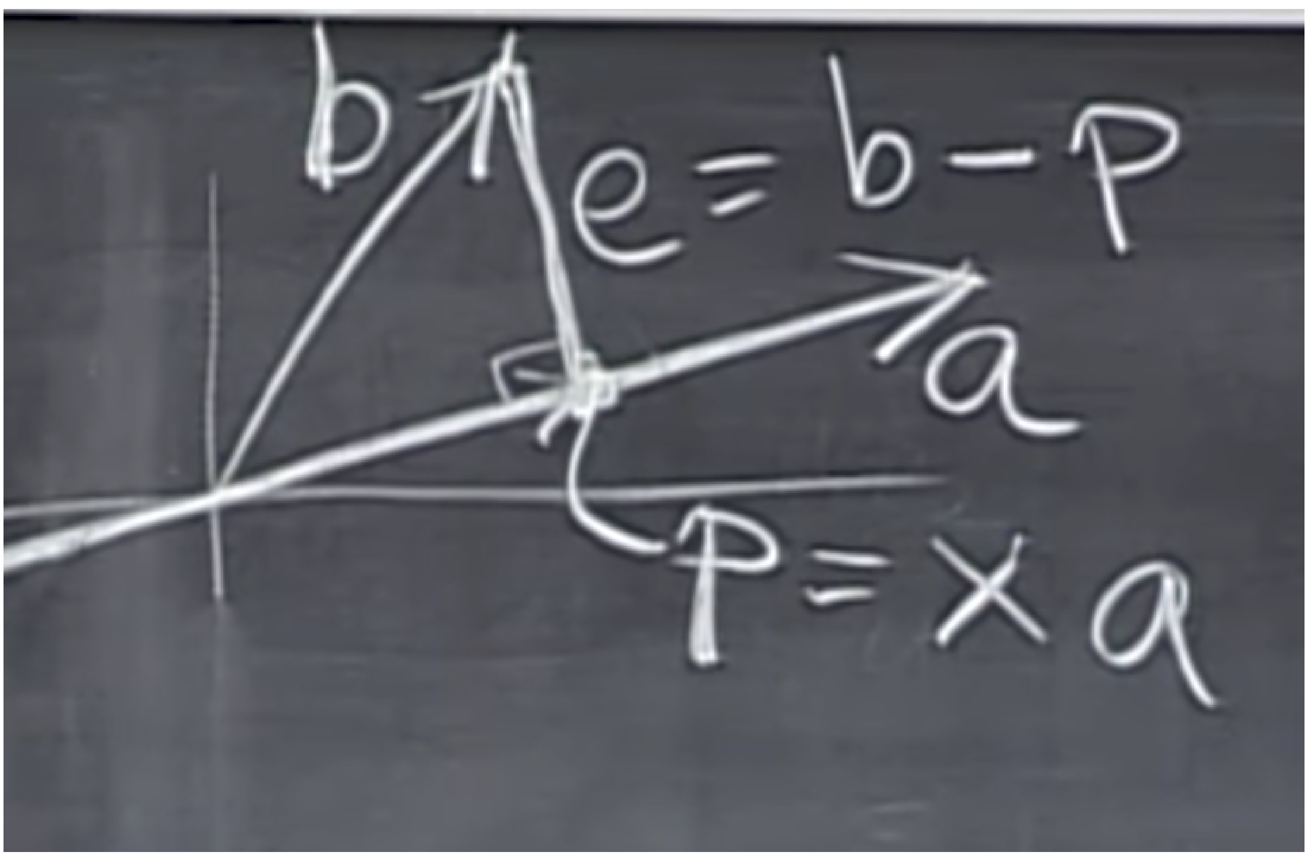
orthogonal, projections

- two vectors are orthogonal (perpendicular) if $X^TY = 0$, where X and Y are vectors. note that this is just the dot product.
- The square of the length of vector $\vec{x}$ is x^Tx (as we saw from herb gross' course)
- for two orthogonal vectors, $|x|^2 + |y|^2 = |x+y|^2$. Also $x^Ty = y^Tx$ for vectors. (see for yourself why)
- The $\vec{0}$ is orthogonal to every vector since $x^Ty = \vec{0}$ when one of them is $\vec{0}$.
- two subspaces being orthogonal means every vector in one subspace is orthogonal with any vector from the other subspace. Note that this does not mean that two vectors within one of the subspaces need be orthogonal.
- row space is orthogonal to the null space. (think about why this is true). hint: any row or any combination of rows multiplied by something in the null space will give $\vec{0}$. similarly, by definition, the column space is orthogonal to the left null space.
- also, not only are the row space and null space orthogonal, but also the null space contains every vector thats orthogonal to the row space.
- Consider that $\vec{e}$ is perpendicular to $\vec{a}$. using this fact, we can show that the projection $p = \frac{aa^T}{a^Ta}b$. Note that $\frac{aa^T}{a^Ta}$ is a matrix which we call the projection matrix since it projects a vector $\vec{b}$ onto $\vec{a}$



- also, note that $p^n = p$ → this is because once the projection is done, if we project again we get no change. You can also verify this algebraically. Also, $p^T = p$. this means the projection matrix is symmetric.
- Note that aa^T is of rank 1 since all the columns are just multiples of a.

- now there are cases where you have $Ax = b$ but there is no solution. That is lets say we have a large number of noisy observations and we are trying to find the relationship equation. We would have something like $Ax = b$, but we wouldn't have any solution (b is not in the column space of A). Here there are many many equations m and a few columns n so that $m >> n$.
- we will project b into the column space of A, so that we can find a $\hat{x}$ that solves $Ax = p$. ($Ax = p$ does have a solution because p is in $C(A)$.
- side note: notice that $A^T A$ is always a square matrix and is always symmetric.
- Now the rank of $A^T A$ is the same as the rank of A. This means that if the columns of A are independent, then $A^T A$ is always invertible (because the dimensions of $A^T A$ is cols(A) x cols(A). proof hint: think about the different ways of looking at matrix multiplication and see which of them is useful to prove that $A^T A$ has the same rank as A.
- alternative proof that if A has independent columns, $A^T A$ is invertible: this is the same as proving that $A^T Ax = 0$ has the only solution $x = \vec{0}$. Now left multiply by x^T to get $x^T A^T Ax = 0$ → $(Ax)^T Ax = 0$ → $||Ax|| = 0$ → $Ax = 0$ → $x = \vec{0}$ (since A has independent columns and isn't $\vec{0}$. hence proved.
- Now instead of solving $Ax = b$ (which is impossible to solve if b is not in the column space of A), we will solve $A^T Ax = A^T b$. The solution x we get does not exactly solve $Ax = b$ but we show why its the closest we can get. here, we say that A^T projects b onto the column space of A.
- why is the solution x obtained above the best solution? it turns out if we want to minimize $||A\hat{x} - b||^2$, we can do so by taking partial derivatives and setting it to 0 and it turns out that is equivalent to solving $A^T Ax = A^T b$! so the reason this is the best solution is in the sense that it minimizes the least square error. we can start with $||A\hat{x} - b||^2$ and take partials and set it to 0 and solve for that x and we will end up exactly with $A^T Ax = A^T b$. the proof is long, it is in cs 229 ML class notes, but is just a bunch of manipulation, after you've accepted that linear least squares is convex.
- Orthonormal vectors are a set of vectors such that they are all of magnitude 1 and are all perpendicular to each other. Notice that if the columns of a matrix are orthonormal (we call it Q), then $Q^T Q = I$.
- an orthonormal matrix that is square is called the orthogonal matrix. for an orthogonal matrix $Q^T = Q^{-1}$.
- If a set of vectors a mutually orthogonal, then they are independent. proof: assume they weren't linearly independent. then we can express $\vec{c_k}$ as linear combination of other $\vec{c}$'s. Let that linear combination be $a_i\vec{c_i} + ..a_j\vec{c_j}$. Now, consider $(a_i\vec{c_i} + ..a_j\vec{c_j}) \cdot \vec{c_k} = \vec{0}$ (because all the vectors are perpendicular to $\vec{c_k}$), but this is a contradiction since $\vec{c_k} \neq \vec{0}$. hence proved.
- Gram Schmidt is a procedure to find an orthonormal basis from an independent set of vectors. Given a, b, c (independent vectors), we first remove components of $a = A$ from b (call this B), and then components of A, B from c. Call this C. Now normalize each of them. the result is an orthonormal set. We can show easily that each step gives us a vector thats perpendicular to the previously obtained ones.