DATA605

Week5 and Week6 Lecture

4 Subspaces of a Matrix A

Four Fundamental Subspaces

1. The ***row space*** is $C(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^n$.

2. The ***column space*** is $C(A)$, a subspace of $\mathbf{R}^m$.

3. The ***nullspace*** is $N(A)$, a subspace of $\mathbf{R}^n$.

4. The ***left nullspace*** is $N(A^{\mathrm{T}})$, a subspace of $\mathbf{R}^m$. This is our new space.

Fundamental Theorem of Linear Algebra pt 1

The column space and row space both have dimension r, which is the rank of the matrix

The nullspaces have dimensions $n - r$ and $m - r$.

Dimensions of The Spaces and Vector Mapping

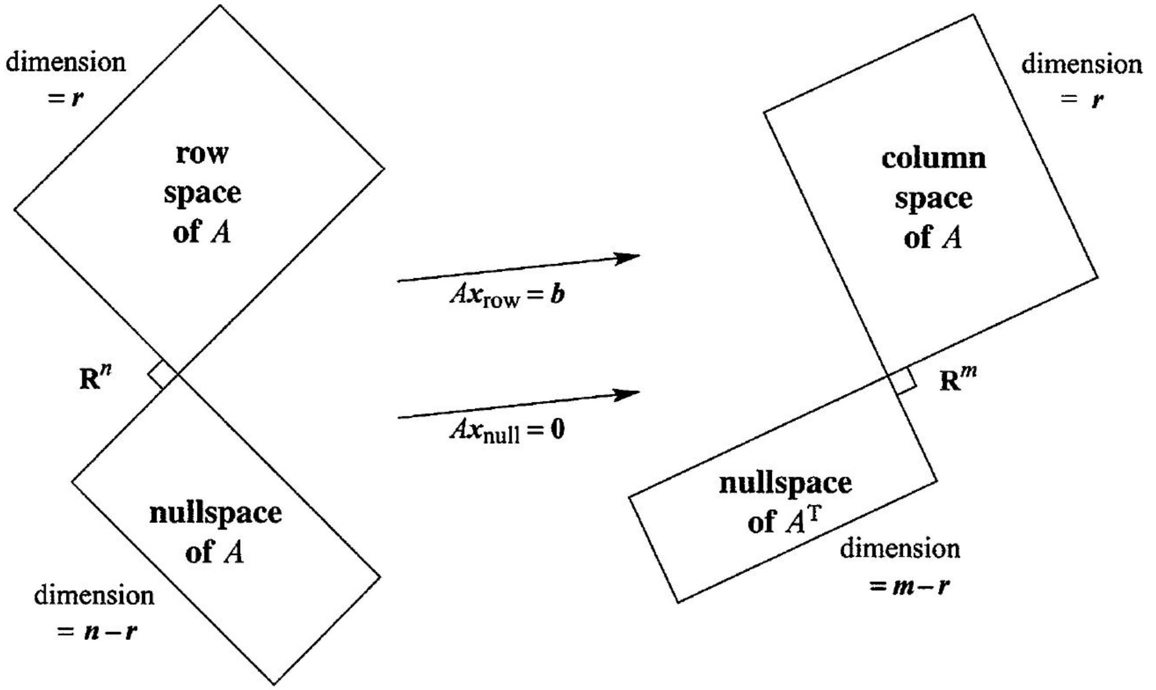


Figure 4.2: Two pairs of orthogonal subspaces. The dimensions add to n and add to m. **This is an important picture**—one pair of subspaces is in $\mathbf{R}^n$ and one pair is in $\mathbf{R}^m$.

Orthogonal Vectors

Orthogonal vectors $\quad v^{\mathrm{T}} w = 0 \quad$ and $\quad \|v\|^2 + \|w\|^2 = \|v + w\|^2.$

This tries to capture the notion of ‘perpendicular’ for any dimensional vector. If you remember the dot product is related to the cos of the angle between vectors, and when cos(angle) = 0, the angle between them is 90 degrees or orthogonal.

In some sense, no shared ‘information’ between the vectors

Vector spaces can be defined as orthogonal, too, which means all pairs of vectors are orthogonal (see diagram p196 for example)

Fundamental Theorem of Linear Algebra pt 2

The orthogonal complement of a subspace V contains every vector that is perpendicular to V.

***Fundamental Theorem of Linear Algebra*, Part 2**

$N(A)$ is the orthogonal complement of the row space $C(A^T)$ (in $\mathbf{R}^n$).

$N(A^T)$ is the orthogonal complement of the column space $C(A)$ (in $\mathbf{R}^m$).

Four Fundamental Spaces

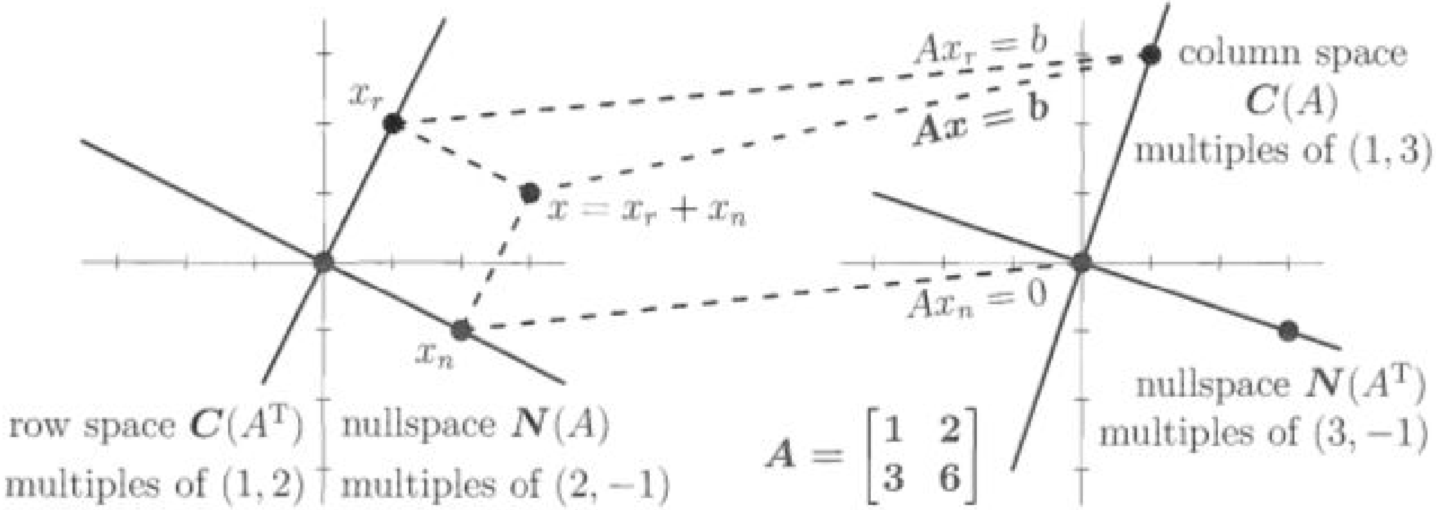


Figure 2.5: The four fundamental subspaces (lines) for the singular matrix A.

Projections

This of how you project a shadow onto the floor when there is a light source behind you. This is similar to how we think of projections in LA. We are trying to find the portion of a vector that lies inside or along some other vector space.

Projection onto One Dimensional Space

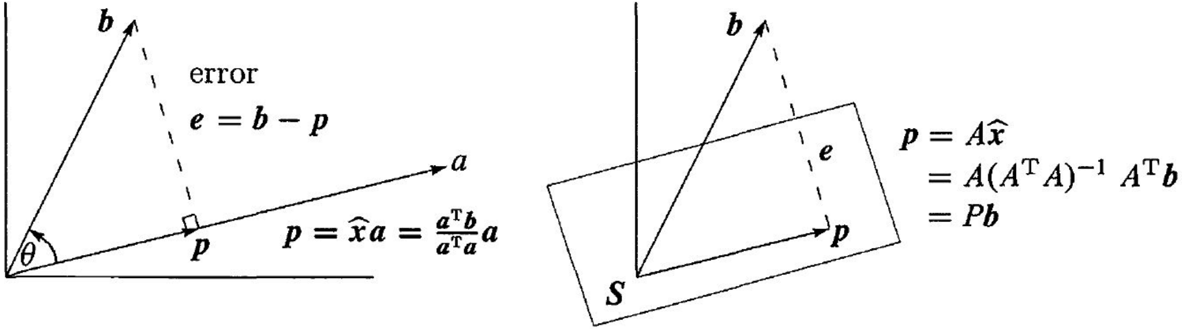


Figure 4.5: The projection p of b onto a line and onto S = column space of A.

The projection of b onto the line through a is the vector $p = \hat{x}a = \frac{a^Tb}{a^Ta}a$.

Special case 1: If $b = a$ then $\hat{x} = 1$. The projection of a onto a is itself. $Pa = a$.

Special case 2: If b is perpendicular to a then $a^Tb = 0$. The projection is $p = 0$.

Projection onto Any Space

The combination $\boldsymbol{p} = \widehat{x}_1\boldsymbol{a}_1 + \cdots + \widehat{x}_n\boldsymbol{a}_n = A\widehat{\boldsymbol{x}}$ that is closest to $\boldsymbol{b}$ comes from

$$A^{\mathrm{T}}(\boldsymbol{b} - A\widehat{\boldsymbol{x}}) = \mathbf{0} \quad \text{or} \quad A^{\mathrm{T}}A\widehat{\boldsymbol{x}} = A^{\mathrm{T}}\boldsymbol{b}. \tag{5}$$

This symmetric matrix $A^{\mathrm{T}}A$ is n by n. It is invertible if the $\boldsymbol{a}$'s are independent. The solution is $\widehat{\boldsymbol{x}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}$. The ***projection*** of $\boldsymbol{b}$ onto the subspace is $\boldsymbol{p}$:

$$\boldsymbol{p} = A\widehat{\boldsymbol{x}} = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{b}. \tag{6}$$

This formula shows the n by n ***projection matrix*** that produces $\boldsymbol{p} = P\boldsymbol{b}$:

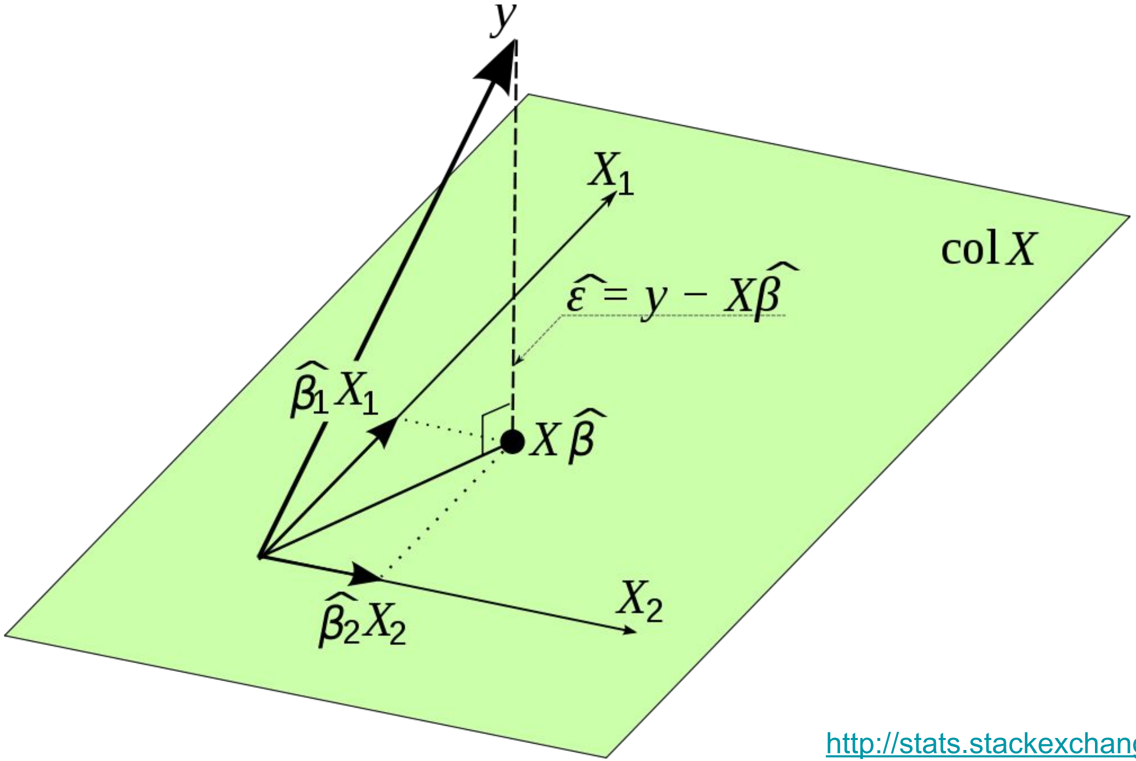
$$P = A(A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}. \tag{7}$$

Least Squares

What happens when b in Ax = b is not in the column space of A? That means we cannot solve exactly but we can come up with a closest solution. One way of going about this is geometrically: the best solution is simply b being projected onto the column space of A. This is called p, or xhat, which in the column space.

The difference between b and xhat is the error, which is orthogonal to the column space (i.e. in the left null space!). See diagram below:

Projection of y into the Column Space



http://stats.stackexchange.com/questions/108591/what-is-the-importance-of-hat-matrix-h-xx-top-x-1-x-top-in-linear-reg

Relation of Spaces when b not in column space

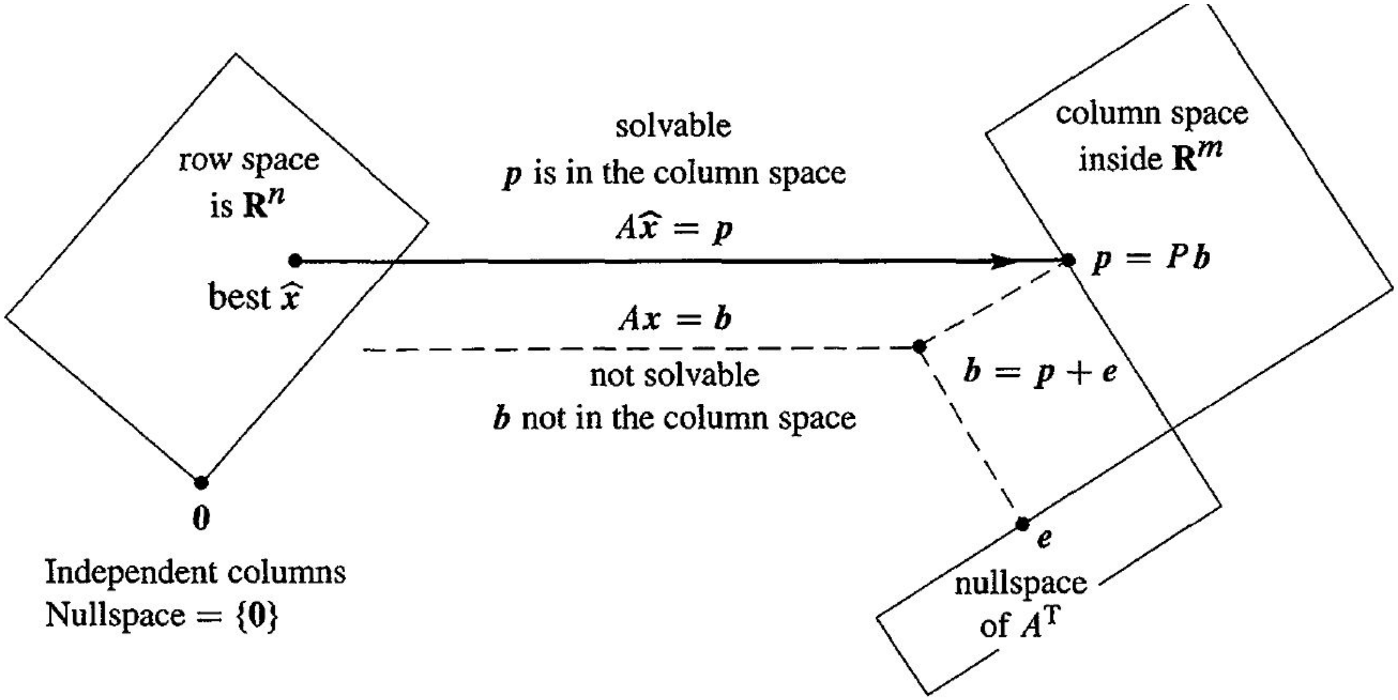


Figure 4.7: The projection $p = A\widehat{x}$ is closest to b, so $\widehat{x}$ minimizes $E = \|b - Ax\|^2$.

Now for a Change of Pace ...

We'll be leaving Linear Algebra now and heading into Probability!

Defining Terms

What is probability? “Probability is usually used to describe outcomes of events that take multiple values and cannot be stated with certainty that a particular outcome will occur at a particular experiment”

A trial refers to a single event such as a coin toss or a roll of a die.

The elementary events of a trail are the possible outcomes of that trial.

Independence means that one coin toss doesn’t tell us anything about subsequent coin tosses. We can say that the first coin toss and the second one are independent. No connective glue of ‘causal dependence’

Random Variables

A random variable is a variable whose value is subject to some randomness or chance. The ‘sample space’ is the space of all possible events of this variable.

Come in two forms: Discrete vs Continuous Random Variables

Discrete: we are restricting ourselves to random variables whose underlying events come from a finite space of possibilities or whose values take countably infinite set of discrete values.

Continuous: Sample space is essentially the real numbers (i.e temperature). Doesn’t make sense to think of probability of a single real number, but of a range of them

Whats Probability?

Associated with the random variable is the probability of a particular value occurring for that variable. For random variable Y, its written as

Pr(set of Y values)

Which is defined as: (number of possible events leading to Y values) / (# total possible Y values)

Example: Similarly, we can have a random variable Y that represents the sum of the faces of two dies rolled together. So, Y can takes values between 2 and 12. We can compute the probability of Y == 6, for instance. Pr(Y == 6) is the probability of getting a sum of 6 by rolling two dies together.

Multiple Events

For **independent** events:

We can compute Pr(Y, Z) the probability of events Y and Z both occurring. This of this like an ‘and’ or a set ‘intersection’. This evals to Pr(Y)Pr(Z)

We can compute Pr(Y U Z) the probability of events Y or Z occurring. This of this like an ‘or’, or a set union. This evals to Pr(Y) + Pr(Z)

Summing to 1

The probabilities across all members of the sample space must all add up to 1.

Trick: Sometimes its easier to compute a ‘complimentary’ probability. If you need to know the probability of an event of a lot of members of the sample space, it might be easier to do 1 - Pr(X’), where X’ is the set complement of X.

Example: Probability of rolling >= 3 can be computing by summing probabilities for 3,4,5 and 6. You could also do Pr(X >= 3) = 1 - Pr(x < 3), which is simpler to compute