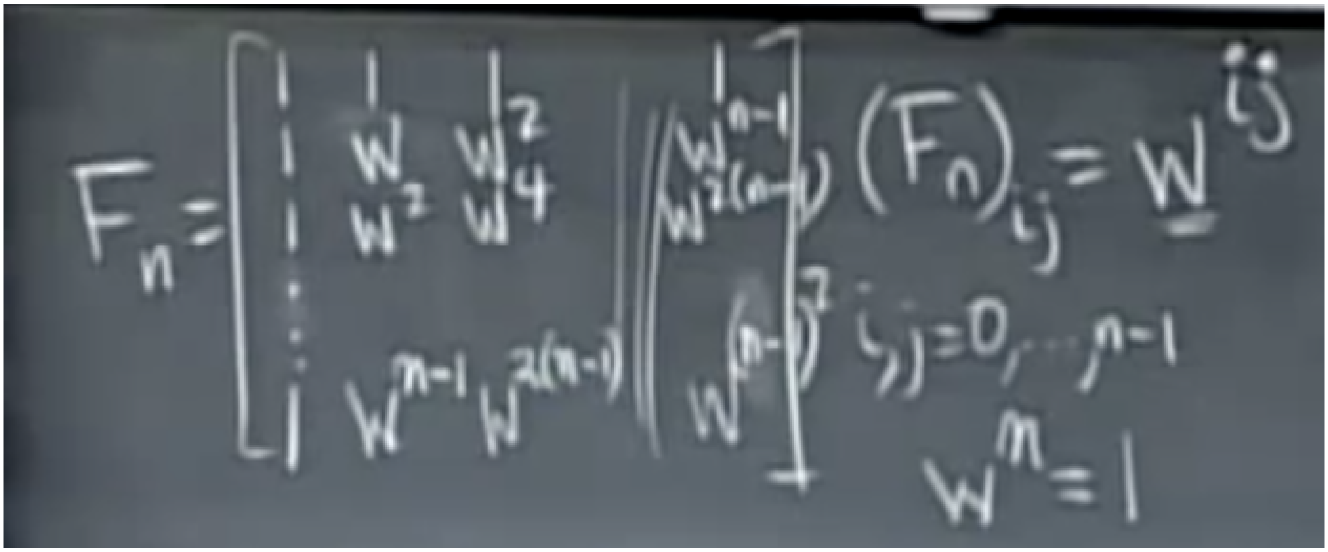
positive definiteness, singular value decomposition

- if a matrix is symmetric, then its eigenvalues are real, eigenvectors are independent and we can write $A = S\Lambda S^{-1}$ as $A = Q\Lambda Q^T$.
  proof:
  $Ax = \lambda x$ → $\overline{x}^T Ax = \lambda \overline{x}^T x$
  $Ax = \lambda x$ → $A\overline{x} = \overline{\lambda}\overline{x}$. (taking conjugates on both sides but A stays as it is since it is real (we are dealing only with real matrices here)) → $\overline{x}^T A = \overline{\lambda}\overline{x}^T$ (transposing) →$\overline{x}^T Ax = \overline{\lambda}\overline{x}^T x$.
  Which means $\lambda \overline{x}^T x = \overline{\lambda}\overline{x}^T x$. Since $\overline{x}^T x$ is not 0, this means $\lambda = \overline{\lambda}$ → λ is real. hence proved.
- So again, a symmetric matrix has all eigenvalues real. Now we can show that the eigenvectors are independent by a method called Schur factorization (see page 335 here, or see an alternate proof). Because the eigenvectors are independent, we can write $A = S\Lambda S^{-1}$ as $A = Q\Lambda Q^T$ where Q is an orthonormal matrix (and hence $Q^T = Q^{-1}$.
- As you read the proof, note that the Schur factorization is for any square matrix but we get an upper triangular T. Because we know A is symmetric, we conclude that T must be diagonal and hence it must be the eigenvector matrix.
- So in summary, a symmetric matrix has real eigenvalues and perpendicular eigenvectors. Diagonalization becomes $A = Q\Lambda Q^T$ with an orthogonal matrix Q. All symmetric matrices are diagonalizable, even with *repeated* eigenvalues.
- We can also write every symmetric matrix as a combination of perpendicular projection matrices: $A = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T .. \lambda_n q_n q_n^T$. This is called the "spectral theorem".
- for symmetric matrices the number of positive eigenvalues = number of positive pivots (and same for number of negatives of each) proof on page 334 here.
- A very practical application of the above is to find if all λ's are > 0 for the symmetric matrix. in general computing λ's for large matrices takes time so we can just reduce the matrix and compute pivots and see if they are all positive.
- positive definite matrices are matrices that are symmetric and have all eigenvalues positive (and hence, all the pivots are also positive).
- this also implies that all n upper left sub-determinants (1x1 matrix, 2x2 matrix.. n x n matrix) are positive. to see why this is true, think about the fact that the pivots don't change when you take a matrix k x k (upper left, k < n).
- to find the magnitude of the a complex vector, note that simply doing $z^T z$ does not give a magnitude number. We need to do $\overline{z}^T z$ → also called $z^H z$, the hermitian.
- Similarly, the inner product of two complex vectors is $y^H x$ (note this could be a complex number)
- similarly, the complex version of symmetry is if $\overline{A}^T = A$ (A is called the hermitian matrices)
- Following the earlier proof we gave, we can show similarly that hermitian matrices have real eigenvalues.
- The orthogonal equivalent of complex matrices is $Q^H = \overline{Q}^T$. We will have $Q^H Q = I$ instead of $Q^T Q = I$.
- the fourier matrix F_n is defined as follows:



- Here w is the n^{th} root of i, where n is the size of the matrix (not 0-indexed). $a_{ij} = w^{i+j}$ (don't confuse complex number i which is the base with the $i + j$ which is the sum of the row and column) → the row and column number here are *0-indexed,* unlike usual.

normalized fourier matrix 4 x 4

- One can show that the inner product $y^H x$ of any column with any other column in the above is 0 → this matrix is orthonormal! so $F_n^H F_n = I$. → the inverse of F_n is the hermitian.
- The above is used in the fast fourier transform which is a way to multiply (some?) matrices in $nlog(n)$ time (much faster than before). "The FFT has revolutionized signal processing. Whole industries are speeded up by this one idea."
- Strang mentions quickly how this matrix is broken down to achieve this but doesn't go too much into the mechanics or proofs 😞 But this clearly looks revolutionary, assuming it achieves what it says it does (of course I trust it does, but he doesn't prove it)
- If A is positive definite, the eigenvalues of A^{-1} are simply the reciprocals of the eigenvalues of A. proof: $Ax = \lambda x$ → $A^{-1}Ax = \lambda A^{-1}x$ → $A^{-1}x = \frac{1}{\lambda}x$. QED (is an initialism of the Latin phrase quod erat demonstrandum, meaning "what was to be demonstrated)
- the above proves that the inverse of a positive definite matrix is also positive definite.
- for positive definite A, we can show that $x^T Ax > 0$ for any non-zero x. proof: express given x as a combination of orthonormal eigenvectors → $x = c_1x_1 + ..c_nx_n$. then $x^T Ax = (c_1x_1 + ..c_nx_n)^T(c_1\lambda_1x_1 + .. + c_n\lambda_nx_n)$. Note that all the x_ix_j terms where $i \neq j$ are 0

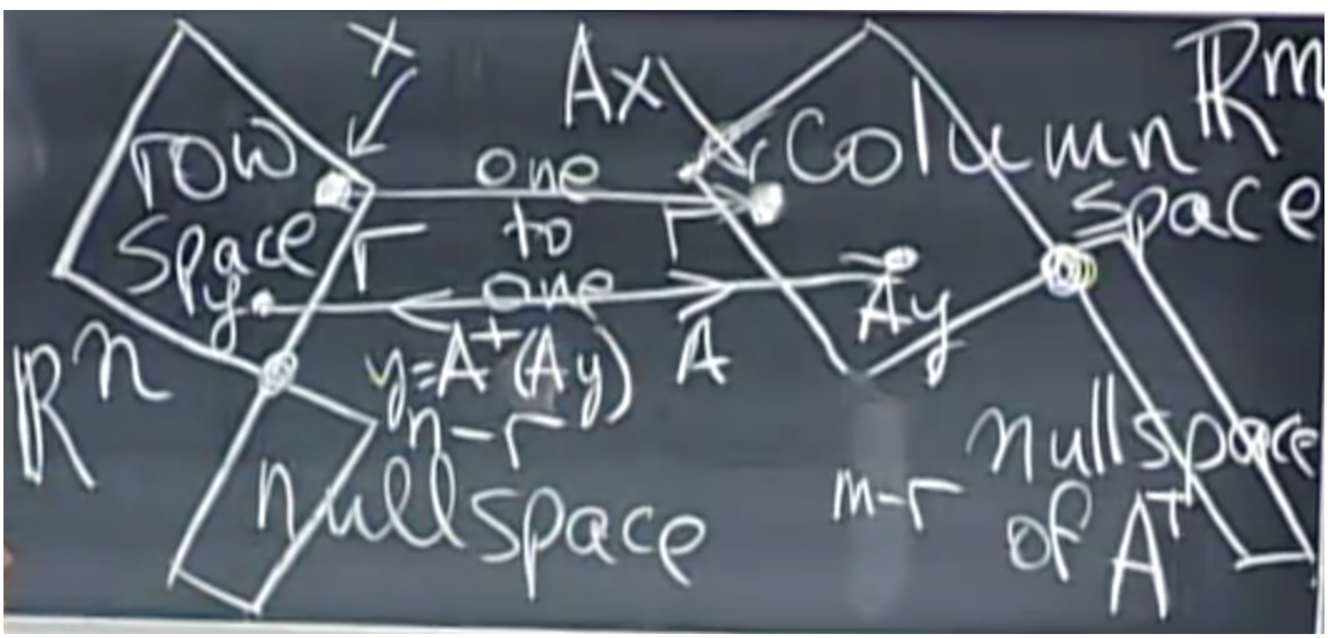
since it is a orthogonal eigenvector basis and we are left with sum of squares which is definitely positive (positive definite 😉)

- so for a symmetric matrix any of these 4 mean the other 3:
  - $x^TAx > 0$ for all non-zero x
  - all λ's > 0
  - all pivots > 0
  - all sub determinants > 0
- for positive semidefinite, just change the above to all ≥ 0 instead of > 0. Note these are allowed to be singular, but positive definite implies invertible.
- if A, B are both positive definite, so is $A + B$. proof: $x^TAx + x^TBx > 0$ → $x^T(A + B)x > 0$
- if a rectangular A (m x n) with no. cols > no. rows with full column rank → A^TA is symmetric and invertible (see why) → is positive definite. proof: $x^T(A^TA)x = (Ax)^TAx = ||Ax||^2$ which is always positive if A is invertible.
- A and B are similar when there exists invertible M such that $B = M^{-1}AM$. Note that this means they have the same eigenvalue. proof: $Bv = \lambda v$ → $M^{-1}AMv = \lambda v$ → $A(Mv) = \lambda(Mv)$ → λ's are shared by A and B.
- the good, easy case is when the λ's are not repeated → we can find independent eigenvectors and we can diagonalize any family in that matrix (any matrix with those λ's) as $A = S\Lambda S^{-1}$. This is the best connection $S = M$ to choose because it connects every matrix in the family to a common matrix Λ and everything is nice and easy.
- note that all matrices in this family have the same determinant (same eigenvalues and det = product of eigenvalues)
- Now if we have repeated eigenvalues, all eigenvectors need not be independent. So we have more than one family → one for which the eigenvectors are independent (think about the matrix with eigenvalues on the diagonals and 0's everywhere - this is already diagonalized) and another for which the eigenvectors aren't (for these because the eigenvectors aren't independent, we can't diagonalize them as $A = S\Lambda S^{-1}$).
- The closest we can get in the second family to a diagonal form is called the Jordan form. This is the best we can do.
- The Jordan form for the distinct eigenvalue case is Λ.
- Some prereq things to realize for the singular value decomposition:
  - Let A be an n x m matrix. Let A^T be the transposed matrix of A. Then AA^T is an n × n matrix and A^TA is an m×m matrix. AA^T and A^TA have the same non-zero eigenvalues, and if one has more eigenvalues than the other, then these are all equal to 0.
    proof:
      Let λ be an eigenvalue of A^TA, i.e.
      $A^TAx = \lambda x$ for some $x \neq 0$. We can multiply A from the left and get
      $AA^T(Ax) = \lambda(Ax)$.
  - Since A^TA and AA^T are symmetric, they have real eigenvalues and full set of orthogonal eigenvectors.
- the singular value decomposition:
  We want to factorize $A_{m\times n} = U_{m\times m}\Sigma_{m\times n}V^T_{n\times n}$ where U, V are orthogonal.
  theorem:
  Find the orthonormal eigenvector set of AA^T and A^TA. This is U and V. Σ is the matrix with the root of the eigenvalues of A^TA (or AA^T) as the first r elements on the leading diagonal. Every other element is 0. Σ is rectangular $m \times n$. this is possible for every A

(rough) **proof:**
start with A → consider A^TA → its eigenvalues can be got by $A^TAv_i = \lambda v_i$ (note that we choose v_i's to be orthonormal and this is possible because A^TA is symmetric → $AA^T(Av_i) = \lambda(Av_i)$ (which means Av_i is an eigenvector of AA^T.
Now multiplying both sides of $A^TAv_i = \lambda v_i$ by v_i^T, we see that $v_i^TA^TAv_i = \lambda v_i^Tv_i$. Because v_i is a unit vector, this means $||Av_i||^2 = \lambda$ → the length of Av_i is $\sqrt{\lambda}$.
The v_i's go in V.
Now because we put the orthonormal eigenvectors of AA^T in U and because $AA^T(Av_i) = \lambda(Av_i)$, Av_i's generate the vectors in U. But because $||Av_i||$ is $\sqrt{\lambda}$ and $||u_i|| = 1$, we have to multiply each u_i by $\sqrt{\lambda}$ to make it equal Av_i. Lets call $\sigma = \sqrt{\lambda}$. So now we have seen that $Av_i = \sigma u_i$. This is the column by column story. Putting together all columns, $AV = U\Sigma$ → $A = U\Sigma V^T$.

- Now notice why between U and V we have a basis for the row space, column space, nullspace and left null space of A. proof idea: Consider that $\frac{Av_i}{\sigma_i} = u_i$ and $A^TAv_i = \lambda v_i$. consider when each of these is 0 and not 0. You will see why U and V cover the basis for the 4 fundamental subspaces of A.
- Notice that the fact that $||Av_i||^2 = \lambda$ means that λ will always be positive for these matrices (for AA^T and A^TA).
- Notice the dimensions work out and that because the last few columns of Σ can be $0's$, these correspond to the last few v_i's in V that are in the nullspace of A.
- The σ's are called the singular values. Notice that the SVD gives us:
  - It gives us a basis for the 4 fundamental subspaces of A
  - A in terms of the product of orthogonal matrices and a diagonal.
  - It gives us the eigenvalues of A^TA (And AA^T)
- A linear transformation T is something that follows these rules:
  - $T(v + w) = T(v) + T(w)$
  - $T(cv) = cT(v)$
- Notice that a matrix multiplication is a linear transformation by the above definition.
- It is easy to see that a projection (seen in either in the conceptual angle sense or the matrix multiplication by a projection matrix sense) is also a linear transformation. Adding a constant vector to an input is not a linear transformation. $T(v) = ||v||$ is also not a linear transformation. Rotation by a fixed angle is a linear transformation.
- Since a matrix multiplication is a linear transformation, any operation that can be expressed as a matrix multiplication, $T(v) = Av$, (projection, rotation etc) is a linear transformation. Matrix multiplication is an important linear transformation.
- How much information do we need to know T? (Knowing T means knowing $T(v)$ for all v). Well, if I gave you what T does to $v_1, v_2..v_n$, then, knowing that T is a linear transformation, you know what T does to any vector in the space spanned by $v_1, ..v_n$. Given a random $\vec{v}$ in this space, we can compute $T(\vec{v})$ by computing the coefficients in $\vec{v} = c_1v_1 + ..c_nv_n$. $c_1, ...c_n$ are called the coordinates of $\vec{v}$ in the basis $v_1, ...v_n$.
- Knowing the above, if we specify T for a orthonormal basis, then we can compute T for any vector in that basis' space. How to compute T for any vector in that space? Put the bases in the rows of a matrix and multiply this matrix by the input vector to get the coefficient vector. Now lay out $T(\vec{b}_1), T(\vec{b}_2)..T(\vec{b}_n)$ in the columns of a matrix and mutiply this with a vector that has the coefficients $c_1..c_n$ as its elements. This gives $T(\vec{v})$ because the last matrix multiplication is the equivalent of taking a linear combination of $T(\vec{b}_1), T(\vec{b}_2)..T(\vec{b}_n)$ with coefficients $c_1..c_n$.

- This proves that any linear transformation can be expressed as a matrix multiplication and the way to construct the matrix is above!
- If I say a vector $[2\ 1\ 3\ 4]^T$ we assume the basis is the standard identity basis and these are the co-efficients. While that is true, we can use $[2\ 1\ 3\ 4]_B^T$ to mean it is the coordinates in basis B, where B is a matrix whose rows are the basis.
- Given a basis expressed as the basis vectors in the columns of a matrix B, to compute the coordinates of a vector in the basis B, we need $Bc = \vec{v}$ where $\vec{v}$ is the coordinates in the identity basis. Often we will have $\vec{v}$ and we want to express this in basis B to find $\vec{c}$. We can just do $B^{-1}\vec{v}$. Because this operation is done often, a good basis is one which is easily invertible.
- Once we have a transformation matrix A, if we can find a eigenvector basis $x_1..x_n$, we can write $A\vec{v} = c_1\lambda_1 x_1 + ... + c_n\lambda_n x_n$. If out of 100 λ's, only 7-8 are significant and the rest are super small, we can approximate that as the sum of just those significant terms. Notice how this speeds up the computation by ~10x. The fewer the significant λ's we keep the faster the computation of the transformation, but more tradeoff of accuracy.
- Note that if a matrix is full column rank ($m \times n$ with $m > n$), then A^TA is invertible. (proof: $x^TA^TAx = ||Ax||^2 > 0$ if $\vec{x} \neq \vec{0}$ because A has independent columns). Now note that $(A^TA)^{-1}A^TA = I$. So we call $(A^TA)^{-1}A^T$ the left inverse of A. Note that A does not have a 2-sided inverse.
- Now from the above example look at A^T. Lets call this our new A. This A has full row rank and the nullspace of A^T is only the $\vec{0}$ vector. $Ax = b$ has infinite solutions ($n - m$ free variables).
- Using the same multiplication as for the left inverse, (except remember we call $A\ A^T$ and we call $A^T\ A$, we can write $AA^T(AA^T)^{-1} = I$ which means $A^T(AA^T)^{-1}$ is the right inverse.
- For a matrix with a nullspace, it takes everything in its nullspace to $\vec{0}$ so there is no hope of an inverse. $A^{-1}\vec{0}$ is not unique.
- theorem: the mapping between row space and column space is 1:1.



proof:
note this is at least possible because both row space and column space have the same dimension (lets call this r for rank). Now lets say x and y are two vectors in the row space that are different. Now lets say $Ax = Ay$. This means $A(x - y) = 0$. That means $x - y$ is in the nullspace. But we know $x - y$ is

in the rowspace since x, y are both in the rowspace. Contradiction. QED. This shows that different x, y in the rowspace map to different vectors in the column space.
Now why does the rowspace cover the full column space? Lets a y in the column space. We know that there must be a solution to $Ax = y$ (since y is in the column space). Find one such solution. This is a hint. Strang does not prove this 😞

- Note that for diagonal matrices of the type $\Sigma_{m\times n}$ (note these need not be square) is $\Sigma_{n\times m}$ with $\frac{1}{\sigma}$ instead of σ's and it is denoted by Σ^+.
- Finding the pseudoinverse:
  - Do $A = U\Sigma V^T$. Write down Σ^+. Now U, V are invertible. Write down $V\Sigma^+U^T$. This is the pseudo inverse. To see why, multiply by $U\Sigma V^T$. You will get I.