the presence and absence of vector norms, but this specialization is beyond the scope of this course.

`ortc(null(A))` Orthonormal basis for the row space; computed as the orthogonal complement of the null space

`ortc(orth(A))` Orthonormal basis for the column complement, also called to co-kernel or left null space of A

An orthogonal matrix, which represents an orthonormal basis set of spanning vectors, has special properties that will be useful later in the course. These properties come from the mutual orthogonality of the vectors and that they are of unit length. To begin, we will observe that for any two vectors having the same number of entries, $\vec{u}$ and $\vec{w}$, the dot product $\vec{u} \cdot \vec{w}$ can be written as the matrix product of a transpose:

$$\vec{u} \cdot \vec{w} = \vec{u}^T \vec{w} \tag{5.1}$$

Suppose that the vectors of an orthonormal basis are written as $\{\vec{q}_j\}$. Using Equation 5.1 to express mutual orthogonality,

$$i \neq j \Rightarrow \vec{q}_i^T \vec{q}_j = 0 \tag{5.2}$$

Using Equation 5.1 to express unit length,

$$i = j \Rightarrow \vec{q}_i^T \vec{q}_j = 1 \tag{5.3}$$

If we assemble the basis vectors $\vec{q}_j$ into a square matrix Q, then the matrix product $A = Q^T Q$ can be found using Equation 5.2 and Equation 5.3 as:

$$m_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases} \tag{5.4}$$

Because the entries described in Equation 5.4, we have shown that $Q^T Q = I$ which means that $Q^T = Q^{-1}$, or:

The inverse of an orthogonal matrix is its transpose.

5.2 Rank-Nullity Theorem

The link between the rank of a matrix and the dimension of its null space it known as the *rank-nullity theorem*. We will take the rank r of a matrix $A \in \mathbb{R}^{m \times n}$ to be the number of independent columns, which is the dimension of its column space, or the number of vectors in its basis. The dimension of the null space is the number of independent vectors in its null space. Here, the rank is determined from the image space (range space) $\mathbb{R}^m$ and the null space is determined from the domain space $\mathbb{R}^n$.

The rank-nullity theorem states:

> *The rank, plus the dimension of the null space, equals the dimension of the domain space:* $r + \dim(\text{null}(A)) = n$

There are many ways to prove the rank-nullity theorem. In this course we do not concentrate on proofs, so we will be content to make observation on how the proof can proceed and the interested student is encouraged to read other source for proof details.

We can begin to appreciate the rank-nullity theorem by considering some examples of $3 \times n$ matrices, for small n. We will think of each matrix as being block-partitioned into columns; a vector $\vec{u}$ in its domain acts on the matrix columns by weighting the sum of the columns, which we recall to be

$$A\vec{u} = \sum_{j=1}^{n} u_j \vec{a}_j = u_1\vec{a}_1 + u_2\vec{a}_2 + \cdots + u_n\vec{a}_n$$

The simplest case is a 3×1 matrix, such as

$$A = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} \qquad \text{null}(A) = \{\} \tag{5.5}$$

The matrix of Equation 5.5 needs $n = 1$ coefficients to create the weighted sum. The column is not the zero vector, so the rank of the matrix is $r = 1$. The null space is empty: if $u_1 \neq 0$ then $u_1\vec{a}_1 \neq \vec{0}$, so there are no non-zero vectors $\vec{u}$ that are mapped to the zero vector.

The next simple case is a 3×2 full-rank matrix, such as

$$A = \begin{bmatrix} 1 & 1 \\ 4 & -4 \\ 1 & 1 \end{bmatrix} \qquad \text{null}(A) = \{\} \tag{5.6}$$

The matrix of Equation 5.6 needs $n = 2$ coefficients to create the weighted sum, so the domain is $\mathbb{R}^2$. The columns are non-zero and linearly independent, meaning that this matrix has rank $r = 2$. The null space is empty, which can be verified by seeing that if both $u_1 \neq 0$ and $u_2 \neq 0$, then $u_1\vec{a}_1 + u_2\vec{a}_2 \neq \vec{0}$.

A 3×2 matrix might be rank-deficient, such as

$$A = \begin{bmatrix} 1 & -2 \\ 4 & -8 \\ 1 & -2 \end{bmatrix} \qquad \text{null}(A) = \begin{bmatrix} -2 \\ 1 \end{bmatrix} \tag{5.7}$$

The matrix of Equation 5.7 also needs $n = 2$ coefficients to create the weighted sum, so the domain is $\mathbb{R}^2$. The columns are non-zero but they are linearly dependent because $\vec{a}_2$ is -2 times $\vec{a}_1$, meaning that this matrix has rank $r = 1$. The null space is spanned by one vector, with $\dim(\text{null}(A)) = 1$; this can be verified by seeing that if $u_1 = -2$ and $u_2 = 1$, then $u_1\vec{a}_1 + u_2\vec{a}_2 = \vec{0}$.

Each student is encouraged to use Matlab or other software to create other sample matrices and verify the rank-nullity properties. Because random real-valued matrices are almost always full rank, linear dependence can be imposed by additional calculations.

5.3 Orthogonal Subspaces

A concept that we will use often in this course is *orthogonality*. For two vectors, $\vec{u}$ and $\vec{v}$, we say that they are orthogonal if and only if their dot product is zero. We can write this as

$$\vec{u} \cdot \vec{v} = \vec{v} \cdot \vec{u} = \vec{u}^T \vec{v} = \vec{v}^T \vec{u} = 0 \tag{5.8}$$

We can also say that vector subspaces are orthogonal. What we mean is that every vector in one space is orthogonal to every vector in the other space; the zero vector is, by specification, orthogonal to any vector because the dot product of the zero vector and any vector is zero.

Definition: orthogonal subspaces

For any vector subspace $\mathbb{U} \subseteq \mathbb{R}^n$, and any vector subspace $\mathbb{V} \subseteq \mathbb{R}^n$, and any vector $\vec{u} \in \mathbb{U}$, and any vector $\vec{v} \in \mathbb{V}$, $\mathbb{U}$ and $\mathbb{V}$ are *orthogonal subspaces* is defined as

$$\vec{u}^T \vec{v} = 0 \tag{5.9}$$

An important related concept is the *orthogonal complement* of a vector subspace. This is the subspace of *every* vector that is orthogonal to the original subspace. For a vector subspace $\mathbb{U}$, we will sometimes write the orthogonal complement as $\mathbb{U}^\perp$ and pronounce the symbol a "$\mathbb{U}perp$.

Definition: orthogonal complement

For any vector subspace $\mathbb{U} \subseteq \mathbb{R}^n$, the *orthogonal complement* is defined as the vector subspace $\mathbb{U}^\perp$ such that, for any vector $\vec{u} \in \mathbb{U}$,

$$\left(\vec{u}^T \vec{v} = 0\right) \rightarrow \left(\vec{v} \in \mathbb{U}^\perp\right) \tag{5.10}$$