ARE Math Review Lecture 5: Linear Independence

Stephen M. Stohs

August 17, 2004

1 Parametric Equations and Subspaces

Yesterday we ended with an example of the parametric equation for a four-dimensional hyperplane. Consider what happens to this equation if we set $\mathbf{x_0} = \mathbf{0}$; what results is the parametric representation of a hyperplane through the origin. In the four dimensional case, it is

$$\mathbf{x}(t, s, r) = t\mathbf{v} + s\mathbf{w} + r\mathbf{u}. \tag{1}$$

In order to explore the nature of the set of points in this hyperplane, we consider the points described by the parametric equation subject to various possible restrictions.

1. Setting $t = s = r = 0$ verifies that the origin itself is in the hyperplane.

2. By permitting exactly one of t, s, or r to be nonzero, we see the line through the origin in the direction of any of the vectors $\mathbf{v}, \mathbf{w}$ or $\mathbf{u}$ is contained in the hyperplane.

3. By restricting exactly one of t, s or r to *equal* zero, we see that the planes through the origin which contain any pair of the vectors $\mathbf{v}, \mathbf{w}$ or $\mathbf{u}$ are all three contained in the hyperplane.

4. Finally, by first focusing attention on the case $r = 0$, then relaxing this restriction, we can think of generating any point in the hyperplane by projecting off the plane described by

$$\mathbf{y}(t, s) = t\mathbf{v} + s\mathbf{w} \tag{2}$$

in the direction of $\mathbf{u}$. In three dimensional Euclidean space, this procedure would enable us to reach any point in the space, since a line through an arbitrary point in the direction $\mathbf{u}$ would cross the plane somewhere[1]. In higher dimensions, this is no longer the case; there will be points included in the space which will be missing from the hyperplane constructed from three independent vectors.

The expression which defines the four-dimensional hyperplane $t\mathbf{v}+s\mathbf{w}+r\mathbf{u}$ is referred to as a **linear combination** of the vectors $\mathbf{v}, \mathbf{w}$, and $\mathbf{u}$. Each of the restricted cases defines what is called a **subspace** of R^4, a subset of R^4 which is **closed**[2] under the operations of addition and scalar multiplication. In general, a subspace is a vector space in its own right whose vectors also inhabit some larger vector space.

1.1 Linear Combinations in Matrix Form

Next consider writing a set of vectors $\{\mathbf{v_i},\ i = 1, 2, \ldots, n\}$ in column vector form and forming a matrix A by stacking them side-by-side:

$A = \begin{bmatrix} \mathbf{v_1} & \mathbf{v_2} & \mathbf{v_3} & \ldots & \mathbf{v_n} \end{bmatrix}$. Let $\mathbf{x}$ represent an arbitrary vector in R^n. Claim: Any linear combination of the $\{\mathbf{v_i}\}$ may be written in the form $A\mathbf{x}$. As already discussed above, the collection of all such linear combinations formed by right-multiplying the matrix A by column vectors in R^n is the column space of A, or equivalently, the span of the $\{\mathbf{v_i}\}$:

$$\mathcal{L}\left[\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\right]. \tag{3}$$

Exercise: Verify this assertion about the matrix form of a linear combination – that is, right-multiplying a matrix A by a column vector $\mathbf{x}$ results in a linear combination of the columns of A, with scalar coefficients given by the components of $\mathbf{x}$. Hint: The definition of A shows how to think of it as a partitioned matrix. Partition $\mathbf{x}$ so that it is compatible for matrix multiplication with the partitioned form of A, then note that the product is in the form of a generalized Euclidean inner product in matrix form (i.e., row "vector" times column vector).

[1] The independence assumption implies that $\mathbf{u}$ cannot be parallel to the plane.

[2] This means that addition of any two vectors in the subspace results in a sum which is also in the subspace, and scalar multiplication of a vector in the subspace results in a new vector which is in the subspace.

1.2 Recasting a Parametric Equation as an Affine Function

The parametric representations of lines, planes, and hyperplanes share a common conceptual foundation: In each case, the parametric equation describes a subspace with a displaced origin. This insight is reinforced by noting that the parametric equations for a line, a plane, and a hyperplane may all be recast as **affine functions**, which are generically described by $f(\mathbf{x}) = A\mathbf{x} + \mathbf{c}$, for some $m \times k$ matrix A and some m-vector $\mathbf{c}$. To tailor the generic affine function to our examples, we would write the vectors $\mathbf{v}, \mathbf{w}$ and/or $\mathbf{u}$ as the columns of A, let $\mathbf{x_0}$ play the role of $\mathbf{c}$, and write the parameters t, s and/or r as the components of $\mathbf{x}$.

The matrix multiplication term $A\mathbf{x}$ is an example of a **linear transformation** of the columns of A; considering all possible values of $\mathbf{x}$, this term describes the smallest possible subspace[3] which contains the column vectors of A. Adding a nonzero vector $\mathbf{x_0}$ has the effect of translating this subspace to a new origin. The **image**[4] of an affine function is thus generally a subspace which has been translated by some displacement from the original origin.

2 Linear Independence

2.1 Linear Combinations and the Definition of Independence

Let $\{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ be a set of vectors in R^m. A **linear combination** of these vectors is a new vector formed by multiplying each vector in the given set by a scalar c_i and adding: $c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_n\mathbf{v_n}$.

We say the set of vectors is **linearly independent** if the only possible linear combination of the $\mathbf{v_i}$ which equals the zero vector is the one with all $c_i = 0$. That is, the vectors are linear independent if

$$c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_n\mathbf{v_n} = 0 \tag{4}$$

implies that $c_i = 0$ for $i = 1, 2, \ldots, n$.

[3] This subspace may be described as the **span** of the columns of A, or the **column space** of A.

[4] The **image** of a function is the set of all possible values which may result when the function is applied to an element in the **domain**.

Equivalently, a linearly independent set of vectors has the property that it is not possible to write any of the vectors in the set as a linear combination of the remaining vectors in the set. To see why this is so, assume the contrary; that is, suppose that there is some vector $\mathbf{v_i}$ for some $i \in \{1, 2, \ldots, n\}$ such that

$$\mathbf{v_i} = \sum_{j \neq i} c_j \mathbf{v_j}. \tag{5}$$

Then if we let $c_i = -1$, then it is clear that

$$c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_n\mathbf{v_n} = 0 \tag{6}$$

without all $c_i = 0$. Thus the vectors are not a linearly independent set.

Conversely, suppose there is a linear combination

$$c_1\mathbf{v_1} + c_2\mathbf{v_2} + \ldots + c_n\mathbf{v_n} = 0 \tag{7}$$

with $c_i \neq 0$ for some $i \in \{1, 2, \ldots, n\}$. Then we can move this term in the linear combination to the left-hand side, divide through by $-c_i$, and write

$$\mathbf{v_i} = \sum_{j \neq i} d_j \mathbf{v_j}, \tag{8}$$

with $d_j \equiv -\frac{c_j}{c_i}$. Thus if a set of vectors is not linearly independent, it is possible to write at least one of the vectors in the set as a linear combination of the remaining vectors.

Exercise: Consider a matrix in row echelon form. Can we say anything about whether the rows are linearly independent?

3 Span, Basis and Dimension

3.1 Definitions

As suggested above, the **span** of a set of vectors $V = \{\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\}$ in R^m is the set of all possible linear combinations of the set:

$$\mathcal{L}\left[\mathbf{v_1}, \mathbf{v_2}, \ldots, \mathbf{v_n}\right]. \tag{9}$$

Suppose the vectors in our representative set V are not independent. From the foregoing remarks, we know that if a set of vectors is not independent, then it is possible to write at least one of them as a linear combination of the other vectors in the set:

$$\mathbf{v_i} = \sum_{j \neq i} c_j \mathbf{v_j}. \tag{10}$$

It follows that the vector on the left hand side lies in the span of the vectors in the right hand side, and the vectors on the RHS hence form a smaller set of vectors which span the subspace. Suppose that we systematically eliminate all dependent vectors from the spanning set; clearly the process cannot go on forever if we have a finite set of vectors, and it will terminate when we are left with an independent set. Once this process of eliminating dependent vectors ends, we will be left with a set of independent vectors which is a spanning set for the subspace of minimal size, called a **basis** for the subspace. Any vector in the subspace may be written as a linear combination of the vectors in the basis, and the remaining vectors in the basis will no longer span the subspace if any further vectors are removed.

The number of vectors in a basis is well-defined, in the sense that any basis for a given subspace must have the same number of members. This number of vectors needed to form a basis is called the **dimension** of the subspace or vector space in question.

3.2 Examples

Next consider some examples in R^2 and R^3 to illustrate the concepts presented above. Starting in the two-dimensional Cartesian plane with any nonzero vector $\mathbf{x} = (x_1, x_2)$, we know that $\{t\mathbf{x} | t \in R\}$ describes a line through the origin with slope x_2/x_1. This line constitutes a one-dimensional subspace of R^2. If we form a set of vectors $V = \{\mathbf{x}, \mathbf{y}\}$, where $\mathbf{y} = (y_1, y_2)$ is another vector in R^2, then the set forms a basis if and only if $\mathbf{y}$ is not a scalar multiple of $\mathbf{x}$. Otherwise the set is dependent and the span is the line through the origin as before, constituting a proper subspace of R^2. If $V = \{\mathbf{x}, \mathbf{y}\}$ is a linearly independent set (and hence a basis for R^2), then any other vector $\mathbf{z} = (z_1, z_2)$ may be written as a linear combination of $\mathbf{x}$ and $\mathbf{y}$, so if $\mathbf{z}$ were added to V, the set would no longer be either linearly independent or a basis.

A similar line of discussion pertains to R^3. Geometrically, the set of all scalar multiples of a single vector $\mathbf{x}$ forms a straight line through the origin. Adding a second vector $\mathbf{y}$ which is not a scalar multiple of $\mathbf{x}$ to the set V of vectors under consideration implies that the span of V is a plane through the origin containing both $\mathbf{x}$ and $\mathbf{y}$. If a third vector $\mathbf{z}$ is added to the set V and $\mathbf{z}$

cannot be written as a linear combination of **x** and **y**, then the three vectors form a linearly independent set and hence a basis: The set has dimension three and spans R^3. Any vector in R^3 can be written as a linear combination of these basis vectors. If we added a fourth vector **w** to V, then since the vectors already present form a basis, it follows that **w** could be written as a linear combination of the other vectors, and the set would no longer either be linearly independent or form a basis.

These examples illustrate the concepts of span, basis, and dimension. In summary:

1. The set of all possible linear combinations of any set V of vectors defines a subspace referred to as the span of V.

2. Any linearly independent set of vectors which spans a subspace (or a vector space) constitutes a basis.

3. A basis is a spanning set of minimal size.

4. The number of elements in any basis for a given vector space (or subspace) is a **well-defined** quantity[5], referred to as the dimension.

5. The dimension of any basis for R^n contains n vectors.

6. If a vector is removed from a basis, the remaining vectors will no longer span associated vector space.

7. If a vector belonging to a given vector space V is added to a basis for V, the resulting set will no longer be linearly independent, as the new vector may be written as a linear combination of the others.

4 Matrix Algebra and Euclidean Spaces

In this last section of the final installment of my notes on linear algebra, I will consider a few miscellaneous topics which I believe help unify the various ideas we have covered thus far.

[5] By **well-defined**, we mean that the value is the same for all possible sets of basis vectors for the vector space.

4.1 Matrix Multiplication and Linear Combinations

We have already seen how the product of a matrix and a column vector $A\mathbf{x}$ creates a linear combination of the columns of A. Now we wish to verify two characterizations of matrix multiplication which were mentioned when this operation was first introduced. Namely, the product of an $m \times k$ matrix A with a $k \times n$ matrix B, $C = AB$, may be described as follows:

1. Row i of C is a linear combination of the rows of B, with the coefficients taken from the corresponding components of row i of A.

2. Column j of C is a linear combinations of the columns of A, with the coefficients given by the corresponding components of column j of B.

Exercise: Prove the above assertion two ways:

1. Write a representative row (column) of C in terms of the summations which define its elements. Then use the rules of vector addition and scalar multiplication to rewrite the row (column) as a linear combination as described.

2. Partition A (or B) appropriately, then apply matrix multiplication to the partitioned matrix to demonstrate the result.

4.2 Column Space and Invertibility

Recall the general system of linear equations written in matrix form:

$$A\mathbf{x} = \mathbf{b}. \tag{11}$$

By previous comments, we can write

$$A\mathbf{x} = \sum_{j=1}^{k} x_j A_{.j}, \tag{12}$$

which shows the left-hand side to be a linear combination of the columns of A (assuming k columns) with coefficients x_j. Hence we see that the existence of a solution to the system requires that $\mathbf{b}$ lies in the column space of A. Now suppose that A is $n \times k$. If $k < n$, we know that the columns of A cannot form a basis for R^n; there are not enough vectors there, no matter whether the columns are independent. Hence we see that a matrix with more rows than columns cannot be nonsingular – there will always be some RHS vectors **b** for which the system will have no solution.

Now assume that A is square, $k = n$, and consider once again the question of whether A is nonsingular, or equivalently, whether the system

$$A\mathbf{x} = \mathbf{b} \tag{13}$$

has a solution for every right hand side vector **b**. Since **b** is allowed to be any vector in R^n, we see that the question is equivalent to asking whether the columns of A form a basis for R^n. Thus A is nonsingular if and only if it is an $n \times n$ (square) matrix whose columns are linearly independent. In other words, the column space of A has dimension n and is, in fact, the entire n-dimensional Euclidean space. By earlier comments, this is equivalent to the statement that A is invertible. So the columns of an invertible matrix A form a basis for R^n.

Finally consider the case $n < k$, where there are fewer equations than unknowns. The columns of A necessarily form a dependent set. If there are n independent vectors among the k columns of A, then we can identify these vectors as a basis for R^n. We can assign arbitrary values to the remaining $k-n$ variables (the independent variables in this scenario) and subtract them from the RHS of the system. What remains on the left may be written as the product of a nonsingular matrix times the vector of variables corresponding to the independent columns of A, and this system has a solution for ever assignment of values to the $k-n$ independent variables. (Think linear implicit function theorem!) This argument shows why there are generally an infinite number of possible solutions to a system with fewer equations than unknowns.

Exercise: How does the last example play out if there are fewer than n independent columns of A? In particular, are there still infinitely many possible solutions?

4.3 The Nonparametric Equation of a Hyperplane as a System of Linear Equations

Yesterday we discussed the parametric representation of lines, planes, and hyperplanes. This section considers nonparametric representations, and connects the ideas of Euclidean spaces back to one of the first topics we discussed, that of the solution to a system of linear equations.

4.3.1 Nonparametric Equations for a Line, a Plane, and a Hyperplane

Recall from elementary analytical geometry that the standard form of a line in x-y space is given by

$$ax + by = c. \tag{14}$$

The line crosses the y-axis at the height $\frac{c}{b}$ (found by setting $x = 0$ and solving for y), while the slope is $m = -\frac{a}{b}$.

The three-dimensional analogue of the two-dimensional formula for a line describes a plane:

$$ax + by + cz = d. \tag{15}$$

In vector notation, a plane may be written $\mathbf{n} \cdot (\mathbf{x} - \mathbf{x_0}) = 0$, which says that the plane consists of all vectors with tails at $\mathbf{x_0}$ which are orthogonal to the normal vector, $\mathbf{n}$. To translate the standard linear equation form given above into vector notation[6], we could write $\mathbf{n} = (a, b, c)$, then take $\mathbf{x_0} = (0, 0, d/c)$.

The extension to n-dimensional Euclidean space is straightforward: The $(n-1)$-dimensional hyperplane has linear equation

$$c_1x_1 + c_2x_2 + \ldots + c_nx_n = d, \tag{16}$$

and the vector representation may be written in the same form as previously, $\mathbf{n} \cdot (\mathbf{x} - \mathbf{x_0}) = 0$, but now with $\mathbf{n} = (c_1, c_2, \ldots, c_n)$ and $\mathbf{x_0} = (0, 0, \ldots, 0, d/c_n)$.

4.3.2 Characterization as a System of Linear Equations

The linear (nonparametric) equations for a line, plane, or hyperplane all may be thought of as constituting a system of one equation in n unknowns (where n depends on the dimensions of the space under consideration). This suggests that we might employ the Gaussian elimination approach to analyze the solutions to these systems.

Exercise:

1. For the four-dimensional case, use Gaussian elimination to characterize the solution to the system consisting of the sole linear restriction that $\mathbf{x} = (x_1, x_2, x_3, x_4)$ lies in the hyperplane

$$a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4 = b. \tag{17}$$

[6] We can also apply this setup to the two-dimensional case, taking $\mathbf{n} = (a, b)$ and $\mathbf{x_0} = (0, c/b)$.

2. Show that this characterization of the solution takes us from the non-parametric representation of the hyperplane to the parametric representation.

   Hint: Assuming that $a_1 \neq 0$, we can divide through both sides by a_1, then subtract all terms other than the first from the RHS. The result is an expression of form

   $$x_1 = c_2 x_2 + c_3 x_3 + c_4 x_4 + b. \tag{18}$$

   Next note that the variables x_2, x_3 and x_4 are free, so we may assign them arbitrary values; say $x_2 = t$, $x_3 = s$, and $x_4 = r$, where $t, s, r \in R$. Substitution for the free variables in the expression for x_1 shows how x_1 depends on t, s, and r.

   After these steps, write down the general expression for a solution (x_1, x_2, x_3, x_4) as a vector in terms of t, s, r, and b. Finally, decompose this vector into the form of a parametric representation of a hyperplane, by expressing it as a linear combination of basis vectors translated away from the origin by some displacement vector.

3. Consider the basis vectors in the linear combination portion of the parametric equation for the hyperplane. What is the angle between the normal vector to the hyperplane, $\mathbf{n} = (a_1, a_2, a_3, a_4)$ and each of these basis vectors? What is the angle between the normal vector and a vector in the subspace defined by the span of this basis?