Vector Spaces and Matrices

Kurt Bryan

Matrices as "Functions"

Up to now matrices have been pretty static objects. We've used them mainly as a bookkeeping tool for doing Gaussian elimination on systems of equations, as a way to avoid writing all the x_k variables when we perform elementary operations on equations. But in fact matrices can be a bit more dynamic—a matrix can actually be thought of as a function, and this point of view can give a lot of insight into the question of whether a system of linear equations is solvable or not, and what to do if there is no solution .

Specifically, suppose we have m by n matrix $\mathbf{A}$. We can compute $\mathbf{Ax}$ for any vector $\mathbf{x}$ in $\mathbb{R}^n$; the result is a vector in $\mathbb{R}^m$. In this way a matrix $\mathbf{A}$ can be thought of as a function from $\mathbb{R}^n$ to $\mathbb{R}^m$. More precisely, an m by n matrix "induces" a function $\mathbf{x} \rightarrow \mathbf{Ax}$ from $\mathbb{R}^n$ to $\mathbb{R}^m$. (We shouldn't actually write $\mathbf{A}$ for this function; $\mathbf{A}$ is the matrix itself. Some people write $T_{\mathbf{A}}$ to denote the actual function from $\mathbb{R}^n$ to $\mathbb{R}^m$, but I won't harp on this).

Example 1: Let

$$\mathbf{A} = \begin{bmatrix} 1 & 4 & 5 \\ -1 & 0 & 3 \end{bmatrix}. \tag{1}$$

Then for a vector $\mathbf{x} =< x_1, x_2, x_3 >$ in $\mathbb{R}^3$,

$$T_{\mathbf{A}}(\mathbf{x}) = \mathbf{Ax} = \begin{bmatrix} 1 & 4 & 5 \\ -1 & 0 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_1 + 4x_2 + 5x_3 \\ -x_1 + 3x_3 \end{bmatrix},$$

a vector in $\mathbb{R}^2$.

There is another very important observation to make about multiplying vectors by matrices. Let $\mathbf{A}$ be the matrix defined by (1). Then you can easily check that the matrix multiplication can be done as

$$\mathbf{Ax} = x_1 \begin{bmatrix} 1 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 4 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 5 \\ 3 \end{bmatrix}.$$

In other words, $\mathbf{Ax}$ consists of *linear combinations of the columns of* $\mathbf{A}$. You can check that this is always true—the product $\mathbf{Ax}$ is a linear combination of the columns of $\mathbf{A}$.

This observation gives another view on why $\mathbf{Ax} = \mathbf{b}$ is not typically solvable if we have more equations than unknowns, say m equations in n unknowns with $n < m$. For $\mathbf{x}$ in $\mathbb{R}^n$ the vector $\mathbf{Ax}$ is a linear combination of a set of n vectors in m dimensional space. Since $n < m$, such a set of vectors *cannot* span $\mathbb{R}^m$. As a result, the range of the function $T_{\mathbf{A}}(\mathbf{x}) = \mathbf{Ax}$ is not $\mathbb{R}^m$, but some proper subspace of $\mathbb{R}^m$, and for most $\mathbf{b} \in \mathbb{R}^m$ there won't be any $\mathbf{x}$ such that $\mathbf{Ax} = \mathbf{b}$.

Example 2: Let

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 0 & -1 \\ -1 & 3 \end{bmatrix} \qquad (2)$$

and $\mathbf{x} =< x_1, x_2 >$ then

$$\mathbf{Ax} = x_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + x_2 \begin{bmatrix} 2 \\ -1 \\ 3 \end{bmatrix}.$$

This generates only a two-dimensional subspace of $\mathbb{R}^3$, i.e., a plane. If a three-dimensional vector $\mathbf{b}$ lies in this plane then $\mathbf{Ax} = \mathbf{b}$ is solvable. But as your geometric intuition should tell you, "most" vectors in $\mathbb{R}^3$ will not lie in such a plane, and for such a vector $\mathbf{b}$, $\mathbf{Ax} = \mathbf{b}$ will have no solution.

Problem 1:

- Let $\mathbf{A}$ be the matrix defined by equation (2) and let $\mathbf{b} =< a, b, c >$. Perform Gaussian elimination to solve $\mathbf{Ax} = \mathbf{b}$ and find conditions on a, b, and c which guarantee consistency of the system, so find which $\mathbf{b}$ are in the range of $T_{\mathbf{A}}$. Hint: this condition involves a linear combination of a, b, and c. What does this subset of $\mathbb{R}^3$ look like?

Column, Null, and Row Spaces

In Example 2, as $\mathbf{x}$ ranges over all of $\mathbb{R}^2$, $\mathbf{Ax}$ generates the span of the two columns of $\mathbf{A}$ in $\mathbb{R}^3$. More generally for an m by n matrix $\mathbf{A}$, as $\mathbf{x}$ ranges over $\mathbb{R}^n$, $\mathbf{Ax}$ generates the span of the columns of $\mathbf{A}$. This span is a subspace of $\mathbb{R}^m$ and is called the *column space* of $\mathbf{A}$. In Example 2 above the column space was spanned by $< 1, 0, -1 >$ and $< 2, -1, 3 >$. In fact, since these two vectors are linearly independent they form a basis for the column space, which is consequently two-dimensional. In general the dimension of the column space of a matrix is called the *column rank* of the matrix.

But be careful. Just because a matrix has n columns doesn't mean the column space is n dimensional. For example, if

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

then the column space of $\mathbf{A}$ is just a one-dimensional line in $\mathbb{R}^2$, since the column vectors are linearly dependent.

Another subspace related to an m by n matrix is the *nullspace*, the subspace of $\mathbb{R}^m$ defined by those vectors $\mathbf{x} \in \mathbb{R}^m$ which satisfy $\mathbf{Ax} = \mathbf{0}$. The dimension of this subspace is called the *nullity* of the matrix.

Problems 2:

- Find a basis for the column spaces of the matrices defined in equations (1) and (2), by throwing out any columns which are linear combinations of other columns.

- Find a basis for the nullspaces of the matrices defined in equations (1) and (2) by solving $\mathbf{Ax} = \mathbf{0}$.

- Show that the nullspace of an m by n matrix $\mathbf{A}$ is in fact a subspace of $\mathbb{R}^n$ by showing that this set is closed under addition and scalar multiplication.
- If $\mathbf{A}$ is an invertible n by n matrix, what is the nullspace of $\mathbf{A}$? What if $\mathbf{A}$ isn't invertible—what can you say about its nullspace?
- What can you say about the nullspace of an m by n matrix when $m < n$? How about when $m > n$? Or $m = n$?
- Suppose we have found a solution $\mathbf{x} = \mathbf{x}_0$ to $\mathbf{Ax} = \mathbf{b}$ for some matrix $\mathbf{A}$. What can you say about the uniqueness of this solution if the nullspace of $\mathbf{A}$ is the trivial subspace $\mathbf{0}$? What if the nullspace of $\mathbf{A}$ has positive dimension?

Finally, another subspace associated to such a matrix is the *row space*, the subspace of $\mathbb{R}^n$ spanned by the rows of $\mathbf{A}$, considered as n dimensional vectors, of course. An m by n matrix has m rows, but (as with the column space) the dimension of the row space could be less than m if the rows are linearly dependent. The dimension of the row space of a matrix is called the *row rank* of the matrix.

Problem 3:

- Let
$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 1 & -1 \\ 0 & 1 & 2 & -2 \\ 3 & -2 & -1 & 1 \end{bmatrix}.$$
Find a basis for the row space of $\mathbf{A}$ by determining which row vectors are linear combinations of the others, and find the row rank of this matrix.

The Dimension Theorems

It turns out that the row rank, column rank, and nullity of a matrix can't be just anything—there are some fundamental relations between these quantities and the dimensions of the matrix. To understand this, we need to look more carefully at the process of finding the row rank, column rank, and nullspace of a matrix.

Finding the Row Rank

We need a concrete way to find the row space for a matrix. As always, the civilized way to specify a subspace is by giving a basis for the subspace, so we want a general procedure for computing basis vectors for the row space.

Let's proceed via a simple example. Take
$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 1 & 5 \\ 1 & 1 & -3 & -7 \\ 2 & 2 & -1 & 1 \end{bmatrix}. \tag{3}$$
The row space of $\mathbf{A}$ is spanned by the row vectors, but that doesn't mean the rows are a basis—they might be linearly dependent. We need to find the dependencies and throw out any "redundancies".

To do this we're going to do Gauss-Jordan elimination to put $\mathbf{A}$ in reduced echelon form (but at this stage we're not trying to solve $\mathbf{Ax} = \mathbf{b}$ for any $\mathbf{b}$). The first step would be the elementary row operation $R_2 \to R_2 - R_1$, leading to the matrix

$$\mathbf{A}' = \begin{bmatrix} 1 & 1 & 1 & 5 \\ 0 & 0 & -4 & -12 \\ 2 & 2 & -1 & 1 \end{bmatrix}.$$

Here's an important observation: *The rows of* $\mathbf{A}'$ *span exactly the same subspace in* $\mathbb{R}^4$ *as the rows of* $\mathbf{A}$. This is clear because:

1. The rows of $\mathbf{A}'$ are constructed from linear combinations of the rows of $\mathbf{A}$, so anything that can be built (linearly) from the rows of $\mathbf{A}'$ can be built from the rows of $\mathbf{A}$.

2. The elementary rows operations we use are reversible. As a result we can build the rows of $\mathbf{A}$ from the rows of $\mathbf{A}'$ (by reversing the row operations). Hence anything that can be built (linearly) from the rows of $\mathbf{A}$ can be built from the rows of $\mathbf{A}'$.

This fact holds through each elementary row operation—the rows of each successive matrix span the same subspace of $\mathbb{R}^4$ as the rows of the previous matrix.

Finishing off the Gauss-Jordan elimination of $\mathbf{A}$ leads us to the reduced echelon matrix

$$\mathbf{E} = \begin{bmatrix} 1 & 1 & 0 & 2 \\ 0 & 0 & 1 & 3 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{4}$$

This matrix has exactly the same row space as the original $\mathbf{A}$. And in fact I claim it's easy to see a basis for the row space of $\mathbf{E}$—it's exactly the non-zero rows of $\mathbf{E}$. Obviously the zero row at the bottom contributes nothing to the row space; the row space is spanned by the non-zero rows of $\mathbf{E}$. I claim that these rows are linearly independent, and so form a basis for the row space. There's no need for an elaborate computation to check the independence. Note that each non-zero row contains a pivot element (which I've circled). Also note that each pivot element is the lone non-zero entry in its column. As a result, if we consider the equation

$$c_1 < 1, 1, 0, 2 > + c_2 < 0, 0, 1, 3 > = < 0, 0, 0, 0 >$$

to check independence, we are led to $< c_1, c_1, c_2, 2c_1 + 3c_2 > = < 0, 0, 0, 0 >$. Now the first component of this vector is just c_1, leading the equation $c_1 = 0$. This is a consequence of the fact that pivot element in row 1 is the only non-zero entry in its column. The third component of $< c_1, c_1, c_2, 2c_1 + 3c_2 >$ is just c_2, stemming from the fact that the row 2 pivot is the only non-zero entry in the third column of $\mathbf{E}$. Of course this forces $c_2 = 0$ and we conclude that the two non-zero rows of $\mathbf{E}$ are linearly independent.

Problem 4:

- Gauss-Jordan elimination is performed on a certain 5 by 5 matrix, leading to

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 4 & 0 & 2 \\ 0 & 1 & -2 & 0 & 3 \\ 0 & 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Verify that the three non-zero rows are linearly independent.

So we now have a methodical way to find a basis for the row space of a matrix—Gauss-Jordan eliminate and take the non-zero rows as the basis. Note that each non-zero row contains exactly one pivot, and every pivot of course lives in some row. As a result

- The row rank is exactly the number of pivot elements in the reduced echelon form of the matrix.

Finding the Column Rank

Now we'll look at how one can find a basis for the column space of a matrix $\mathbf{A}$. The column space is spanned by the columns of $\mathbf{A}$, but as with the row space these vectors might not be linearly independent. What we need to do (as we did with the row space) is root out the linear dependencies. Again, let's use the matrix we used in the row space example, defined by equation (3). To test the linear independence of the columns of $\mathbf{A}$ we look for non-zero solutions $< x_1, x_2, x_3, x_4 >$ to

$$x_1 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix} + x_3 \begin{bmatrix} 1 \\ -3 \\ -1 \end{bmatrix} + x_4 \begin{bmatrix} 5 \\ -7 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

But in fact, this is nothing more than the equation $\mathbf{Ax} = \mathbf{0}$! We can solve this by forming the augmented matrix $[\mathbf{A}|\mathbf{0}]$ and doing Gauss-Jordan elimination. The computations are exactly the same as the row-space case, but with an additional column of zeros tacked on at the right (and this column stays all zeros through all elementary row operations). We are led to the matrix

$$\begin{bmatrix} 1 & 1 & 0 & 2 & 0 \\ 0 & 0 & 1 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{5}$$

corresponding to the equation $\mathbf{Ex} = \mathbf{0}$ (which has exactly the same solutions as $\mathbf{Ax} = \mathbf{0}$). Notice that the pivot elements live in columns 1 and 3. I claim that all other columns can be built from these two columns.

To see this, cast $\mathbf{Ex} = \mathbf{0}$ back into the "linear combinations of columns" form, i.e.,

$$x_1 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + x_3 \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} + x_4 \begin{bmatrix} 2 \\ 3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}. \tag{6}$$

Now notice that taking $x_1 = 1, x_2 = -1$, and all other $x_k = 0$ gives a solution to the equation. In terms of the original matrix, this means that

$$\mathbf{A}\begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \end{bmatrix} = \mathbf{0}$$

so the first column of **A** minus the second column of **A** is zero, i.e., the second column equals the first. This means that the second column of **A** is a linear combination of the other columns. We can throw out the second column as a basis element.

Now look back at equation (6) and note that $x_1 = 2, x_2 = 0, x_3 = 3, x_4 = -1$ gives a solution to $\mathbf{Ex} = \mathbf{0}$, and hence also to $\mathbf{Ax} = \mathbf{0}$. This means that the 4th column in **A** is twice the first column plus three times the third column, i.e., a linear combination of these columns. We can throw out column 4 of **A** as a basis element.

What remains as basis elements for the column space are exactly those columns of **A** in which the pivot elements reside (in **E**).

Problem 5:

- Let **E** be the 5 by 5 matrix reduced echelon matrix in the problem at the bottom of page 4. Verify that any column without a pivot can be written as a linear combination of those columns which contain pivots.
- Make up a few matrices in Maple, then hit them with the `rref` command to get them to reduced echelon form. Look at the structure of the reduced echelon form and convince yourself that any column without a pivot can be written as a linear combination of columns with pivots. Also convince yourself that the pivot columns are always linearly independent.
- True or False: The column space of a matrix is the same as the column space of the reduced echelon matrix.

What we've now shown is that the dimension of the column space is the number of pivot elements in the reduced echelon form of the matrix. But this was exactly the dimension of the row space as well! We've proved that

Theorem: The row rank of a matrix equals the column rank of the matrix.

This is an especially surprising fact, given that the row and column spaces are subspaces of entirely different vectors spaces, $\mathbb{R}^m$ and $\mathbb{R}^n$ with different m and n. Since the row and column rank must be the same, we often just refer to the *rank* of a matrix.

The Rank and Nullity

Look back at the matrix defined by equation (3) and consider the problem of finding a basis for the nullspace. Finding the nullspace means finding all solutions to $\mathbf{Ax} = \mathbf{0}$. But we've already done the work of forming $[\mathbf{A}|\mathbf{0}]$ and doing Gauss-Jordan elimination. It led to the matrix in (5). Now if we backsubstitute to find the solutions, we see that columns 4 and 2 lack pivots and hence give rise to free variables—I'll call these variables s and t. The backsubstitution actually produces vectors $< -2s - t, t, -3s, s >$, or

$$\mathbf{x} = s \begin{bmatrix} -2 \\ 0 \\ -3 \\ 1 \end{bmatrix} + t \begin{bmatrix} -1 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

In this case (since the vectors are independent) you can see that the nullspace is two dimensional.

More generally, the vectors produced by this procedure are ALWAYS independent, for a reason very similar to the reason that the non-zero rows in the reduced echelon form are independent: If the nullspace computation produces vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$, (coming from k free variables) then a free variable in the jth column will give rise to a nullspace vector which has a 1 in the jth position, while all other vectors will have zeros in this position.

Problem 6:

- Again consider the matrix $\mathbf{E}$ at the bottom of page 4 (already in reduced echelon form). Backsubstitute $\mathbf{Ex} = \mathbf{0}$ to find vectors which span the nullspace. Verify that the vectors are linearly independent—as asserted above, each will contain a 1 in a position where all other vectors so produced contain 0.

Now given an m by n matrix $\mathbf{A}$, our column/row rank computations show that each column with a pivot in the reduce echelon form contributes one to the rank of the matrix. The nullspace computations show that each column without a pivot contributes one to the nullity of the matrix. Since each column either contains a pivot or doesn't, we shown the

Rank-Nullity Theorem: The rank of a matrix plus the nullity of the matrix equals the number of columns in the matrix.

This theorem is also sometimes called the *Dimension Theorem*, or even the *Fundamental Theorem of Linear Algebra.*

Problems 7:

- Verify that this was the case in Problem 3.
- Suppose $\mathbf{A}$ is a 3 by 7 matrix. Why must $\mathbf{A}$ have a non-trivial nullspace?
- Suppose $\mathbf{A}$ is a 3 by 7 matrix and $\mathbf{Ax} = \mathbf{b}$ is not solvable for a certain vector $\mathbf{b}$ in $\mathbb{R}^3$. What can you say about the nullity of $\mathbf{A}$?

Least Squares Solutions

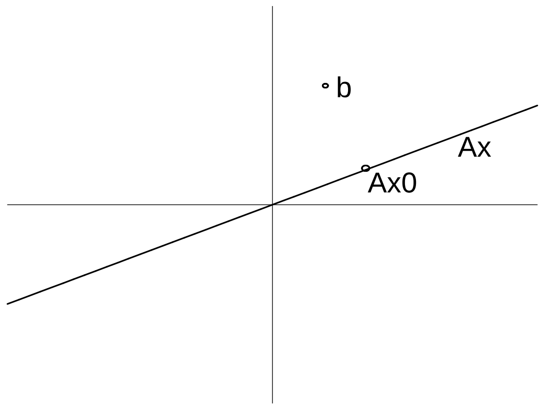
Consider the points $(-1,4),(-1,3),(0,2),(1,1),(3,7)$. Suppose we want to interpolate these points with a quadratic polynomial $p(x) = a + bx + cx^2$. Of course a quadratic typically can't be drawn through 5 points, and in the present case it's clearly hopeless (for a polynomial of any degree), for the first two points have the same x coordinates but different y coordinates. Nonetheless, we ought to do the best we can; if you plot the points, they do lie approximately on some parabola.

If we were to try to find an interpolating polynomial, we'd be led to $\mathbf{Ax} = \mathbf{b}$, with

$$\mathbf{A} = \begin{bmatrix} 1 & -1 & 1 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 3 \\ 2 \\ 1 \\ 7 \end{bmatrix} \tag{7}$$

a system with no solution.

The problem is that the column space of $\mathbf{A}$, a subspace of $\mathbb{R}^5$, doesn't contain $\mathbf{b}$. Here's a picture to help you visualize the situation:



Our "solution" to this problem will consist of choosing $\mathbf{x}_0$ in $\mathbb{R}^3$ so that $\mathbf{Ax}_0$ is as close to $\mathbf{b}$ as possible.

Now it's easy to see geometrically (and we'll prove it later) that if $\mathbf{Ax}_0$ is in the column space of $\mathbf{A}$ and lies as close to $\mathbf{b}$ as possible, then the vector $\mathbf{Ax}_0 - \mathbf{b}$ must be orthogonal to every single vector in the column space of $\mathbf{A}$. Thus $\mathbf{x}_0$ has to satisfy

$$(\mathbf{Ax}_0 - \mathbf{b}) \cdot (\mathbf{Ax}) = 0 \tag{8}$$

for EVERY vector $\mathbf{x}$ in $\mathbb{R}^3$. The task now is to deduce what this tells us about $\mathbf{x}_0$. But first, we need to cast the dot product in terms of matrix multiplication.

The Transpose of a Matrix

If $\mathbf{A}$ is an m by n matrix with entries a_{ij} then the *transpose* of $\mathbf{A}$, written $\mathbf{A}^T$ is the n by m matrix with entries a_{ji}.

Problems 8

- Write out the transpose of the matrix $\mathbf{A}$ and the vector $\mathbf{b}$ in equation (7).
- Consider two vectors
$$\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}, \quad \mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix}$$
in $\mathbb{R}^n$. Write out $\mathbf{v}^T\mathbf{w}$ and $\mathbf{w}^T\mathbf{v}$.

The main reason for introducing the transpose is that it gives us an easy way to cast the dot product in terms of matrix multiplication. In fact, in the last problem you figured out how to do this: $\mathbf{v} \cdot \mathbf{w} = \mathbf{v}^T\mathbf{w} = \mathbf{w}\mathbf{v}^T$.

There's one last fact we need about transposes. You should recall that for matrices we have $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. A similar fact holds for the transpose: $(\mathbf{AB})^T = \mathbf{B}^T\mathbf{A}^T$.

Problem 9:

- Suppose that $\mathbf{A}$ is m by n and $\mathbf{B}$ is n by r, so $\mathbf{AB}$ can be formed. What are the dimensions of $\mathbf{A}^T$ and $\mathbf{B}^T$? Verify that the product $\mathbf{B}^T\mathbf{A}^T$ can be formed.

To prove the transpose product formula, start with the brutally rigorous definition of matrix multiplication, specifically, if $\mathbf{C} = \mathbf{AB}$ (where $\mathbf{A}$ is m by n and $\mathbf{B}$ is n by r) then the i,jth entry of $\mathbf{C}$ is

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

for $1 \leq i \leq m, 1 \leq j \leq r$. Let $\tilde{c}_{ij}$ denote the entries of $\mathbf{C}^T$, so $\tilde{c}_{ij} = c_{ji}$. The transpose $\mathbf{C}^T$ thus has entries obtained by interchanging the i and j's in the above summation,

$$\tilde{c}_{ij} = \sum_{k=1}^{n} a_{jk}b_{ki}. \tag{9}$$

Now compare this to the product $\mathbf{D} = \mathbf{B}^T\mathbf{A}^T$. Let $\tilde{b}_{ij}$ denote the entries of $\mathbf{B}^T$ and $\tilde{a}_{ij}$ denote the entries of $\mathbf{A}^T$. Of course then $\tilde{b}_{ij} = b_{ji}$ and $\tilde{a}_{ij} = a_{ji}$. If we write d_{ij} for the entries of $\mathbf{D} = \mathbf{B}^T\mathbf{A}^T$ then

$$d_{ij} = \sum_{k=1}^{n} \tilde{b}_{ik}\tilde{a}_{kj} = \sum_{k=1}^{n} b_{ki}a_{jk} \tag{10}$$

for $1 \leq i \leq r, 1 \leq j \leq m$, which are exactly the entries $\tilde{c}_{ij}$ of $(\mathbf{AB})^T$.

Problem 10:

- Suppose that $\mathbf{v}$ is some vector in $\mathbb{R}^n$ and $\mathbf{v}^T\mathbf{w} = 0$ for EVERY vector $\mathbf{w}$ in $\mathbb{R}^n$. Explain why $\mathbf{v}$ must in fact be the zero vector.

Back to Least Squares

In light of our work with the transpose, we can write equation (8) as

$$(\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}_0 - \mathbf{b}) = 0.$$

But since $(\mathbf{A}\mathbf{x})^T = \mathbf{x}^T\mathbf{A}^T$, this is really

$$\mathbf{x}^T\mathbf{A}^T(\mathbf{A}\mathbf{x}_0 - \mathbf{b}) = 0. \tag{11}$$

Now remember, this must hold for EVERY vector $\mathbf{x}$ in $\mathbb{R}^n$. The quantity $\mathbf{A}^T(\mathbf{A}\mathbf{x}_0 - \mathbf{b})$ in equation (11) is just some vector $\mathbf{v}$ in $\mathbb{R}^n$; since $\mathbf{x}^T\mathbf{v} = 0$ for all $\mathbf{x}$, we conclude (based on Problem 10) that $\mathbf{v} = \mathbf{0}$, i.e., we must have

$$\mathbf{A}^T(\mathbf{A}\mathbf{x}_0 - \mathbf{b}) = \mathbf{0}.$$

or with a little rearrangement,

$$\mathbf{A}^T\mathbf{A}\mathbf{x}_0 = \mathbf{A}^T\mathbf{b}. \tag{12}$$

This is a condition that $\mathbf{x}_0$ must satisfy if it is to minimize the difference $\mathbf{A}\mathbf{x}_0 - \mathbf{b}$. If the matrix $\mathbf{A}^T\mathbf{A}$ is invertible, we can find the least squares solution $\mathbf{x}_0$ to $\mathbf{A}\mathbf{x} = \mathbf{b}$.

Example: Let's go back to the polynomial curve fitting problem as the start of this section, embodied in equation (7). You can easily compute that

$$\mathbf{A}^T\mathbf{A} = \begin{bmatrix} 5 & 2 & 12 \\ 2 & 12 & 26 \\ 12 & 26 & 84 \end{bmatrix}, \quad \mathbf{A}^T\mathbf{b} = \begin{bmatrix} 17 \\ 15 \\ 71 \end{bmatrix}.$$

You can solve $(\mathbf{A}^T\mathbf{A})\mathbf{x} = \mathbf{A}^T\mathbf{b}$ to find that $\mathbf{x}_0 =< 318/211, -469/422, 411/422 >$ is the least squares solution, so that $p(x) = \frac{318}{211} - \frac{496}{422} + \frac{411}{422}x^2$ is the quadratic polynomial which does the best job (in the sense of least squares) to fitting the data points.

Problems 11:

- If $\mathbf{A}$ is m by n (so $\mathbf{x}_0$ is in $\mathbb{R}^n$, $\mathbf{b}$ is in $\mathbb{R}^m$) check that everything in equation (12) is dimensionally compatible.

- Find the line $f(x) = a+bx$ which does the best job of fitting the points $(0,1), (1,1), (3,5), (4,4)$, and $(5,4)$. Plot the line and these points.

- Find that cubic polynomial $p(x) = a + bx + cx^2 + dx^3$ which best fits the data in the previous problem.

One last word: it sometimes occurs that $\mathbf{A}^T\mathbf{A}$ is NOT invertible, meaning it has a non-trivial nullspace. At first glance, this might seem to suggest that the system $\mathbf{A}^T\mathbf{A}\mathbf{x} = \mathbf{A}^T\mathbf{b}$ has no solution, but in fact it always does (if you believe the figure above, there has to be *some* point in the column space of $\mathbf{A}$ which is closest to $\mathbf{b}$). What does happen in this case is that there are an infinite number of least squares solutions, any one as good as another, at least from the point of view of minimizing the difference $\mathbf{A}\mathbf{x} - \mathbf{b}$. In this case one often chooses that least squares solutions which has minimum length.