Week 10

Vector Spaces, Orthogonality, and Linear Least Squares

10.1 Opening Remarks

10.1.1 Visualizing Planes, Lines, and Solutions

Consider the following system of linear equations from the opener for Week 9:

$$\begin{array}{rcrcrcr} \chi_0 & - & 2\chi_1 & + & 4\chi_2 & = & -1 \\ \chi_0 & & & & & = & 2 \\ \chi_0 & + & 2\chi_1 & + & 4\chi_2 & = & 3 \end{array}$$

We solved this to find the (unique) solution

$$\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix}$$

Let us look at each of these equations one at a time, and then put them together.

Example 10.1 Find the general solution to

$$\begin{array}{ccccccc} \chi_0 & - & 2\chi_1 & + & 4\chi_2 & = & -1 \end{array}$$

We can write this as an appended system:

$$\left(\begin{array}{ccc|c} 1 & -2 & 4 & -1 \end{array} \right).$$

Now, we would perform Gaussian or Gauss-Jordan elimination with this, except that there really isn't anything to do, other than to identify the pivot, the free variables, and the dependent variables:

$$\left(\begin{array}{ccc|c} \boxed{1} & -2 & 4 & -1 \end{array} \right).$$
$$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ \text{dependent variable} & \text{free variable} & \text{free variable} \end{array}$$

Here the pivot is highlighted with the box. There are two free variables, χ_1 and χ_2, and there is one dependent variable, χ_0. To find a specific solution, we can set χ_1 and χ_2 to any value, and solve for χ_0. Setting $\chi_1 = \chi_2 = 0$ is particularly convenient, leaving us with $\chi_0 - 2(0) + 4(0) = -1$, or $\chi_0 = -1$, so that the specific solution is given by

$$x_s = \begin{pmatrix} \boxed{\text{-1}} \\ 0 \\ 0 \end{pmatrix}.$$

To find solutions (a basis) in the null space, we look for solutions of $\left(\begin{array}{ccc|c} \boxed{1} & -2 & 4 & 0 \end{array} \right)$ in the form

$$x_{n_0} = \begin{pmatrix} \boxed{\chi_0} \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad x_{n_1} = \begin{pmatrix} \boxed{\chi_0} \\ 0 \\ 1 \end{pmatrix}$$

which yields the vectors

$$x_{n_0} = \begin{pmatrix} \boxed{2} \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad x_{n_1} = \begin{pmatrix} \boxed{\text{-4}} \\ 0 \\ 1 \end{pmatrix}.$$

This then gives us the general solution

$$\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \end{pmatrix} = x_s + \beta_0 x_{n_0} + \beta_1 x_{n_1} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}.$$

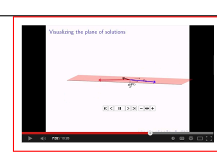
Homework 10.1.1.1 Consider, again, the equation from the last example:

$$\chi_0 \quad - \quad 2\chi_1 \quad + \quad 4\chi_2 \quad = \quad -1$$

Which of the following represent(s) a general solution to this equation? (Mark all)

- $\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}.$
- $\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix} + \beta_0 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}.$
- $\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} -5 \\ 0 \\ 1 \end{pmatrix} + \beta_0 \begin{pmatrix} 2 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}.$

The following video helps you visualize the results from the above exercise:



☛ View at edX

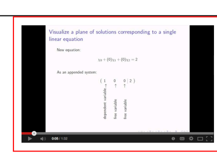
Homework 10.1.1.2 Now you find the general solution for the **second** equation in the system of linear equations with which we started this unit. Consider

$$\chi_0 \qquad\qquad = \quad 2$$

Which of the following is a true statement about this equation:

- $\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is a general solution.
- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix} + \beta_0 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$ is a general solution.
- $\begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 \\ 0 \\ 2 \end{pmatrix}$ is a general solution.

The following video helps you visualize the message in the above exercise:



☛ View at edX

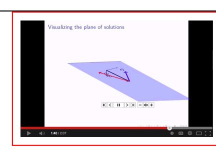
Homework 10.1.1.3 Now you find the general solution for the **third** equation in the system of linear equations with which we started this unit. Consider

$$\chi_0 + 2\chi_1 + 4\chi_2 = 3$$

Which of the following is a true statement about this equation:

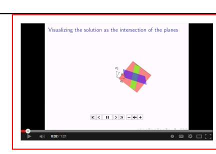
- $\begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}$ is a general solution.
- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix} + \beta_0 \begin{pmatrix} -2 \\ 1 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}$ is a general solution.
- $\begin{pmatrix} 3 \\ 0 \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} -4 \\ 2 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}$ is a general solution.

The following video helps you visualize the message in the above exercise:



☛ **View at edX**

Now, let's put the three planes together in one visualization.



☛ **View at edX**

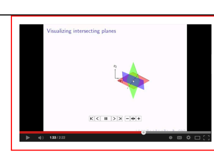
Homework 10.1.1.4 We notice that it would be nice to put lines where planes meet. Now, let's start by focusing on the first two equations: Consider

$$\begin{array}{ccccccc} \chi_0 & - & 2\chi_1 & + & 4\chi_2 & = & -1 \\ \chi_0 & & & & & = & 2 \end{array}$$

Compute the general solution of this system with two equations in three unknowns and indicate which of the following is true about this system?

- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 2 \\ 3/2 \\ 0 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 2 \\ 3/2 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ is a general solution.
- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ 2 \\ 1 \end{pmatrix}$ is a general solution.

The following video helps you visualize the message in the above exercise:



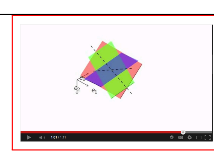
☛ **View at edX**

Homework 10.1.1.5 Similarly, consider

$$\begin{array}{rcrcrcl} \chi_0 & & & & & = & 2 \\ \chi_0 & + & 2\chi_1 & + & 4\chi_2 & = & 3 \end{array}$$

Compute the general solution of this system that has two equations with three unknowns and indicate which of the following is true about this system?

- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 2 \\ 1/2 \\ 0 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 2 \\ 1/2 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}$ is a general solution.
- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -2 \\ 1 \end{pmatrix}$ is a general solution.



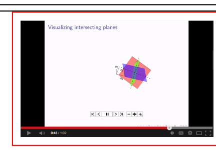
☛ **View at edX**

Homework 10.1.1.6 Finally consider

$$\begin{array}{rcrcrcr} \chi_0 & - & 2\chi_1 & + & 4\chi_2 & = & -1 \\ \chi_0 & + & 2\chi_1 & + & 4\chi_2 & = & 3 \end{array}$$

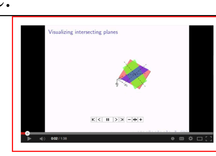
Compute the general solution of this system with two equations in three unknowns and indicate which of the following is true about this system? UPDATE

- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$ is a specific solution.
- $\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}$ is a general solution.
- $\begin{pmatrix} 2 \\ 1 \\ -0.25 \end{pmatrix} + \beta \begin{pmatrix} -4 \\ 0 \\ 1 \end{pmatrix}$ is a general solution.



☛ View at edX

The following video helps you visualize the message in the above exercise:



☛ View at edX

10.1.2 Outline

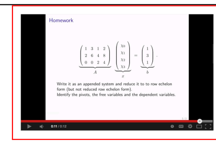
10.1.3 What You Will Learn

Upon completion of this unit, you should be able to

- Determine when linear systems of equations have a unique solution, an infinite number of solutions, or only approximate solutions.
- Determine the row-echelon form of a system of linear equations or matrix and use it to
  - find the pivots,
  - decide the free and dependent variables,
  - establish specific (particular) and general (complete) solutions,
  - find a basis for the column space, the null space, and row space of a matrix,
  - determine the rank of a matrix, and/or
  - determine the dimension of the row and column space of a matrix.
- Picture and interpret the fundamental spaces of matrices and their dimensionalities.
- Indicate whether vectors are orthogonal and determine whether subspaces are orthogonal.
- Determine the null space and column space for a given matrix and connect the row space of A with the column space of A^T.
- Identify, apply, and prove simple properties of vector spaces, subspaces, null spaces and column spaces.
- Determine when a set of vectors is linearly independent by exploiting special structures. For example, relate the rows of a matrix with the columns of its transpose to determine if the matrix has linearly independent rows.
- Approximate the solution to a system of linear equations of small dimension using the method of normal equations to solve the linear least-squares problem.

10.2 How the Row Echelon Form Answers (Almost) Everything

10.2.1 Example



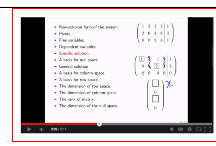
☛ View at edX

Homework 10.2.1.1 Consider the linear system of equations

$$\underbrace{\begin{pmatrix} 1 & 3 & 1 & 2 \\ 2 & 6 & 4 & 8 \\ 0 & 0 & 2 & 4 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \\ \chi_3 \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}}_{b}.$$

Write it as an appended system and reduce it to to row echelon form (but not reduced row echelon form). Identify the pivots, the free variables and the dependent variables.

10.2.2 The Important Attributes of a Linear System



☛ View at edX

We now discuss how questions about subspaces can be answered once it has been reduced to its row echelon form. In particular, you can identify:

- The row-echelon form of the system.
- The pivots.
- The free variables.
- The dependent variables.
- A specific solution
  Often called a particular solution.
- A general solution
  Often called a complete solution.
- A basis for the column space.
  Something we should have mentioned before: The column space is often called the *range* of the matrix.
- A basis for the null space.
  Something we should have mentioned before: The null space is often called the *kernel* of the matrix.
- A basis for the row space.
  The row space is the subspace of all vectors that can be created by taking linear combinations of the rows of a matrix. In other words, the row space of A equals $\mathcal{C}(A^T)$ (the column space of A^T).
- The dimension of the row and column space.
- The rank of the matrix.
- The dimension of the null space.

Motivating example

Consider the example from the last unit.

$$\underbrace{\begin{pmatrix} 1 & 3 & 1 & 2 \\ 2 & 6 & 4 & 8 \\ 0 & 0 & 2 & 4 \end{pmatrix}}_{A} \begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \\ \chi_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix}$$

which, when reduced to row echelon form, yields

$$\left(\begin{array}{cccc|c} 1 & 3 & 1 & 2 & 1 \\ 2 & 6 & 4 & 8 & 3 \\ 0 & 0 & 2 & 4 & 1 \end{array}\right) \rightarrow \left(\begin{array}{cccc|c} \boxed{1} & 3 & 1 & 2 & 1 \\ 0 & 0 & \boxed{2} & 4 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{array}\right).$$

Here the boxed entries are the pivots (the first nonzero entry in each row) and they identify that the corresponding variables (χ_0 and χ_2) are dependent variables while the other variables (χ_1 and χ_3) are free variables.

Various dimensions

Notice that inherently the matrix is $m \times n$. In this case

- $m = 3$ (the number of rows in the matrix which equals the number of equations in the linear system); and
- $n = 4$ (the number of columns in the matrix which equals the number of equations in the linear system).

Now

- There are two pivots. Let's say that in general there are k pivots, where here $k = 2$.
- There are two free variables. In general, there are $n-k$ free variables, corresponding to the columns in which no pivot reside. **This means that the null space dimension equals $n-k$, or two in this case.**
- There are two dependent variables. In general, there are k dependent variables, corresponding to the columns in which the pivots reside. **This means that the column space dimension equals k, or also two in this case. This also means that the row space dimension equals k, or also two in this case.**
- The dimension of the row space always equals the dimension of the column space which always equals the number of pivots in the row echelon form of the equation. This number, k, is called the **rank** of matrix A, $\text{rank}(A)$.

Format of a general solution

To find a general solution to problem, you recognize that there are two free variables (χ_1 and χ_3) and a general solution can be given by

$$\begin{pmatrix} \square \\ 0 \\ \square \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} \square \\ 1 \\ \square \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} \square \\ 0 \\ \square \\ 1 \end{pmatrix}.$$

Computing a specific solution

The specific (particular or special) solution is given by $x_s = \begin{pmatrix} \boxed{} \\ 0 \\ \boxed{} \\ 0 \end{pmatrix}$. It solves the system. To obtain it, you set the free variables to zero and solve the row echelon form of the system for the values in the boxes:

$$\begin{pmatrix} 1 & 3 & 1 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} \chi_0 \\ 0 \\ \chi_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

or

$$\begin{array}{rrcl} \chi_0 & +\chi_2 & = & 1 \\ & 2\chi_2 & = & 1 \end{array}$$

so that $\chi_2 = 1/2$ and $\chi_0 = 1/2$ yielding a specific solution $x_p = \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{pmatrix}$.

Computing a basis for the null space

Next, we have to find two linearly independent vectors in the null space of the matrix. (There are two because there are two free variables. In general, there are $n-k$.)

To obtain the first, we set the first free variable to one and the other(s) to zero, and solve the row echelon form of the system *with the right-hand side set to zero*:

$$\begin{pmatrix} 1 & 3 & 1 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} \chi_0 \\ 1 \\ \chi_2 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

or

$$\begin{array}{rrrcl} \chi_0 & +3\times 1 & +\chi_2 & = & 0 \\ & & 2\chi_2 & = & 0 \end{array}$$

so that $\chi_2 = 0$ and $\chi_0 = -3$, yielding the first vector in the null space $x_{n_0} = \begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \end{pmatrix}$.

To obtain the second, we set the second free variable to one and the other(s) to zero, and solve the row echelon form of the system *with the right-hand side set to zero*:

$$\begin{pmatrix} 1 & 3 & 1 & 2 \\ 0 & 0 & 2 & 4 \end{pmatrix} \begin{pmatrix} \chi_0 \\ 0 \\ \chi_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

or

$$\begin{array}{rrrcl} \chi_0 & +\chi_2 & +2\times 1 & = & 0 \\ & 2\chi_2 & +4\times 1 & = & 0 \end{array}$$

so that $\chi_2 = -4/2 = -2$ and $\chi_0 = -\chi_2 - 2 = 0$, yielding the second vector in the null space $x_{n_1} = \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}$.

Thus,

$$\mathcal{N}(A) = \text{Span}\left(\left\{\begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \end{pmatrix}\right\}\right).$$

A general solution

Thus, a general solution is given by

$$\begin{pmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{pmatrix} + \beta_0 \begin{pmatrix} -3 \\ 1 \\ 0 \\ 0 \end{pmatrix} + \beta_1 \begin{pmatrix} 0 \\ 0 \\ -2 \\ 1 \end{pmatrix},$$

where $\beta_0, \beta_1 \in \mathbb{R}$.

Finding a basis for the column space of the original matrix

To find the linearly independent columns, you look at the row echelon form of the matrix:

$$\begin{pmatrix} \boxed{1} & 3 & 1 & 2 \\ 0 & 0 & \boxed{2} & 4 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

with the pivots highlighted. The columns that have pivots in them are linearly independent. The corresponding columns in the original matrix are also linearly independent:

$$\left(\begin{array}{|c|c|c|c} 1 & 3 & 1 & 2 \\ 2 & 6 & 4 & 8 \\ 0 & 0 & 2 & 4 \end{array}\right).$$

Thus, in our example, the answer is $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}$ (the first and third column).

Thus,

$$\mathcal{C}(A) = \text{Span}\left(\left\{\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 4 \\ 2 \end{pmatrix}\right\}\right).$$

Find a basis for the row space of the matrix.

The row space (we will see in the next chapter) is the space spanned by the rows of the matrix (viewed as column vectors). Reducing a matrix to row echelon form merely takes linear combinations of the rows of the matrix. What this means is that the space spanned by the rows of the original matrix is the same space as is spanned by the rows of the matrix in row echelon form. Thus, all you need to do is list the rows in the matrix in row echelon form, as column vectors.

For our example this means a basis for the row space of the matrix is given by

$$\mathcal{R}(A) = \text{Span}\left(\left\{\begin{pmatrix} 1 \\ 3 \\ 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 2 \\ 4 \end{pmatrix}\right\}\right).$$

Summary observation

The following are all equal:

- The dimension of the column space.
- The rank of the matrix.
- The number of dependent variables.
- The number of nonzero rows in the upper echelon form.
- The number of columns in the matrix minus the number of free variables.
- The number of columns in the matrix minus the dimension of the null space.
- The number of linearly independent columns in the matrix.
- The number of linearly independent rows in the matrix.

Homework 10.2.2.1 Consider $\begin{pmatrix} 1 & 2 & 2 \\ 2 & 4 & 5 \end{pmatrix}\begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$.

- Reduce the system to row echelon form (but not reduced row echelon form).
- Identify the free variables.
- Identify the dependent variables.
- What is the dimension of the column space?
- What is the dimension of the row space?
- What is the dimension of the null space?
- Give a set of linearly independent vectors that span the column space
- Give a set of linearly independent vectors that span the row space.
- What is the rank of the matrix?
- Give a general solution.

Homework 10.2.2.2 Which of these statements is a correct definition of the rank of a given matrix $A \in \mathbb{R}^{m\times n}$?

1. The number of nonzero rows in the reduced row echelon form of A. **True/False**
2. The number of columns minus the number of rows, $n-m$. **True/False**
3. The number of columns minus the number of free columns in the row reduced form of A. (Note: a free column is a column that does not contain a pivot.) **True/False**
4. The number of 1s in the row reduced form of A. **True/False**

Homework 10.2.2.3 Compute

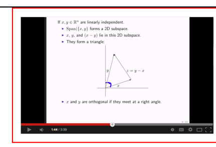
$$\begin{pmatrix} -1 \\ 2 \\ 3 \end{pmatrix} \begin{pmatrix} 3 & -1 & 2 \end{pmatrix}.$$

Reduce it to row echelon form. What is the rank of this matrix?

Homework 10.2.2.4 Let $u \in \mathbb{R}^m$ and $v \in \mathbb{R}^n$ so that uv^T is a $m \times n$ matrix. What is the rank, k, of this matrix?

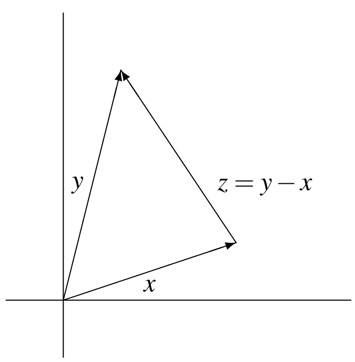
10.3 Orthogonal Vectors and Spaces

10.3.1 Orthogonal Vectors



☛ View at edX

If nonzero vectors $x, y \in \mathbb{R}^n$ are linearly independent then the subspace of all vectors $\alpha x + \beta y$, $\alpha, \beta \in \mathbb{R}$ (the space spanned by x and y) form a plane. All three vectors x, y, and $(x-y)$ lie in this plane and they form a triangle:



where this page represents the plane in which all of these vectors lie.

Vectors x and y are considered to be orthogonal (perpendicular) if they meet at a right angle. Using the Euclidean length

$$\|x\|_2 = \sqrt{\chi_0^2 + \cdots + \chi_{n-1}^2} = \sqrt{x^T x},$$

we find that the Pythagorean Theorem dictates that *if* the angle in the triangle where x and y meet is a right angle, then $\|z\|_2^2 = \|x\|_2^2 + \|y\|_2^2$. In this case,

$$\begin{aligned} \|z\|_2^2 = \|x\|_2^2 + \|y\|_2^2 &= \|y-x\|_2^2 \\ &= (y-x)^T(y-x) \\ &= (y^T - x^T)(y-x) \end{aligned}$$

$$
\begin{aligned}
&= (y^T - x^T)y - (y^T - x^T)x \\
&= \underbrace{y^T y}_{\|y\|_2^2} - \underbrace{(x^T y + y^T x)}_{2x^T y} + \underbrace{x^T x}_{\|x\|_2^2} \\
&= \|x\|_2^2 - 2x^T y + \|y\|_2^2.
\end{aligned}
$$

In other words, when x and y are perpendicular (orthogonal)

$$\|x\|_2^2 + \|y\|_2^2 = \|x\|_2^2 - 2x^T y + \|y\|_2^2.$$

Cancelling terms on the left and right of the equality, this implies that $x^T y = 0$. This motivates the following definition:

Definition 10.2 *Two vectors $x, y \in \mathbb{R}^n$ are said to be orthogonal if and only if $x^T y = 0$.*

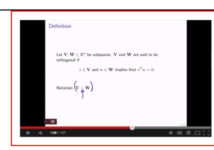
Sometimes we will use the notation $x \perp y$ to indicate that x is perpendicular to y.

Homework 10.3.1.1 For each of the following, indicate whether the vectors are orthogonal:

$\begin{pmatrix} 1 \\ -1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$	True/False
$\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $\begin{pmatrix} 0 \\ 1 \end{pmatrix}$	True/False
The unit basis vectors e_i and e_j.	Always/Sometimes/Never
$\begin{pmatrix} c \\ s \end{pmatrix}$ and $\begin{pmatrix} -s \\ c \end{pmatrix}$	Always/Sometimes/Never

Homework 10.3.1.2 Let $A \in \mathbb{R}^{m \times n}$. Let a_i^T be a row of A and $x \in \mathcal{N}(A)$. Then a_i is orthogonal to x.
Always/Sometimes/Never

10.3.2 Orthogonal Spaces



☛ **View at edX**

We can extend this to define orthogonality of two subspaces:

Definition 10.3 *Let $\mathbf{V}, \mathbf{W} \subset \mathbb{R}^n$ be subspaces. Then $\mathbf{V}$ and $\mathbf{W}$ are said to be orthogonal if and only if $v \in \mathbf{V}$ and $w \in \mathbf{W}$ implies that $v^T w = 0$.*

We will use the notation $\mathbf{V} \perp \mathbf{W}$ to indicate that subspace $\mathbf{V}$ is orthogonal to subspace $\mathbf{W}$.

In other words: Two subspaces are orthogonal if all the vectors from one of the subspaces are orthogonal to all of the vectors from the other subspace.

Homework 10.3.2.1 Let $\mathbf{V} = \{0\}$ where 0 denotes the zero vector of size n. Then $\mathbf{V} \perp \mathbb{R}^n$.
Always/Sometimes/Never

Homework 10.3.2.2 Let

$$\mathbf{V} = \mathrm{Span}\left(\left\{\begin{pmatrix}1\\0\\0\end{pmatrix},\begin{pmatrix}0\\1\\0\end{pmatrix}\right\}\right) \quad \text{and} \quad \mathbf{W} = \mathrm{Span}\left(\left\{\begin{pmatrix}0\\0\\1\end{pmatrix}\right\}\right)$$

Then $\mathbf{V} \perp \mathbf{W}$.

True/False

The above can be interpreted as: the "x-y" plane is orthogonal to the z axis.

Homework 10.3.2.3 Let $\mathbf{V}, \mathbf{W} \subset \mathbb{R}^n$ be subspaces. If $\mathbf{V} \perp \mathbf{W}$ then $\mathbf{V} \cap \mathbf{W} = \{0\}$, the zero vector.

Always/Sometimes/Never

Whenever $S \cap T = \{0\}$ we will sometimes call this the *trivial intersection* of two subspaces. Trivial in the sense that it only contains the zero vector.

Definition 10.4 *Given subspace* $\mathbf{V} \subset \mathbb{R}^n$*, the set of all vectors in* $\mathbb{R}^n$ *that are orthogonal to* $\mathbf{V}$ *is denoted by* $\mathbf{V}^\perp$ *(pronounced as "V-perp").*

Homework 10.3.2.4 If $\mathbf{V} \in \mathbb{R}^n$ is a subspace, then $\mathbf{V}^\perp$ is a subspace.

True/False

☛ **View at edX**

10.3.3 Fundamental Spaces

☛ **View at edX**

Let us recall some definitions:

- The column space of a matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{C}(\mathcal{A})$, equals the set of all vectors in $\mathbb{R}^m$ that can be written as Ax: $\{y \mid y = Ax\}$.
- The null space of a matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{N}(A)$, equals the set of all vectors in $\mathbb{R}^n$ that map to the zero vector: $\{x \mid Ax = 0\}$.
- The row space of a matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{R}(A)$, equals the set of all vectors in $\mathbb{R}^n$ that can be written as A^Tx: $\{y \mid y = A^Tx\}$.

Theorem 10.5 *Let* $A \in \mathbb{R}^{m\times n}$*. Then* $\mathcal{R}(A) \perp \mathcal{N}(A)$*.*

Proof: Let $y \in \mathcal{R}(A)$ and $z \in \mathcal{N}(A)$. We need to prove that $y^T z = 0$.

$$
\begin{array}{ll}
y^T z & \\
\quad = & < y \in \mathcal{R}(A) \text{ implies that } y = A^T x \text{ for some } x \in \mathbb{R}^m > \\
(A^T x)^T z & \\
\quad = & < (AB)^T = B^T A^T > \\
x^T (A^T)^T z & \\
\quad = & < (A^T)^T = A > \\
x^T A z & \\
\quad = & < z \in \mathcal{N}(A) \text{ implies that } Az = 0 > \\
x^T 0 & \\
\quad = & < \text{algebra} > \\
0 &
\end{array}
$$

Theorem 10.6 *Let $A \in \mathbb{R}^{m \times n}$. Then every $x \in \mathbb{R}^n$ can be written as $x = x_r + x_n$ where $x_r \in \mathcal{R}(A)$ and $x_n \in \mathcal{N}(A)$.*

Proof: Recall that if $\dim(\mathcal{R}(A) = k$, then $\dim(\mathcal{N}(A)) = n - k$. Let $\{v_0, \ldots, v_{k-1}$ be a basis for $\mathcal{R}(A)$ and $\{v_k, \ldots, v_{n-1}$ be a basis for $\mathcal{N}(A)$. It can be argued, via a proof by contradiction that is beyond this course, that the set of vectors $\{v_0, \ldots, v_{n-1}\}$ are linearly independent.

Let $x \in \mathbb{R}^n$. This is then a basis for $\mathbb{R}^n$, which in turn means that $x = \sum_{i=0} \alpha_i v_i$, some linear combination. But then

$$
x = \underbrace{\sum_{i=0}^{k-1} \alpha_i v_i}_{x_r} + \underbrace{\sum_{i=k}^{n-1} \alpha_i v_i}_{x_n},
$$

where by construction $x_r \in \mathcal{R}(A)$ and $x_n \in \mathcal{N}(A)$.

Let $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$, with $Ax = b$. Then there exist $x_r \in \mathcal{R}(A)$ and $x_n \in \mathcal{N}(A)$ such that $x = x_r + x_n$. But then

$$
\begin{aligned}
& Ax_r \\
& \quad = \quad < 0 \text{ of size } n > \\
& Ax_r + 0 \\
& \quad = \quad < Ax_n = 0 > \\
& Ax_r + Ax_n \\
& \quad = \quad < A(y+z) = Ay + Az > \\
& A(x_r + x_n) \\
& \quad = \quad < x = x_r + x_n > \\
& Ax \\
& \quad = \quad < Ax = b > \\
& b.
\end{aligned}
$$

We conclude that if $Ax = b$ has a solution, then there is a $x_r \in \mathcal{R}(A)$ such that $Ax_r = b$.

Theorem 10.7 *Let $A \in \mathbb{R}^{m \times n}$. Then A is a one-to-one, onto mapping from $\mathcal{R}(A)$ to $\mathcal{C}(\mathcal{A})$.*

Proof: Let $A \in \mathbb{R}^{m \times n}$. We need to show that

- A maps $\mathcal{R}(A)$ to $\mathcal{C}(\mathcal{A})$. This is trivial, since any vector $x \in \mathbb{R}^n$ maps to $\mathcal{C}(\mathcal{A})$.
- Uniqueness: We need to show that if $x, y \in \mathcal{R}(A)$ and $Ax = Ay$ then $x = y$. Notice that $Ax = Ay$ implies that $A(x - y) = 0$, which means that $(x - y)$ is both in $\mathcal{R}(A)$ (since it is a linear combination of x and y, both of which are in $\mathcal{R}(A)$) and in $\mathcal{N}(A)$. Since we just showed that these two spaces are orthogonal, we conclude that $(x - y) = 0$, the zero vector. Thus $x = y$.
- Onto: We need to show that for any $b \in \mathcal{C}(\mathcal{A})$ there exists $x_r \in \mathcal{R}(A)$ such that $Ax_r = b$. Notice that if $b \in \mathcal{C}$, then there exists $x \in \mathbb{R}^n$ such that $Ax = b$. By Theorem 10.6, $x = x_r + x_n$ where $x_r \in \mathcal{R}(A)$ and $x_n \in \mathcal{N}(A)$. Then $b = Ax = A(x_r + x_n) = Ax_r + Ax_n = Ax_r$.

We define one more subpace:

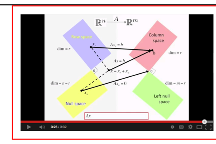
Definition 10.8 *Given $A \in \mathbb{R}^{m \times n}$ the left null space of A is the set of all vectors x such that $x^T A = 0$.*

Clearly, the left null space of A equals the null space of A^T.

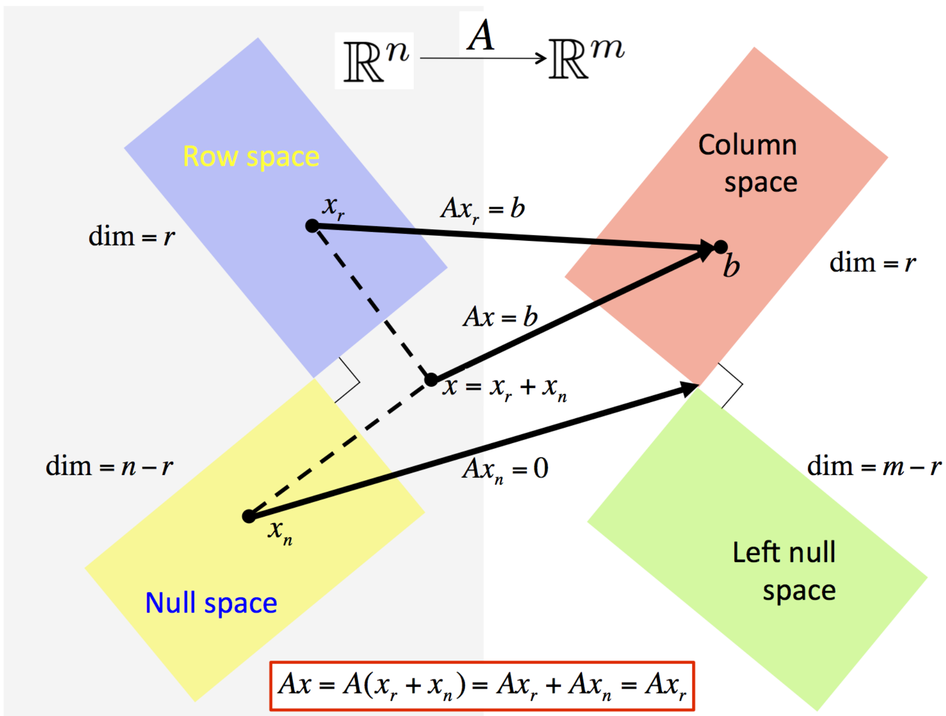
Theorem 10.9 *Let $A \in \mathbb{R}^{m \times n}$. Then the left null space of A is orthogonal to the column space of A and the dimension of the left null space of A equals $m - r$, where r is the dimension of the column space of A.*

Proof: This follows trivially by applying the previous theorems to A^T.

The observations in this unit are summarized by the following video and subsequent picture:



☛ **View at edX**

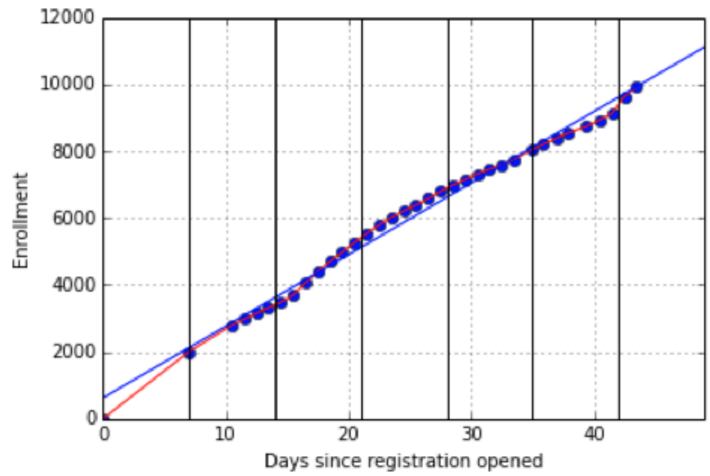


10.4 Approximating a Solution

10.4.1 A Motivating Example

☛ **View at edX**

Consider the following graph:

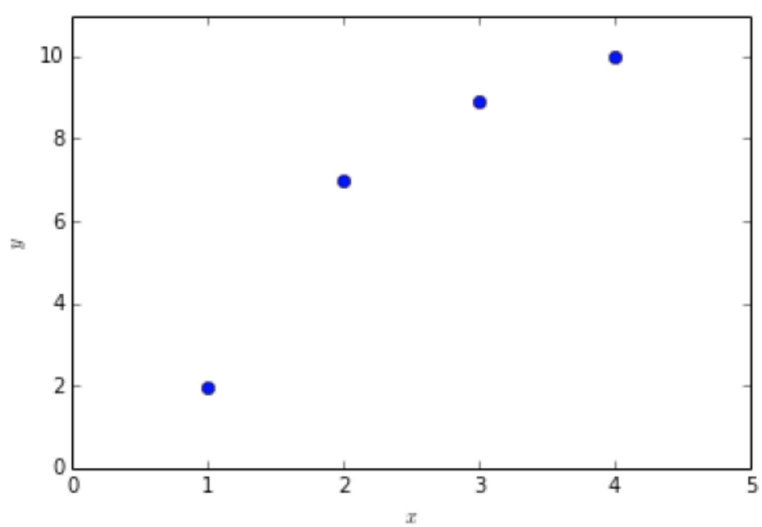


It plots the number of registrants for our "Linear Algebra - Foundations to Frontiers" course as a function of days that have passed since registration opened (data for the first offering of LAFF in Spring 2014), for the first 45 days or so (the course opens after 107 days). The blue dots represent the measured data and the blue line is the best straight line fit (which we will later call the linear least-squares fit to the data). By fitting this line, we can, for example, extrapolate that we will likely have more than 20,000 participants by the time the course commences.

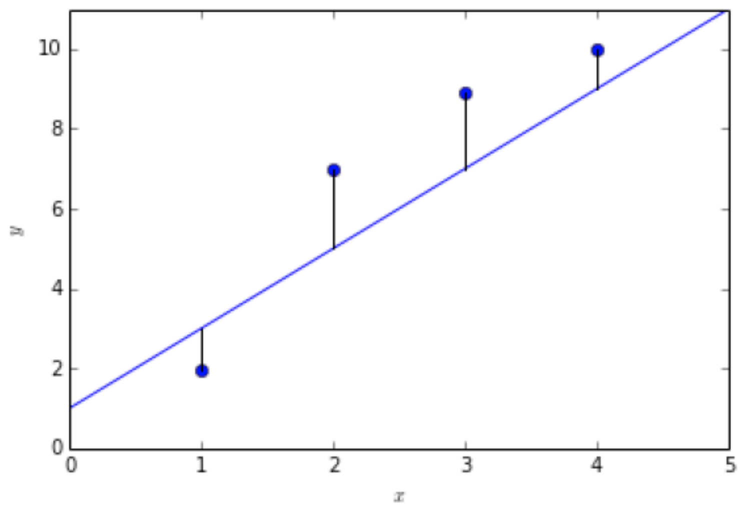
Let us illustrate the basic principles with a simpler, artificial example. Consider the following set of points:

$$(\chi_0, \psi_0) = (1, 1.97), (\chi_1, \psi_1) = (2, 6.97), (\chi_2, \psi_2) = (3, 8.89), (\chi_3, \psi_3) = (4, 10.01),$$

which we plot in the following figure:



What we would like to do is to find a line that interpolates these points. Here is a rough approximation for such a line:



Here we show with the vertical lines the distance from the points to the line that was chosen. The question becomes, what is the best line? We will see that "best" is defined in terms of minimizing the sum of the square of the distances to the line. The above line does **not** appear to be "best", and it isn't.

Let us express this with matrices and vectors. Let

$$x = \begin{pmatrix} \chi_0 \\ \chi_1 \\ \chi_2 \\ \chi_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \quad \text{and} \quad y = \begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} 1.97 \\ 6.97 \\ 8.89 \\ 10.01 \end{pmatrix}.$$

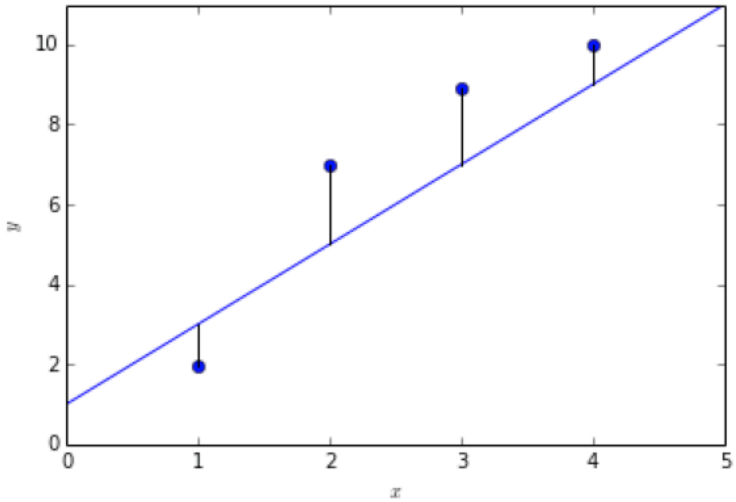
If we give the equation of the line as $y = \gamma_0 + \gamma_1 x$ then, **IF** this line **COULD** go through all these points **THEN** the following equations would have to be simultaneously satified:

$$\begin{array}{rcl} \psi_0 & = & \gamma_0 + \gamma_1 \chi_1 \\ \psi_1 & = & \gamma_0 + \gamma_1 \chi_2 \\ \psi_2 & = & \gamma_0 + \gamma_1 \chi_3 \\ \psi_3 & = & \gamma_0 + \gamma_1 \chi_4 \end{array} \quad \text{or, specifically,} \quad \begin{array}{rcl} 1.97 & = & \gamma_0 + \gamma_1 \\ 6.97 & = & \gamma_0 + 2\gamma_1 \\ 8.89 & = & \gamma_0 + 3\gamma_1 \\ 10.01 & = & \gamma_0 + 4\gamma_1 \end{array}$$

which can be written in matrix notation as

$$\begin{pmatrix} \psi_0 \\ \psi_1 \\ \psi_2 \\ \psi_3 \end{pmatrix} = \begin{pmatrix} 1 & \chi_0 \\ 1 & \chi_1 \\ 1 & \chi_2 \\ 1 & \chi_3 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} \quad \text{or, specifically,} \quad \begin{pmatrix} 1.97 \\ 6.97 \\ 8.89 \\ 10.01 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix}.$$

Now, just looking at



it is obvious that these points do not lie on the same line and that therefore all these equations cannot be simultaneously satified. **So, what do we do now?**

How does it relate to column spaces?

The first question we ask is "For what right-hand sides could we have solved all four equations simultaneously?" We would have had to choose y so that $Ac = y$, where

$$A = \begin{pmatrix} 1 & \chi_0 \\ 1 & \chi_1 \\ 1 & \chi_2 \\ 1 & \chi_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ 1 & 3 \\ 1 & 4 \end{pmatrix} \quad \text{and} \quad c = \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix}.$$

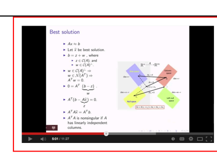
This means that y must be in the column space of A. It must be possible to express it as $y = \gamma_0 a_0 + \gamma_1 a_1$, where $A = \left(\begin{array}{c|c} a_0 & a_1 \end{array} \right)$! What does this mean if we relate this back to the picture? Only if $\{\psi_0, \cdots, \psi_3\}$ have the property that $\{(1, \psi_0), \cdots, (4, \psi_3)\}$ lie on a line can we find coefficients γ_0 and γ_1 such that $Ac = y$.

How does this problem relate to orthogonality?

The problem is that the given y does **not** lie in the column space of A. So a question is, what vector z, that **does** lie in the column space should we use to solve $Ac = z$ instead so that we end up with a line that best interpolates the given points?

If z solves $Ac = z$ exactly, then $z = \left(\begin{array}{c|c} a_0 & a_1 \end{array} \right) \begin{pmatrix} \gamma_0 \\ \gamma_1 \end{pmatrix} = \gamma_0 a_0 + \gamma_1 a_1$, which is of course just a repeat of the observation that z is in the column space of A. Thus, what we want is $y = z + w$, where w is as small (in length) as possible. This happens when w is orthogonal to z! So, $y = \gamma_0 a_0 + \gamma_1 a_1 + w$, with $a_0^T w = a_1^T w = 0$. The vector z in the column space of A that is closest to y is known as the **projection** of y onto the column space of A. So, it would be nice to have a way of finding a way to compute this projection.

10.4.2 Finding the Best Solution



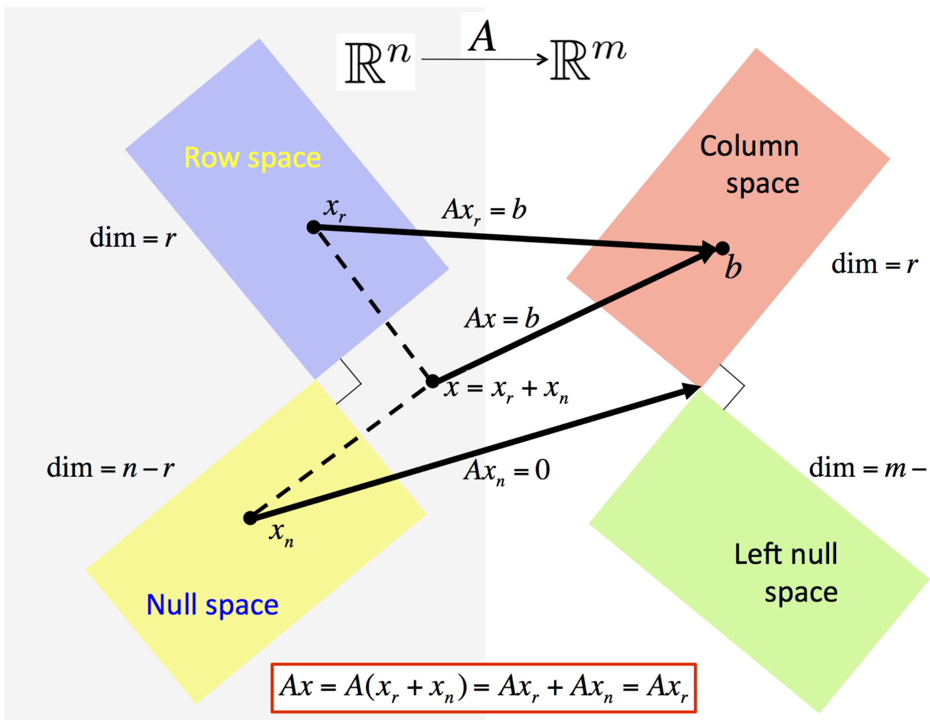
☛ View at edX

The last problem motivated the following general problem: Given m equations in n unknowns, we end up with a system $Ax = b$ where $A \in \mathbb{R}^{m \times n}$, $x \in \mathbb{R}^n$, and $b \in \mathbb{R}^m$.

- This system of equations may have no solutions. This happens when b is not in the column space of A.
- This system may have a unique solution. This happens only when $r = m = n$, where r is the rank of the matrix (the dimension of the column space of A). Another way of saying this is that it happens only if A is square and nonsingular (it has an inverse).
- This system may have many solutions. This happens when b is in the column space of A and $r < n$ (the columns of A are linearly dependent, so that the null space of A is nontrivial).

Let us focus on the first case: b is not in the column space of A.

In the last unit, we argued that what we want is an approximate solution $\hat{x}$ such that $A\hat{x} = z$, where z is the vector in the column space of A that is "closest" to b: $b = z + w$ where $w^T v = 0$ for all $v \in \mathcal{C}(\mathcal{A})$. From



we conclude that this means that w is in the left null space of A. So, $A^T w = 0$. But that means that

$$0 = A^T w = A^T(b - z) = A^T(b - A\hat{x})$$

which we can rewrite as

$$A^T A\hat{x} = A^T b. \tag{10.1}$$

This is known as the **normal equation** associated with the problem $A\hat{x} \approx b$.

Theorem 10.10 *If $A \in \mathbb{R}^{m\times n}$ has linearly independent columns, then $A^T A$ is nonsingular (equivalently, has an inverse, $A^T A\hat{x} = A^T b$ has a solution for all b, etc.).*

Proof: Proof by contradiction.

- Assume that $A \in \mathbb{R}^{m\times n}$ has linearly independent columns and $A^T A$ is singular.
- Then there exists $x \neq 0$ such that $A^T Ax = 0$.

- Hence, there exists $y = Ax \neq 0$ such that $A^T y = 0$ (because A has linearly independent columns and $x \neq 0$).
- This means y is in the left null space of A.
- But y is also in the column space of A, since $Ax = y$.
- Thus, $y = 0$, since the intersection of the column space of A and the left null space of A only contains the zero vector.
- This contradicts the fact that A has linearly independent columns.

Therefore $A^T A$ cannot be singular.

This means that if A has linearly independent columns, then the desired $\hat{x}$ that is the best approximate solution is given by

$$\hat{x} = (A^T A)^{-1} A^T b$$

and the vector $z \in \mathcal{C}(\mathcal{A})$ closest to b is given by

$$z = A\hat{x} = A(A^T A)^{-1} A^T b.$$

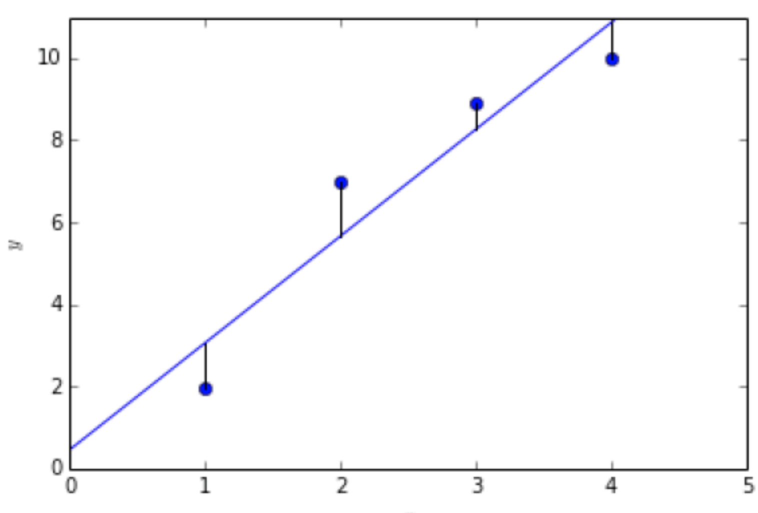
This shows that *if* A has linearly independent columns, then $z = A(A^T A)^{-1} A^T b$ is the vector in the columns space closest to b. **This is the projection of b onto the column space of A.**

Let us now formulate the above observations as a special case of a *linear least-squares* problem:

Theorem 10.11 *Let $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, and $x \in \mathbb{R}^n$ and assume that A has linearly independent columns. Then the solution that minimizes the length of the vector $b - Ax$ is given by $\hat{x} = (A^T A)^{-1} A^T b$.*

Definition 10.12 *Let $A \in \mathbb{R}^{m \times n}$. If A has linearly independent columns, then $A^\dagger = (A^T A)^{-1} A^T$ is called the (left) pseudo inverse. Note that this means $m \geq n$ and $A^\dagger A = (A^T A)^{-1} A^T A = I$.*

If we apply these insights to the motivating example from the last unit, we get the following approximating line



Homework 10.4.2.1 Consider $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$.

1. Is b in the column space of A?

   True/False

2. $A^T b =$
3. $A^T A =$
4. $(A^T A)^{-1} =$
5. $A^\dagger =$.
6. $A^\dagger A =$.
7. Compute the approximate solution, in the least squares sense, of $Ax \approx b$.

   $x = \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} =$

8. What is the project of b onto the column space of A?

   $\widehat{b} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} =$

Homework 10.4.2.2 Consider $A = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix}$ and $b = \begin{pmatrix} 4 \\ 5 \\ 9 \end{pmatrix}$.

1. b is in the column space of A, $\mathcal{C}(A)$.

   True/False

2. Compute the approximate solution, in the least squares sense, of $Ax \approx b$.

   $x = \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} =$

3. What is the project of b onto the column space of A?

   $\widehat{b} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} =$

4. $A^\dagger =$.
5. $A^\dagger A =$.

Homework 10.4.2.3 What 2×2 matrix B projects the x-y plane onto the line $x + y = 0$?

Homework 10.4.2.4 Find the line that best fits the following data:

x	y
-1	2
1	-3
0	0
2	-5

Homework 10.4.2.5 Consider $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ -2 & 4 \end{pmatrix}$ and $b = \begin{pmatrix} 1 \\ 2 \\ 7 \end{pmatrix}$.

1. b is in the column space of A, $\mathcal{C}(A)$.

   True/False

2. Compute the approximate solution, in the least squares sense, of $Ax \approx b$.

   $x = \begin{pmatrix} \chi_0 \\ \chi_1 \end{pmatrix} =$

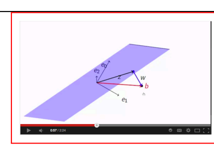
3. What is the projection of b onto the column space of A?

   $\widehat{b} = \begin{pmatrix} \widehat{\beta}_0 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} =$

4. $A^\dagger =$.

5. $A^\dagger A =$.

10.4.3 Why It is Called Linear Least-Squares



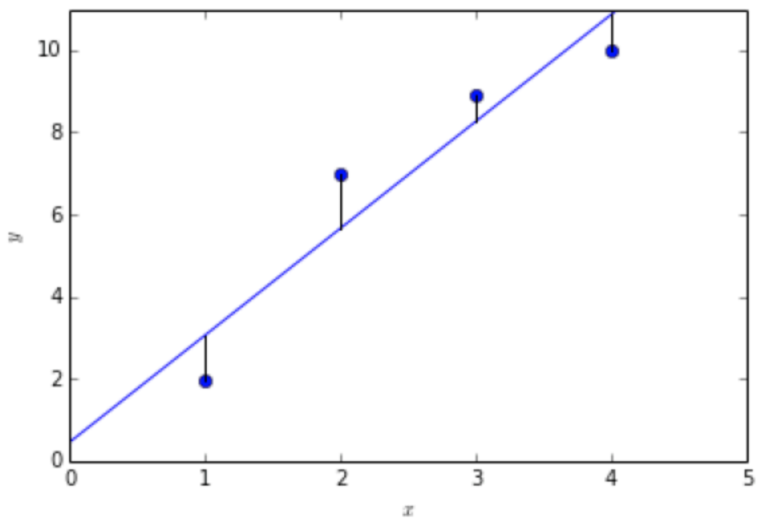
☛ **View at edX**

The "best" solution discussed in the last unit is known as the "linear least-squares" solution. Why?

Notice that we are trying to find $\hat{x}$ that minimizes the length of the vector $b - Ax$. In other words, we wish to find $\hat{x}$ that minimizes $\min_x \|b - Ax\|_2$. Now, if $\hat{x}$ minimizes $\min_x \|b - Ax\|_2$, it also minimizes the function $\|b - Ax\|_2^2$. Let $y = A\hat{x}$. Then

$$\|b - A\hat{x}\|^2 = \|b - y\|^2 = \sum_{i=0}^{n-1} (\beta_i - \psi_i)^2.$$

Thus, we are trying to minimize the sum of the squares of the differences. If you consider, again,



then this translates to minimizing the sum of the lenghts of the vertical lines that connect the linear approximation to the original points.

10.5 Enrichment

10.5.1 Solving the Normal Equations

In our examples and exercises, we solved the normal equations

$$A^T A x = A^T b,$$

where $A \in \mathbb{R}^{m \times n}$ has linear independent columns, via the following steps:

- Form $y = A^T b$
- Form $A^T A$.
- Invert $A^T A$ to compute $B = (A^T A)^{-1}$.
- Compute $\hat{x} = By = (A^T A)^{-1} A^T b$.

This involves the inversion of a matrix, and we claimed in Week 8 that one should (almost) never, ever invert a matrix.

In practice, this is not how it is done for larger systems of equations. Instead, one uses either the Cholesky factorization (which was discussed in the enrichment for Week 8), the QR factorization (to be discussed in Week 11), or the Singular Value Decomposition (SVD, which is briefly mentioned in Week 11).

Let us focus on how to use the Cholesky factorization. Here are the steps:

- Compute $C = A^T A$.
- Compute the Cholesky factorization $C = LL^T$, where L is lower triangular.
  This allows us to take advantage of symmetry in C.
- Compute $y = A^T b$.
- Solve $Lz = y$.
- Solve $L^T \hat{x} = z$.

The vector $\hat{x}$ is then the best solution (in the linear least-squares sense) to $Ax \approx b$.

The Cholesky factorization of a matrix, C, exists if and only if C has a special property. Namely, it must be symmetric positive definite (SPD).

Definition 10.13 *A symmetric matrix $C \in \mathbb{R}^{m\times m}$ is said to be symmetric positive definite if $x^T Cx \geq 0$ for all nonzero vectors $x \in \mathbb{R}^m$.*

We started by assuming that A has linearly independent columns and that $C = A^T A$. Clearly, C is symmetric: $C^T = (A^T A)^T = A^T (A^T)^T = A^T A = C$. Now, let $x \neq 0$. Then

$$x^T Cx = x^T (A^T A)x = (x^T A^T)(Ax) = (Ax)^T (Ax) = \|Ax\|_2^2.$$

We notice that $Ax \neq 0$ because the columns of A are linearly independent. But that means that its length, $\|Ax\|_2$, is not equal to zero and hence $\|Ax\|_2^2 > 0$. We conclude that $x \neq 0$ implies that $x^T Cx > 0$ and that therefore C is symmetric positive definite.

10.6 Wrap Up

10.6.1 Homework

No additional homework this week.

10.6.2 Summary

Solving underdetermined systems

Important attributes of a linear system $Ax = b$ and associated matrix A:

- The row-echelon form of the system.
- The pivots.
- The free variables.
- The dependent variables.
- A specific solution
  Also called a *particular* solution.
- A general solution
  Also called a *complete* solution.
- A basis for the null space.
  Also called the *kernel* of the matrix. This is the set of all vectors that are mapped to the zero vector by A.
- A basis for the column space, $\mathcal{C}(A)$.
  Also called the *range* of the matrix. This is the set of linear combinations of the columns of A.
- A basis for the row space, $\mathcal{R}(A) = \mathcal{C}(A^T)$.
  This is the set of linear combinations of the columns of A^T.
- The dimension of the row and column space.
- The rank of the matrix.
- The dimension of the null space.

Various dimensions Notice that, in general, a matrix is $m \times n$. In this case

- Start the linear system of equations $Ax = y$.
- Reduce this to row echelon form $Bx = \hat{y}$.
- If any of the equations are inconsistent ($0 \neq \hat{\psi}_i$, for some row i in the row echelon form $Bx = \hat{y}$), then the system does not have a solution, and y is not in the column space of A.

- If this is not the case, assume there are k pivots in the row echelon reduced form.
- Then there are $n-k$ free variables, corresponding to the columns in which no pivots reside. **This means that the null space dimension equals** $n-k$
- There are k dependent variables corresponding to the columns in which the pivots reside. **This means that the column space dimension equals** k **and the row space dimension equals** k**.**
- The dimension of the row space always equals the dimension of the column space which always equals the number of pivots in the row echelon form of the equation, k. This number, k, is called the **rank** of matrix A, $\text{rank}(A)$.
- To find a specific (particular) solution to system $Ax = b$, set the free variables to zero and solve $Bx = \hat{y}$ for the dependent variables. Let us call this solution x_s.
- To find $n-k$ linearly independent vectors in $\mathcal{N}(A)$, follow the following procedure, assuming that $n_0, \ldots n_{n-k-1}$ equal the indices of the free variables. (In other words: $\chi_{n_0}, \ldots, \chi_{n_{n-k-1}}$ equal the free variables.)
  - Set χ_{n_j} equal to one and χ_{n_k} with $n_k \neq n_j$ equal to zero. Solve for the dependent variables.

  This yields $n-k$ linearly independent vectors that are a basis for $\mathcal{N}(A)$. Let us call these $x_{n_0}, \ldots, x_{n_{n-k-1}}$.
- The general (complete) solution is then given as

$$x_s + \gamma_0 x_{n_0} + \gamma_1 x_{n_1} + \cdots + \gamma_{n-k-1} x_{n_{n-k-1}}.$$

- To find a basis for the column space of A, $\mathcal{C}(A)$, you take the columns of A that correspond to the columns with pivots in B.
- To find a basis for the row space of A, $\mathcal{R}(A)$, you take the rows of B that contain pivots, and transpose those into the vectors that become the desired basis. (Note: you take the rows of B, not A.)
- The following are all equal:
  - The dimension of the column space.
  - The rank of the matrix.
  - The number of dependent variables.
  - The number of nonzero rows in the upper echelon form.
  - The number of columns in the matrix minus the number of free variables.
  - The number of columns in the matrix minus the dimension of the null space.
  - The number of linearly independent columns in the matrix.
  - The number of linearly independent rows in the matrix.

Orthogonal vectors

Definition 10.14 *Two vectors $x, y \in \mathbb{R}^m$ are orthogonal if and only if $x^T y = 0$.*

Orthogonal subspaces

Definition 10.15 *Two subspaces $\mathbf{V}, \mathbf{W} \subset \mathbb{R}^m$ are orthogonal if and only if $v \in \mathbf{V}$ and $w \in \mathbf{W}$ implies $v^T w = 0$.*

Definition 10.16 *Let $\mathbf{V} \subset \mathbb{R}^m$ be a subspace. Then $\mathbf{V}^\perp \subset \mathbb{R}^m$ equals the set of all vectors that are orthogonal to $\mathbf{V}$.*

Theorem 10.17 *Let $\mathbf{V} \subset \mathbb{R}^m$ be a subspace. Then $\mathbf{V}^\perp$ is a subspace of $\mathbb{R}^m$.*

The Fundamental Subspaces

- The column space of a matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{C}(\mathcal{A})$, equals the set of all vectors in $\mathbb{R}^m$ that can be written as Ax: $\{y \mid y = Ax\}$.
- The null space of a matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{N}(A)$, equals the set of all vectors in $\mathbb{R}^n$ that map to the zero vector: $\{x \mid Ax = 0\}$.
- The row space of a matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{R}(A)$, equals the set of all vectors in $\mathbb{R}^n$ that can be written as $A^T x$: $\{y \mid y = A^T x\}$.
- The left null space of a matrix $A \in \mathbb{R}^{m\times n}$, $\mathcal{N}(A^T)$, equals the set of all vectors in $\mathbb{R}^m$ described by $\{x \mid x^T A = 0\}$.

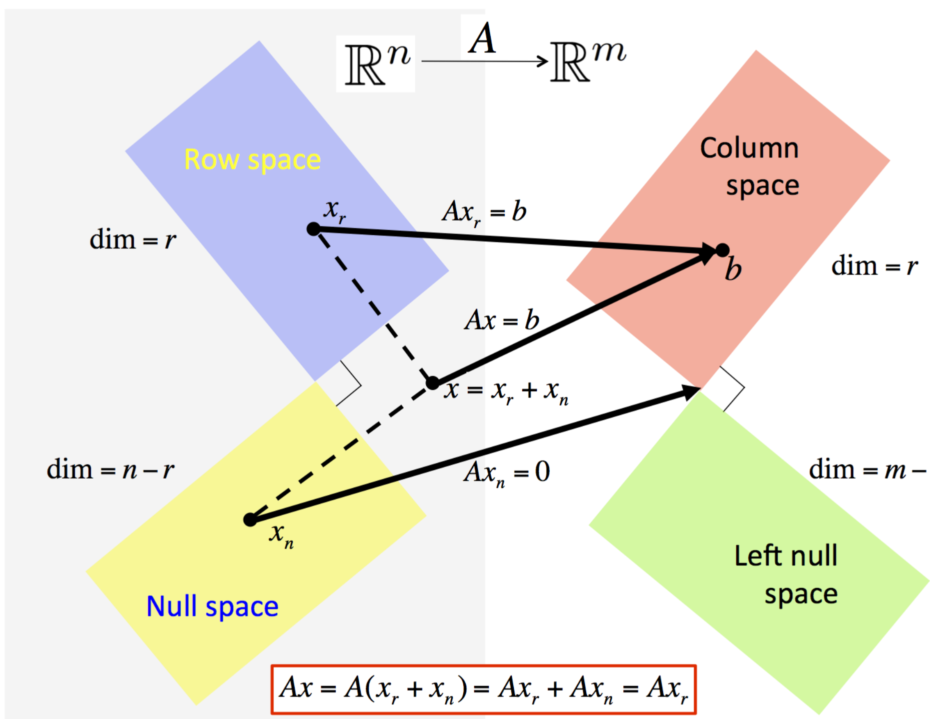
Theorem 10.18 *Let $A \in \mathbb{R}^{m\times n}$. Then $\mathcal{R}(A) \perp \mathcal{N}(A)$.*

Theorem 10.19 *Let $A \in \mathbb{R}^{m\times n}$. Then every $x \in \mathbb{R}^n$ can be written as $x = x_r + x_n$ where $x_r \in \mathcal{R}(A)$ and $x_n \in \mathcal{N}(A)$.*

Theorem 10.20 *Let $A \in \mathbb{R}^{m\times n}$. Then A is a one-to-one, onto mapping from $\mathcal{R}(A)$ to $\mathcal{C}(\mathcal{A})$.*

Theorem 10.21 *Let $A \in \mathbb{R}^{m\times n}$. Then the left null space of A is orthogonal to the column space of A and the dimension of the left null space of A equals $m - r$, where r is the dimension of the column space of A.*

An important figure:



Overdetermined systems

- $Ax = b$ has a solution if and only if $b \in \mathcal{C}(A)$.
- Let us assume that A has linearly independent columns and we wish to solve $Ax \approx b$. Then
  - The solution of the normal equations
  $$A^T Ax = A^T b$$
  is the best solution (in the linear least-squares sense) to $Ax \approx b$.
  - The pseudo inverse of A is given by $A^\dagger = (A^T A)^{-1} A^T$.

- The best solution (in the linear least-squares sense) of $Ax = b$ is given by $\hat{x} = A^{\dagger} b = (A^T A)^{-1} A^T b$.
- The orthogonal projection of b onto $\mathcal{C}(A)$ is given by $\hat{b} = A(A^T A)^{-1} A^T b$.
- The vector $(b - \hat{b})$ is the component of b orthogonal to $\mathcal{C}(A)$.
- The orthogonal projection of b onto $\mathcal{C}(A)^{\perp}$ is given by $b - \hat{b} = [I - A(A^T A)^{-1} A^T] b$.