Unit III : Lecture - 1

Special matrices have special eigenvalues & eigenvectors. Symmetric & positive definite matrices have extremely nice properties, and studying these matrices brings together everything we have learned about pivots, determinants & eigenvalues. In this session we also practice doing linear algebra with complex numbers and learn how the pivots give information about the eigenvalues of a symmetric matrix.

# Symmetric matrices and positive definiteness

Symmetric matrices are good – their eigenvalues are real and each has a complete set of orthonormal eigenvectors. Positive definite matrices are even better.

## Symmetric matrices

A *symmetric matrix* is one for which $A = A^T$. If a matrix has some special property (e.g. it's a Markov matrix), its eigenvalues and eigenvectors are likely to have special properties as well. For a symmetric matrix with real number entries, the eigenvalues are real numbers and it's possible to choose a complete set of eigenvectors that are perpendicular (or even orthonormal).

If $A$ has $n$ independent eigenvectors we can write $A = S\Lambda S^{-1}$. If $A$ is symmetric we can write $A = Q\Lambda Q^{-1} = Q\Lambda Q^T$, where $Q$ is an orthogonal matrix. Mathematicians call this the *spectral theorem* and think of the eigenvalues as the "spectrum" of the matrix. In mechanics it's called the *principal axis theorem.*

In addition, any matrix of the form $Q\Lambda Q^T$ will be symmetric.

### Real eigenvalues

Why are the eigenvalues of a symmetric matrix real? Suppose $A$ is symmetric and $A\mathbf{x} = \lambda\mathbf{x}$. Then we can conjugate to get $\overline{A}\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. If the entries of $A$ are real, this becomes $A\overline{\mathbf{x}} = \overline{\lambda}\overline{\mathbf{x}}$. (This proves that complex eigenvalues of real valued matrices come in conjugate pairs.)

Now transpose to get $\overline{\mathbf{x}}^T A^T = \overline{\mathbf{x}}^T\overline{\lambda}$. Because $A$ is symmetric we now have $\overline{\mathbf{x}}^T A = \overline{\mathbf{x}}^T\overline{\lambda}$. Multiplying both sides of this equation on the right by $\mathbf{x}$ gives:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x}.$$

On the other hand, we can multiply $A\mathbf{x} = \lambda\mathbf{x}$ on the left by $\overline{\mathbf{x}}^T$ to get:

$$\overline{\mathbf{x}}^T A\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}.$$

Comparing the two equations we see that $\overline{\mathbf{x}}^T\overline{\lambda}\mathbf{x} = \overline{\mathbf{x}}^T\lambda\mathbf{x}$ and, unless $\overline{\mathbf{x}}^T\mathbf{x}$ is zero, we can conclude $\lambda = \overline{\lambda}$ is real.

How do we know $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$?

$$\overline{\mathbf{x}}^T\mathbf{x} = \begin{bmatrix} \overline{x}_1 & \overline{x}_2 & \cdots & \overline{x}_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2.$$

If $\mathbf{x} \neq \mathbf{0}$ then $\overline{\mathbf{x}}^T\mathbf{x} \neq 0$.



t.

With complex vectors, as with complex numbers, multiplying by the conjugate is often helpful.

Symmetric matrices with real entries have $A = A^T$, real eigenvalues, and perpendicular eigenvectors. If $A$ has complex entries, then it will have real eigenvalues and perpendicular eigenvectors if and only if $A = \overline{A}^T$. (The proof of this follows the same pattern.)

**Projection onto eigenvectors**

If $A = A^T$, we can write:

$$\begin{aligned} A &= Q\Lambda Q^T \\ &= \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{q}_1^T \\ \mathbf{q}_2^T \\ \vdots \\ \mathbf{q}_n^T \end{bmatrix} \\ &= \lambda_1 \mathbf{q}_1 \mathbf{q}_1^T + \lambda_2 \mathbf{q}_2 \mathbf{q}_2^T + \cdots + \lambda_n \mathbf{q}_n \mathbf{q}_n^T \end{aligned}$$

The matrix $\mathbf{q}_k \mathbf{q}_k^T$ is the projection matrix onto $\mathbf{q}_k$, so every symmetric matrix is a combination of perpendicular projection matrices.

**Information about eigenvalues**

If we know that eigenvalues are real, we can ask whether they are positive or negative. (Remember that the signs of the eigenvalues are important in solving systems of differential equations.)

For very large matrices $A$, it's impractical to compute eigenvalues by solving $|A - \lambda I| = 0$. However, it's not hard to compute the pivots, and the signs of the pivots of a symmetric matrix are the same as the signs of the eigenvalues:

$$\text{number of positive pivots} = \text{number of positive eigenvalues.}$$

Because the eigenvalues of $A + bI$ are just $b$ more than the eigenvalues of $A$, we can use this fact to find which eigenvalues of a symmetric matrix are greater or less than any real number $b$. This tells us a lot about the eigenvalues of $A$ even if we can't compute them directly.

## Positive definite matrices

A *positive definite matrix* is a symmetric matrix $A$ for which all eigenvalues are positive. A good way to tell if a matrix is positive definite is to check that all its pivots are positive.

Let $A = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$. The pivots of this matrix are 5 and $(\det A)/5 = 11/5$. The matrix is symmetric and its pivots (and therefore eigenvalues) are positive, so $A$ is a positive definite matrix. Its eigenvalues are the solutions to:

$$|A - \lambda I| = \lambda^2 - 8\lambda + 11 = 0,$$

i.e. $4 \pm \sqrt{5}$.

The determinant of a positive definite matrix is always positive but the determinant of $\begin{bmatrix} -1 & 0 \\ 0 & -3 \end{bmatrix}$ is also positive, and that matrix isn't positive definite. If all of the subdeterminants of $A$ are positive (determinants of the $k$ by $k$ matrices in the upper left corner of $A$, where $1 \leq k \leq n$), then $A$ is positive definite.

The subject of positive definite matrices brings together what we've learned about pivots, determinants and eigenvalues of square matrices. Soon we'll have a chance to bring together what we've learned in this course and apply it to non-square matrices.

MIT OpenCourseWare
http://ocw.mit.edu

18.06SC Linear Algebra
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

Linear Algebra Lecture 25

Symmetric Matrices

Eigenvalues/EigenVectors

Start: Positive Definite Matrices

$A = A^T$ (For Real Symmetric Matrices)

① The eigenvalues are REAL

② The eigenvectors are PERPENDICULAR or can be chosen perpendicular (case of repeated eigenvalues)

we can have orthonormal unit vectors as well (columns of $Q$)

In usual case: $A = S\Lambda S^{-1}$

In Symmetric case: $A = S\Lambda S^{-1}$

Now since we have symmetric "A", we know that $\Lambda$ contain real eigen values & $S$ = matrix of eigenvector of A, contains orthonormal eigen vector. Hence "S" can be written as Q.

$\Rightarrow A = Q\Lambda Q^{-1}$ $\quad Q$ = square matrix, $Q^{-1} = Q^T$ } Reality

$$\boxed{A = Q\Lambda Q^T}$$

famous theorem of Linear Algebra:

"Any real symmetric matrix can be factorised into an orthogonal matrix times diagonal matrix time transpose of orthogonal matrix"

Its spectral theorem (In Mathematics). Spectrum is the set of eigen values of a matrix. It comes from the idea of spectrum of light which describe light as a combination of pure things where our matrix is broken down into pure eigenvalues & eigenvectors. In mechanics it is also called <u>principal axis theorem</u>.

Q: Why are the eigenvalues of A ~~are~~ always real?

$$(i) — \boxed{Ax = \lambda x} \Longrightarrow \bar{A}\bar{x} = \bar{\lambda}\bar{x} \quad (\text{conjugate of } Ax = \lambda x).$$

since we have real A $\qquad \overline{a+ib} = a - ib$

$$Ax = \lambda x \Rightarrow A\bar{x} = \bar{\lambda}\bar{x}$$

$x, \lambda$ could be complex & thus $\bar{\lambda}$ might be complex as well.

Now we will use symmetry.

$$A\bar{x} = \bar{\lambda}\bar{x} \Longrightarrow (A\bar{x})^T = (\bar{\lambda}\bar{x})^T$$

$$\bar{x}^T A^T = \bar{x}^T \bar{\lambda}$$

If A is symmetric

then $\boxed{\bar{x}^T A = \bar{x}^T \bar{\lambda}}$ — (ii)

Multiply eq(i) by $\bar{x}^T$

$$\boxed{\bar{x}^T A x = \lambda \bar{x}^T x} \text{ — (iii)}$$

Multiply eq(ii) by x

$$\boxed{\bar{x}^T A x = \bar{x}^T \bar{\lambda} x} \text{ — (iv)}$$

L.H.S of (iii) & (iv) are same

$$\Rightarrow \lambda \bar{x}^T x = \bar{\lambda} \bar{x}^T x$$

$$\Rightarrow \boxed{\lambda = \bar{\lambda}}$$

$\lambda$ is real as $\lambda$ is equal to its conjugate. i.e it doesn't have complex part.

In $\lambda \bar{x}^T x = \bar{\lambda} \bar{x}^T x$, we know that $\bar{x}^T x \neq 0$. How to prove it? $\bar{x}^T x$, Here x is complex.

Now when $X$ is complex then:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} \qquad \bar{X} = \text{conjugate of } X = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \\ \vdots \\ \bar{x}_n \end{bmatrix}$$

Then $\bar{X}^T X = \begin{bmatrix} \bar{x}_1 & \bar{x}_2 & \dots & \bar{x}_N \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}$

$$\cancel{\begin{bmatrix} \bar{x}_1 x_1 \\ \bar{x}_2 x_2 \\ \vdots \\ \bar{x}_N x_N \end{bmatrix}} = \bar{x}_1 x_1 + \bar{x}_2 x_2 + \dots$$

$(a-ib)(a+ib) = a^2+b^2$

As
$(a-ib)(a+ib)$
$= a^2 - (ib)^2$
$= a^2 + b^2$

Thus $\underbrace{\bar{X}^T X}_{\substack{(\text{length})^2 \\ \text{of vector } \bar{X}}} > 0$ where $x$ is vector

Now we can define our good matrices, By good means our matrix $A$ is real & symmetric i.e $A = A^T$ then:

i) They have real $\lambda$'s (eigenvalues)

ii) Perpendicular $x$'s (eigenvectors)

And we just proved (i) with $A$ being symmetric & real.

Now if $A$ had been complex then our above proof will also work if $\bar{A}^T = A$. That is the good matrices if complex are those whose conjugate transpose ($\bar{A}^T$) is equal to the original complex matrix $A$.

If $\underline{\underline{A = A^T}}$ $\quad A = Q \Lambda Q^T$

$Q$ — orthonormal eigenvector; $\Lambda$ — real eigen values

$$= \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} q_1^T \\ q_2^T \\ \vdots \end{bmatrix}$$

$$= \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \lambda_3 q_3 q_3^T + \cdots$$

Every symmetric matrix break into ~~these~~ above parts.

Here $q_1 q_1^T$, $q_2 q_2^T \ldots$ are matrices b/c $q_1$ = column vector & $q_2^T$ is row vector. where $q_1, q_2$, are orthonormal unit vectors ~~thus~~ $q_1 q_1^T$, $q_2 q_2^T$ ~~are~~ projection matrices. Thus

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 3 \end{bmatrix}$$

$(3 \times 1)$ $(1 \times 3)$

every symmetric matrix is a combination of perpendicular projection matrices

Another way of spectral theorem

**Another great fact of symmetric Matrices:**

When we have symmetric matrices we know that their eigen values are real. Then we can ask whether these eigen values are positive or negative And it is important b/c we have seen that sign of eigenvalues decides stability & instability in differential equation

Thus in $A = A^T$

Signs of pivots are same as signs of $\lambda$'s.

i.e # of positive pivots = # of positive $\lambda$'s

This tells us:

Product of pivots = product of $\lambda$'s

b/c determinant = (product of pivots) = product of $\lambda$'s.

# Positive Definite (symmetric) Matrix :

Three facts about positive definite matrices

1. They are symmetric
2. Their eigenvalues are ~~real~~ positive
3. They are subclass of symmetric matrices
4. All pivots are positive (because of rule on last page)
5. All subdeterminants are positive

For Instance

$$A = \begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$$

~~Pivots → 5 2~~

Pivots

$5, \frac{11}{5}$ (~~Product of pivots =~~ Product of pivots = $|A|$)

we got $\frac{11}{5}$ by the following fact

$|A| = \text{pivot \#1} \cdot \text{pivot \#2}$

$11 = 5 \times ?$

$\Rightarrow \boxed{\text{pivot \#2} = 11/5}$

Now For eigenvalues

$\lambda_s = ??$

$\lambda^2 - 8\lambda + 11 = 0$

$\lambda = \frac{4 \pm \sqrt{16-11}}{2}$

$\boxed{\lambda = 2 \pm \frac{\sqrt{5}}{2}}$ → $\lambda$ are $> 0$

## Elaborating point # 5

Meaning of subdeterminants to be positive.

Ex#1

$$\underbrace{\begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}}_{\det = 11 > 0} \rightarrow \underbrace{\begin{vmatrix} 5 \end{vmatrix}}_{\substack{1\times1 \\ \det = 5 > 0}} \rightarrow \underbrace{\begin{vmatrix} 5 & 2 \\ 2 & 3 \end{vmatrix}}_{\substack{2\times2 \\ \det = 11 > 0}}$$

All determinants $> 0$ hence $\begin{bmatrix} 5 & 2 \\ 2 & 3 \end{bmatrix}$ is P.D

Ex#2

$$\underbrace{\begin{vmatrix} -1 & 0 \\ 0 & -3 \end{vmatrix}}_{\det = 3 > 0} \rightarrow \underbrace{\begin{vmatrix} -1 \end{vmatrix}}_{\det = -1 < 0} \rightarrow \underbrace{\begin{vmatrix} -1 & 0 \\ 0 & -3 \end{vmatrix}}_{\det = 3 > 0}$$

Although $\det \begin{vmatrix} -1 & 0 \\ 0 & -3 \end{vmatrix} > 0$, we can't say it's PD unless all its subdeterminants are +ve.

Recitation?

Unit III: Lecture 2

The Fourier matrices have complex valued entries and many nice properties. This session covers the basic of working with complex matrices & vectors and concludes with a description of FFT.

## Complex matrices; fast Fourier transform

Matrices with all real entries can have complex eigenvalues! So we can't avoid working with complex numbers. In this lecture we learn to work with complex vectors and matrices.

The most important complex matrix is the Fourier matrix $F_n$, which is used for Fourier transforms. Normally, multiplication by $F_n$ would require $n^2$ multiplications. The fast Fourier transform (FFT) reduces this to roughly $n \log_2 n$ multiplications, a revolutionary improvement.

### Complex vectors

#### Length

Given a vector $\mathbf{z} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} \in \mathbb{C}^n$ with complex entries, how do we find its length? Our old definition:

$$\mathbf{z}^T\mathbf{z} = \begin{bmatrix} z_1 & z_2 & \cdots & z_n \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

is no good; this quantity isn't always positive. For example:

$$\begin{bmatrix} 1 & i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 0.$$

We don't want to define the length of $\begin{bmatrix} 1 \\ i \end{bmatrix}$ to be 0. The correct definition is: $|\mathbf{z}|^2 = \overline{\mathbf{z}}^T\mathbf{z} = |z_1|^2 + |z_2|^2 + \cdots + |z_n|^2$. Then we have:

$$\left(\text{length} \begin{bmatrix} 1 \\ i \end{bmatrix}\right)^2 = \begin{bmatrix} 1 & -i \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = 2.$$

To simplify our notation we write $|\mathbf{z}|^2 = \mathbf{z}^H\mathbf{z}$, where $\mathbf{z}^H = \overline{\mathbf{z}}^T$. The $H$ comes from the name Hermite, and $\mathbf{z}^H\mathbf{z}$ is read "$\mathbf{z}$ Hermitian $\mathbf{z}$".

#### Inner product

Similarly, the inner or dot product of two complex vectors is not just $\mathbf{y}^T\mathbf{x}$. We must also take the complex conjugate of $\mathbf{y}$:

$$\mathbf{y}^H\mathbf{x} = \overline{\mathbf{y}}^T\mathbf{x} = \overline{y}_1x_1 + \overline{y}_2x_2 + \cdots + \overline{y}_nx_n.$$

## Complex matrices

### Hermitian matrices

Symmetric matrices are real valued matrices for which $A^T = A$. If $A$ is complex, a nicer property is $\overline{A}^T = A$; such a matrix is called *Hermitian* and we abbreviate $\overline{A}^T$ as $A^H$. Note that the diagonal entries of a Hermitian matrix must be real. For example,

$$\overline{A}^T = A = \begin{bmatrix} 2 & 3+i \\ 3-i & 5 \end{bmatrix}.$$

Similar to symmetric matrices, Hermitian matrices have real eigenvalues and perpendicular eigenvectors.

### Unitary matrices

What does it mean for complex vectors $\mathbf{q}_1, \mathbf{q}_2, ..., \mathbf{q}_n$ to be perpendicular (or orthonormal)? We must use our new definition of the inner product. For a collection of $\mathbf{q}_j$ in complex space to be orthonormal, we require:

$$\overline{\mathbf{q}}_j \mathbf{q}_k = \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

We can again define $Q = \begin{bmatrix} \mathbf{q}_1 & \mathbf{q}_2 & \cdots & \mathbf{q}_n \end{bmatrix}$, and then $Q^H Q = I$. Just as "Hermitian" is the complex equivalent of "symmetric", the term "*unitary*" is analogous to "orthogonal". A *unitary matrix* is a square matrix with perpendicular columns of unit length.

## Discrete Fourier transform

A *Fourier series* is a way of writing a periodic function or *signal* as a sum of functions of different frequencies:

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots.$$

When working with finite data sets, the *discrete Fourier transform* is the key to this decomposition.

In electrical engineering and computer science, the rows and columns of a matrix are numbered starting with 0, not 1 (and ending with $n-1$, not $n$). We'll follow this convention when discussing the Fourier matrix:

$$F_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & w & w^2 & & w^{n-1} \\ 1 & w^2 & w^4 & & w^{2(n-1)} \\ \vdots & & & \ddots & \vdots \\ 1 & w^{n-1} & w^{2(n-1)} & \cdots & w^{(n-1)^2} \end{bmatrix}.$$

Notice that $F_n = F_n^T$ and $(F_n)_{jk} = w^{jk}$, where $j,k = 0,1,...,n-1$ and the complex number $w$ is $w = e^{i\cdot 2\pi/n}$ (so $w^n = 1$). The columns of this matrix are orthogonal.

All the entries of $F_n$ are on the unit circle in the complex plane, and raising each one to the $n$th power gives 1. We could write $w = \cos(2\pi/n) + i\sin(2\pi/n)$, but that would just make it harder to compute $w^{jk}$.

Because $w^4 = 1$ and $w = e^{2\pi i/4} = i$, our best example of a Fourier matrix is:

$$F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}.$$

To find the Fourier transform of a vector with four components (four data points) we multiply by $F_4$.

It's easy to check that the columns of $F_4$ are orthogonal, as long as we remember to conjugate when computing the inner product. However, $F_4$ is not quite unitary because each column has length 2. We could divide each entry by 2 to get a matrix whose columns are orthonormal:

$$\frac{1}{4}F_4^H F_4 = I.$$

**An example**

The signal corresponding to a single impulse at time zero is (roughly) described by $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$. To find the Fourier transform of this signal we compute:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

A single impulse has all frequencies in equal amounts.

If we multiply by $F_4$ again we almost get back to $(1,0,0,0)$:

$$\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 0 \\ 0 \\ 0 \end{bmatrix} = 4 \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Because $\frac{1}{\sqrt{n}}F_n$ is unitary, multiplying by $F_n$ and dividing by the scalar $n$ inverts the transform.

## Fast Fourier transform

Fourier matrices can be broken down into chunks with lots of zero entries; Fourier probably didn't notice this. Gauss did, but didn't realize how significant a discovery this was.

There's a nice relationship between $F_n$ and $F_{2n}$ related to the fact that $w_{2n}^2 = w_n$:

$$F_{2n} = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_n & 0 \\ 0 & F_n \end{bmatrix} P,$$

where $D$ is a diagonal matrix and $P$ is a $2n$ by $2n$ permutation matrix:

$$P = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & 1 & \cdots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \cdots & 0 & 1 \end{bmatrix}.$$

So, a $2n$ sized Fourier transform $F$ times $\mathbf{x}$ which we might think would require $(2n)^2 = 4n^2$ operations can instead be performed using two size $n$ Fourier transforms ($2n^2$ operations) plus two very simple matrix multiplications which require on the order of $n$ multiplications. The matrix $P$ picks out the even components $x_0, x_2, x_4, \ldots$ of a vector first, and then the odd ones – this calculation can be done very quickly.

Thus we can do a Fourier transform of size 64 on a vector by separating the vector into its odd and even components, performing a size 32 Fourier transform on each half of its components, then recombining the two halves through a process which involves multiplication by the diagonal matrix $D$.

$$D = \begin{bmatrix} 1 & & & & \\ & w & & & \\ & & w^2 & & \\ & & & \ddots & \\ & & & & w^{n-1} \end{bmatrix}.$$

Of course we can break each of those copies of $F_{32}$ down into two copies of $F_{16}$, and so on. In the end, instead of using $n^2$ operations to multiply by $F_n$ we get the same result using about $\frac{1}{2}n\log_2 n$ operations.

A typical case is $n = 1024 = 2^{10}$. Simply multiplying by $F_n$ requires over a million calculations. The fast Fourier transform can be completed with only $\frac{1}{2}n\log_2 n = 5 \cdot 1024$ calculations. This is 200 times faster!

This is only possible because Fourier matrices are special matrices with orthogonal columns. In the next lecture we'll return to dealing exclusively with real numbers and will learn about positive definite matrices, which are the matrices most often seen in applications.

MIT OpenCourseWare
http://ocw.mit.edu

18.06SC Linear Algebra
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

# LA Lec 26

Complex Vectors & matrices → Inner Products

Discrete Fourier

FFT (FAST Fourier transform)

$\boxed{\text{Real matrices can have complex eigenvalues \& vectors}}$

This lecture will deal with complex vector & matrices.

---

An n by n matrix (with orthogonal columns) has $n^2$ multiplication

FFT will reduce it to $n \log_2 n$.

Let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ $x_1, \dots x_n$ are in $\mathbb{C}^n$ (n-dimensional complex domain)

Length$(x)$ = $x^T x$ is no good

$= \bar{x}^T x$ is good (because it will give the length)

$\bar{x}$: complex conjugat

Let $x = \begin{bmatrix} 1 \\ i \end{bmatrix}$ => $\bar{x} = \begin{bmatrix} 1 \\ -i \end{bmatrix}$ => $\bar{x}^T x = [1 \ \ i]\begin{bmatrix} 1 \\ -i \end{bmatrix} = 2$

Now with complex vectors we will denoted $x^H x$ as $\bar{x}^T x$.

$(x^{\text{conjugate}})^{\text{Transpose}} = x^{\text{Hermitian}}$ => $\boxed{\bar{x}^T x = x^H x}$

$x$ Hermitian

## Inner Product

It is no longer $y^T x$.

Now it is $\bar{y}^T x$ or $y^H x$.

$$\boxed{x^H x = |x_1|^2 + |x_2|^2 + \cdots + |x_n|^2}$$

Symmetric $A^T = A$ No good if $A$ complex

Here $\bar{A}^T = A$

Ex. $A = \begin{bmatrix} 2 & 3+i \\ 3-i & 5 \end{bmatrix}$ then $\bar{A}^T = \begin{bmatrix} 2 & \overline{3+i} \\ \overline{3-i} & 5 \end{bmatrix}$

They are Hermitian matrix $A^H = A$. They have real eigen values & perpendicular eigen vectors.

Perpendicular $q_1, q_2, \ldots q_n$ (Also they are unit length)

$$\bar{q}_i^T q_j = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$$

$$Q = \begin{bmatrix} q_1 & q_2 & \ldots & q_n \end{bmatrix}$$

$Q^T Q = I$ (orthogonal real matrix)

If $Q$ is complex then

$$\boxed{\bar{Q}^T Q = Q^H Q = I}$$

Orthogonal will be called unitary now.

Unitary matrix is $N \times N$ matrix have $N$ orthogonal ~~ve~~ columns.

Fourier Matrix

$$F_n = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & \omega & \omega^2 & \cdots & \omega^{n-1} \\ 1 & \omega^2 & \omega^4 & \cdots & \omega^{2(n-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \omega^{n-1} & \omega^{2(n-1)} & \cdots & \omega^{(n-1)(n-1)} \end{bmatrix}$$

$(n \times n)$ ~~(symmetric)~~

(Very important matrix)

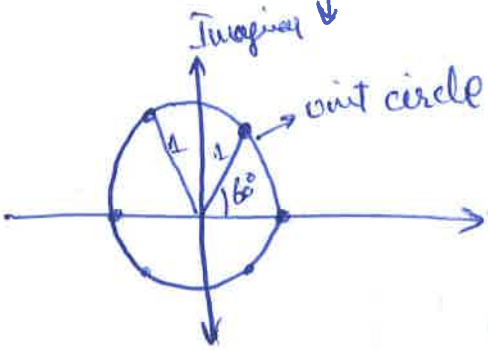
$$\Rightarrow (F_n)_{ij} = \omega^{ij} \qquad i,j = 0, \dots, n-1$$

Full matrix. None of them is zero.

$\omega \rightarrow$ special number whose $n^{th}$ power $= 1$

$$\Rightarrow \omega^n = 1$$

$$\omega = e^{i(2\pi/n)} = \cos\left(\frac{2\pi}{n}\right) + i \sin\left(\frac{2\pi}{n}\right)$$




when $n=6$ : $e^{i\left(\frac{2\pi}{6}\right)}$

All powers of $\omega$ lies on unit circle

$$|e^{i\theta}|^2 = |\cos\theta|^2 + |\sin\theta|^2$$

when $n=4$ what $\omega$ then?

then $\omega^4 = 1$

$$\omega = e^{i2\pi/4} = e^{i\pi/2} = i$$

$$i,\ i^2 = -1,\ i^3 = -i,\ i^4 = 1$$

(we can raise powers of $\omega$ & they all lie on unit circle)

Now write $F_4$:

$$F_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & i^2 & i^3 \\ 1 & i^2 & i^4 & i^6 \\ 1 & i^3 & i^6 & i^9 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}$$

Why $F_4$ is remarkable? Its columns are orthogonal. You can break it into lots of zeroes.

Inner product of col 2 & col 4: $\left.\begin{aligned}&(1)(1)+(i)(-i)+(-1)(-1)+(-i)(i) \\ &= 1 + 1 + 1 + 1\end{aligned}\right\}$ Not zero.

we have to conjugate on column

$$= (1)(1)+(-i)(-i)+(-1)(-1)+(i)(i)$$
$$= 1 - 1 + 1 - 1$$
$$= 0$$

$|col1| = 4$
$|col2| = \cancel{2}4$
$|col3| = 4$
$|col4| = 4$

To make $F_4$ orthonormal.

$$F_4 = \underbrace{\frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & -1 & 1 & -1 \\ 1 & -i & -1 & i \end{bmatrix}}_{\text{orthonormal}}$$

Now $F_4^H F_4 = I$ (As $F_4$ is orthonormal)

$$\boxed{F_4^{-1} = F_4^H}$$

# FFT

Idea $w_n = e^{i 2\pi/n}$

$w_{32} = (w_{64})^2$

$F_{64}$ is connected to two copies of $F_{32}$.

$$\Rightarrow F_{64} = \underbrace{\begin{bmatrix} & \\ & \end{bmatrix}}_{\text{Fixup matrix}} \begin{bmatrix} F_{32} & 0 \\ 0 & F_{32} \end{bmatrix} \underbrace{\begin{bmatrix} & \\ & \end{bmatrix}}_{\text{Fixup } \cancel{\text{Factor}} \text{ matrix}}$$

The beauty is : $F_{64}$ has $64^2$ calculations

But $\begin{bmatrix} & \\ & \end{bmatrix} \begin{bmatrix} F_{32} & 0 \\ 0 & F_{32} \end{bmatrix} \begin{bmatrix} & \\ & \end{bmatrix}$ has $2(32)^2 +$ Fixup calculations

$$\begin{bmatrix} F_{64} \end{bmatrix} = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} F_{32} & 0 \\ 0 & F_{32} \end{bmatrix} \underbrace{\begin{bmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{bmatrix}}_{P}$$

Permutation matrix
(Takes even elements then odd)

Fixup cost is lying in D only.

$$\Rightarrow \text{cost of decomposition} = 2(32)^2 + 32$$

$$D = \begin{bmatrix} 1 & & & \\ & w^2 & & \\ & & \ddots & \\ & & & w^{31} \end{bmatrix}$$

$$F_{32} = \begin{bmatrix} I & D \\ I & -D \end{bmatrix} \begin{bmatrix} & 0 \\ 0 & \end{bmatrix} \begin{bmatrix} F_{16}\ F_{16} & \\ & F_{16}\ F_{16} \end{bmatrix} \begin{bmatrix} P & \\ & P \end{bmatrix} \begin{bmatrix} P \\ P \end{bmatrix}$$

$2(32)^2+32$ is now $2\left[2(16)^2+16\right]+32$

Doing recursion more & more ~~with~~

finally we will get $6 \times 32$

$$\boxed{\text{Final Count} = \frac{1}{2} n \log_2 n}$$

Suppos $n = 1024 = 2^{10}$

$n^2 > 1000\,000$

$$\frac{1}{2} n \log_2 n = \frac{1}{2}(1024) \log_2(1024)$$
$$= \frac{1}{2}(1024)(10)$$
$$= 5(1024)$$ as compared to $1024 \times 1024$

Recitation??

Unit III :- Lecture III

"In calculus, the second derivative decides whether a critical point of $y(x)$ is minimum. For functions of multiple variables, the test is whether a matrix of second derivatives is positive definite. In this session we learn several ways of testing for positive definiteness and also how the shape of the graph of $f(x) = x^T A x$ is determined by the entries of $A$."

# Positive definite matrices and minima

Studying positive definite matrices brings the whole course together; we use pivots, determinants, eigenvalues and stability. The new quantity here is $\mathbf{x}^T A\mathbf{x}$; watch for it.

This lecture covers how to tell if a matrix is positive definite, what it means for it to be positive definite, and some geometry.

## Positive definite matrices

Given a symmetric two by two matrix $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$, here are four ways to tell if it's positive definite:

1. Eigenvalue test: $\lambda_1 > 0, \lambda_2 > 0$.
2. Determinants test: $a > 0, ac - b^2 > 0$.
3. Pivot test: $a > 0, \dfrac{ac - b^2}{a} > 0$.
4. $\mathbf{x}^T A\mathbf{x}$ is positive except when $\mathbf{x} = \mathbf{0}$ (this is usually the definition of positive definiteness).

### 2 by 2

Using the determinants test, we know that $\begin{bmatrix} 2 & 6 \\ 6 & y \end{bmatrix}$ is positive definite when $2y - 36 > 0$ or when $y > 18$.

The matrix $\begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ is on the borderline of positive definiteness and is called a *positive semidefinite* matrix. It's a singular matrix with eigenvalues 0 and 20. Positive semidefinite matrices have eigenvalues greater than or equal to 0. For a singular matrix, the determinant is 0 and it only has one pivot.

$$\begin{aligned}
\mathbf{x}^T A\mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\
&= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 2x_1 + 6x_2 \\ 6x_1 + 18x_2 \end{bmatrix} \\
&= 2x_1^2 + 12x_1x_2 + 18x_2^2 \\
&= ax_1^2 + 2bx_1x_2 + cx_2^2.
\end{aligned}$$

If this *quadratic form* is positive for every (real) $x_1$ and $x_2$ then the matrix is positive definite. In this positive semi-definite example, $2x_1^2 + 12x_1x_2 + 18x_2^2 = 2(x_1 + 3x_2)^2 = 0$ when $x_1 = 3$ and $x_2 = -1$.

## Tests for minimum

If we apply the fourth test to the matrix $\begin{bmatrix} 2 & 6 \\ 6 & 7 \end{bmatrix}$ which is not positive definite, we get the quadratic form $f(x,y) = 2x^2 + 12xy + 7y^2$. The graph of this function has a saddle point at the origin; see Figure 1.

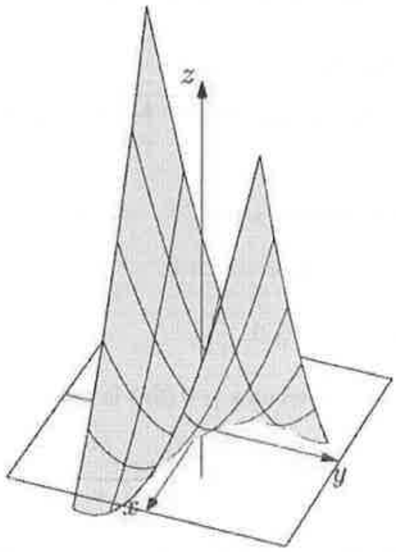



Figure 1: The graph of $f(x,y) = 2x^2 + 12xy + 7y^2$.

The matrix $\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$ is positive definite – its determinant is 4 and its trace is 22 so its eigenvalues are positive. The quadratic form associated with this matrix is $f(x,y) = 2x^2 + 12xy + 20y^2$, which is positive except when $x = y = 0$. The level curves $f(x,y) = k$ of this graph are ellipses; its graph appears in Figure 2. If $a > 0$ and $c > 0$, the quadratic form $ax^2 + 2bxy + cy^2$ is only negative when the value of $2bxy$ is negative and overwhelms the (positive) value of $ax^2 + cy^2$.

The first derivatives $f_x$ and $f_y$ of this function are zero, so its graph is tangent to the $xy$-plane at $(0,0,0)$; but this was also true of $2x^2 + 12xy + 7y^2$. As in single variable calculus, we need to look at the second derivatives of $f$ to tell whether there is a minimum at the critical point.

We can prove that $2x^2 + 12xy + 20y^2$ is always positive by writing it as a sum of squares. We do this by completing the square:

$$2x^2 + 12xy + 20y^2 = 2(x + 3y)^2 + 2y^2.$$

Note that $2(x+3y)^2 = 2x^2 + 12xy + 18y^2$, and 18 was the "borderline" between passing and failing the tests for positive definiteness.

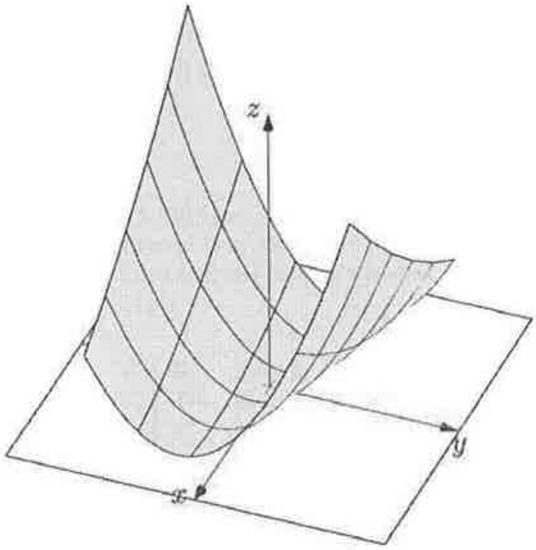



Figure 2: The graph of $f(x,y) = 2x^2 + 12xy + 20y^2$.

When we complete the square for $2x^2 + 12xy + 7y^2$ we get:

$$2x^2 + 12xy + 7y^2 = 2(x+3y)^2 - 11y^2$$

which may be negative; e.g. when $x = -3$ and $y = 1$.

The coefficients that appear when completing the square are exactly the entries that appear when performing elimination on the original matrix. The two pivots are multiplied by the squares, and the coefficient $c$ in the term $(x - cy)^2$ is the multiple of the first row that's subtracted from the second row.

$$\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix} \xrightarrow{\text{subtract 3 times row 1}} \begin{bmatrix} 2 & 6 \\ 0 & 2 \end{bmatrix}.$$

We can see the terms that appear when completing the square in:

$$U = \begin{bmatrix} 2 & 6 \\ 0 & 2 \end{bmatrix}, \quad \text{and} \quad L = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}.$$

When we complete the square, the numbers multiplied by the squares are the pivots; if the pivots are all positive then the sum of squares will always be positive.

**Hessian matrix**

The matrix of second derivatives of $f(x,y)$ is:

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}.$$

This matrix is symmetric because $f_{xy} = f_{yx}$. Its determinant is positive when the matrix is positive definite, which matches the $f_{xx}f_{yy} > f_{xy}^2$ test for a minimum that we learned in calculus.

### $n$ by $n$

A function of several variables $f(x_1, x_2, ..., x_n)$ has a minimum when its matrix of second derivatives is positive definite, and identifying minima of functions is often important. The tests we've just learned for 2 by 2 matrices also apply to $n$ by $n$ matrices.

A 3 by 3 example:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}.$$

Is this matrix positive definite? Our tests will say *yes*. What's the function $\mathbf{x}^T A\mathbf{x}$ associated with this matrix? Does that function have a minimum at $\mathbf{x} = \mathbf{0}$? What does the graph of its quadratic form look like?

Looking at determinants we see:

$$\det[\,2\,] = 2, \qquad \det\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} = 5, \qquad \det\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = 4.$$

These are all positive, so $A$ is positive definite.

The pivots of $A$ are $2, 3/2$ and $4/3$ (all positive) because the products of the pivots equal the determinants.

The eigenvalues of $A$ are positive and their product is 4. It's not difficult to check that they are $2 - \sqrt{2}, 2$ and $2 + \sqrt{2}$ (all positive).

## Ellipsoids in $\mathbb{R}^n$

$$f(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = 2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3.$$

Because $A$ is positive definite, we expect $f(\mathbf{x})$ to be positive except when $\mathbf{x} = \mathbf{0}$. Its graph is a sort of four dimensional bowl or *paraboloid*. If we wrote $f(\mathbf{x})$ as a sum of three squares, those squares would be multiplied by the (positive) pivots of $A$. Earlier, we said that a horizontal slice of our three dimensional bowl shape would be an ellipse. Here, a horizontal slice of the four dimensional bowl is an ellipsoid – a little bit like a rugby ball. For example, if we cut the graph at height 1 we get a surface whose equation is: $2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 = 1$.

Just as an ellipse has a major and minor axis, an ellipsoid has three axes. If we write $A = Q\Lambda Q^T$, as the principal axis theorem tells us we can, the eigenvectors of $A$ tell us the directions of the principal axes of the ellipsoid. The eigenvalues tell us the lengths of those axes.

MIT OpenCourseWare
http://ocw.mit.edu

18.06SC Linear Algebra
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

## LA Lecture 27

Positive Definite Matrix (Tests)
Test for Minimum ($x^TAx > 0$)
Ellipsoids in $\mathbb{R}^N$

TEST 2x2 case

① $A = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$

⋮

Geometrically, ellipses are connected to positive definitness while ~~hyperbola~~ Parabolas aren't.

Now all our matrices are symmetric. Consider 2x2 symmetric matrix. To qualify for positive-definite matrix, A has to satisfy following tests: If $A = \begin{bmatrix} a & b \\ b & c \end{bmatrix}$

① $\lambda_1 > 0 \quad \lambda_2 > 0$

② Sub-determinant should be +ve
i.e $a > 0 \quad ac - b^2 > 0$

③ pivots are +ve i.e
$a > 0$
$\dfrac{ac - b^2}{a} > 0$

④ $x^TAx > 0$ ( In most application, $x^TAx > 0$ will be the definition of PD & 1,2,3 will be

1st Pivot $= a$
2nd Pivot $=$ 1st Row $\times \frac{b}{a}$

$\begin{bmatrix} a & b \\ b - a\left(\frac{b}{a}\right) & c - b\left(\frac{b}{a}\right) \end{bmatrix}$

$\begin{bmatrix} \boxed{a} & b \\ b - b & \boxed{c - \frac{b^2}{a}} \end{bmatrix}$

## Examples

$$A = \begin{bmatrix} 2 & 6 \\ 6 & \boxed{?} \end{bmatrix} \rightarrow \text{what should be here for it to be P.D?}$$

i) If $\begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix}$ Its positive-semidefinite. Its a singular matrix & thus one of its eigenvalue is zero.

Trace $= \lambda_1 + \lambda_2$ If $\lambda_1 = 0$
$= 20$ $\Rightarrow \lambda_2 = 20$

Since its Eigen Value $\geq 0$, therefore it is called PSD matrix.

Pivots Are 2, No Pivot (since its a singular matrix). Its rank $= 1$

Novelty $X \rightarrow$ any vector

Let $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$\Rightarrow X^T A X = [x_1 \ x_2] \begin{bmatrix} 2 & 6 \\ 6 & 18 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$ (we will get some function of $x_1$ & $x_2$)

$$= [x_1 \ x_2] \begin{bmatrix} 2x_1 + 6x_2 \\ 6x_1 + 18x_2 \end{bmatrix}$$

$$= 2x_1^2 + 6x_1x_2 + 6x_1x_2 + 18x_2^2$$

$$= \underbrace{2}_{a}x_1^2 + \underbrace{12}_{2b}x_1x_2 + \underbrace{18}_{c}x_2^2$$

$$= ax^2 + 2bxy + cy^2$$ (Quadratic form = degree 2, poly)

Is $2x_1^2 + 12x_1x_2 + 18x_2^2 > 0$ for every $x_1, x_2$? This is my definition for Positive definitness. If it is positive for all $(x_1, x_2)$ then my matrix A is PD.

Lets assume that $A=\begin{bmatrix}2 & 6\\ 6 & 7\end{bmatrix}$. This Matrix is not P.D. B/c $|A|<0$. Its pivot are 2, -ve. So how can we look at quantity $x^TAx$?

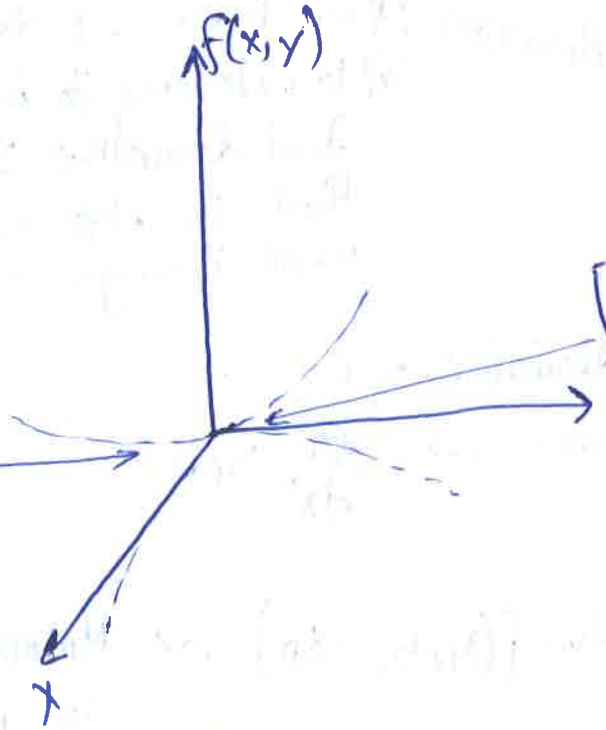
$$x^TAx = [x_1\ x_2]\begin{bmatrix}2x_1+6x_2\\ 6x_1+7x_2\end{bmatrix} = 2x_1^2+6x_1x_2+6x_1x_2+7x_2^2$$

Graph of $f(x,y) = \vec{x}^T A \vec{x}$

$$= ax^2+2bxy+y^2$$

$$\Rightarrow \boxed{2x^2+12xy+7y^2 = 0}$$

Draw it




i) 1st derivative $=0$ @ saddle point

2) 2nd derivative

Saddle point: At this point graph ~~is going up~~ has maximum in one direction & ~~going~~ minimum in other direction.

So Actually, the perfect direction to look at is the direction given by eigenvectors.

Now $A=\begin{bmatrix}2 & 6\\ 6 & 20\end{bmatrix}$ will lead to $2x_1^2+12x_1x_2+20x_2^2$.

It passes the test for:

i) Determinant $>0$

ii) Subdeterminant $>0$

ii) How do I know that its eigenvalues are +ve?

$\boxed{\text{Trace}=22 \quad \lambda_1+\lambda_2=22}$ $\boxed{\text{Det}=4 \quad \lambda_1\lambda_2=4}$ ← see here:

Thus $\lambda_1, \lambda_2 > 0$

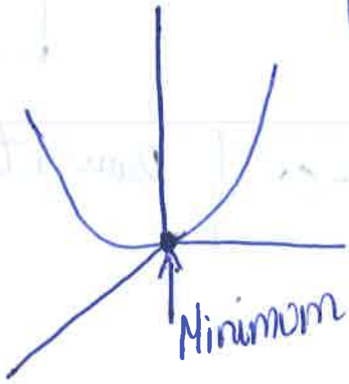
iii) Pivots are +ve

iv) $x^TAx$ ??

Graph of $2x_1^2 + 12x_1x_2 + 20x_2^2$ (It does not have saddle point)

$$f(x,y) = 2x^2 + 12xy + 20y^2$$

$\boxed{\text{Its a bowl}}$



Minimum : i) we know 1st derivative $= 0$
ii) To make sure its is minimum, its 2nd derivative $> 0$ (which means that its slope is increasing as you went through origin).

In Calculus : 1st derivative $= 0$
Minimum $\sim$ $\frac{d^2v}{dx^2} > 0$

In Linear Algebra. Min $f(x_1, x_2, \dots x_n)$ $\sim$ MATRIX of 2nd derivatives is Positive definite

Now how to ~~...~~ make sure that $f(x,y) = 2x^2 + 12xy + 20y^2$ is always positive? Completing the square method gives us.

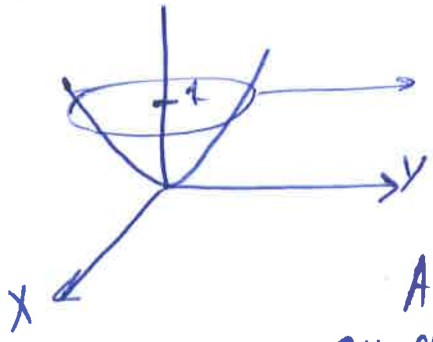
$$f(x,y) = 2(x+3y)^2 + 2y^2$$

This means for $\forall (x,y)$ $f(x,y) \geq 0$
Hence $x^TAx$ for $A = \begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}$ is +ve &
A is PSD.

what if $f(x,y) = 2x^2 + 12xy + 18y^2$ (Marginal case)
$= 2(x+3y)^2$

when $f(x,y) = 2x^2 + 12xy + 7y^2$ (Not PD)
$= 2(x+3y)^2 - 11y^2$

Now see, $f(x,y) = 2x^2+12xy+20y^2 = 2(x+3y)^2+2y^2$




If I cut this bowl at $z=1$ I will get curve whose equation would be

$$2x^2+12xy+20y^2=1$$

And that curve would be an ellipse. whereas if I cut through saddle point of $2x^2+12xy+7y^2=0$, we get hyperbola.

* Completing the square is Gaussian elimination we we have seen in $f(x,y) = 2x^2+12xy+20y^2$

Cut it at $f(x,y)=1 \Rightarrow 2(x+3y)^2+2y^2=1$

<u>Elimination</u>

$$\overset{A}{\begin{bmatrix} 2 & 6 \\ 6 & 20 \end{bmatrix}} \longrightarrow \overset{U}{\begin{bmatrix} 2 & 6 \\ 0 & 2 \end{bmatrix}}$$

$$L = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix}$$

we can see pivots outsides

$$\underset{\text{+ve pivots}}{2}(x+3y)^2 + \underset{\text{+ve pivots}}{2}y^2$$

<u>2nd Derivative</u>

In 2D: $\begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$

$\boxed{\text{Matrix of 2nd Derivatives}}$

$f_{xy} = f_{yx}$ (Partial derivate is symmetric)

$$f_{xy} = \frac{d}{dx}\left[\frac{d}{dy}(f)\right]$$

$$f_{xx} = \frac{d}{dx}\left(\frac{d}{dx}f\right)$$

$$f_{yy} = \frac{d}{dy}\left(\frac{d}{dy}f(x,y)\right)$$

For $A$ to be P.D, Its matrix $x^TAx$ for 2nd derivative has to P.D

## In Calculus:

Minimum for $f(x,y)$

i) $\frac{d}{dx} f(x,y) = 0$ , $\frac{d}{dy} f(x,y) = 0$

ii) Matrix of 2nd derivative has to be P.D

i.e $f_{xx} f_{yy} > f_{xy} f_{yx}$

## For 3x3 or nxn Matrices

i) 3x3 Example

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

Q. Is it PD?

Q. What $x^TAx$?

Q. Do we have minimum at $x = [0,0,0]^T$

$|A_1| = |2| = 2$ $\quad |A_2| = \begin{vmatrix} 2 & -1 \\ -1 & 2 \end{vmatrix} = 3$ $\quad |A_3| = |A| = 4$

Pivots $= 2, \frac{3}{2}, \frac{4}{3}$ (B/c product of pivots is determinant)

Eigen Values = ??

How to get them? Eigen Values $\lambda_1, \lambda_2, \lambda_3 > 0$

$$\lambda_1 = 2-\sqrt{2} \quad \lambda_2 = 2, \quad \lambda_3 = 2+\sqrt{2}$$

$* \ \lambda_1+\lambda_2+\lambda_3 = \text{trace} = 6$
$* \ \lambda_1\lambda_2\lambda_3 = \text{determinant} = 4$ } Test for eigenvalues

$$f = x^T A x = 2x_1^2 + 2x_2^2 + 2\lambda_3^2 - 2x_1x_2 - 2x_2x_3 \underbrace{> 0}_{\text{we belive } x^TAx > 0}$$

Its graph is bowl (paraboloid)?

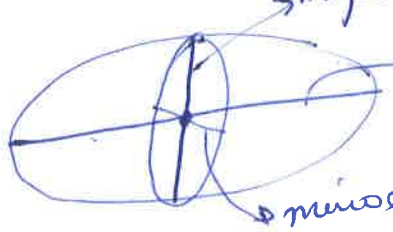
This "f" will go up b/c all pivots are +ve & we can see that when we will recomplete the squares (Gaussian elimination), we will get something.

$$f = (\text{pivots})(\text{some number})^2 + (\text{pivots})(\text{some number})^2$$

If I ~~cut~~ cut through height "1" that is:

$$2x_1^2 + 2x_2^2 + 2x_3^2 - 2x_1x_2 - 2x_2x_3 = 1$$

we will get ellipsoid (we get ellipse in 2D when $A = 2\times 2$). Its Equation of football. It has gotten three principal axis.




Sphere = 3 eigen values (all same)
Rugby Ball = 3 eigen values (2 same, 1 different) } length of axis is given by eigenvalue

* Principal Axis is determined by eigenVector.

## Principal Axis theorem

$$A = \underbrace{Q \Lambda Q^T}$$

Matrix factorization / diagonalization with symmetric matrices.

## Recitation

Unit III : Lecture IV

"After a final discussion of Positive definite matrices, we learn about similar matrices: $B = M^{-1}AM$ for some invertible matrix M. Square matrices can be grouped by similarity, and each group has a "nicest" representative in Jordan normal form. This form tells at a glance the eigenvalues & the number of eigenvectors."

# Similar matrices and Jordan form

We've nearly covered the entire heart of linear algebra – once we've finished singular value decompositions we'll have seen all the most central topics.

## $A^TA$ is positive definite

A matrix is *positive definite* if $\mathbf{x}^TA\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$. This is a very important class of matrices; positive definite matrices appear in the form of $A^TA$ when computing least squares solutions. In many situations, a rectangular matrix is multiplied by its transpose to get a square matrix.

Given a symmetric positive definite matrix $A$, is its inverse also symmetric and positive definite? Yes, because if the (positive) eigenvalues of $A$ are $\lambda_1, \lambda_2, \cdots \lambda_d$ then the eigenvalues $1/\lambda_1, 1/\lambda_2, \cdots 1/\lambda_d$ of $A^{-1}$ are also positive.

If $A$ and $B$ are positive definite, is $A + B$ positive definite? We don't know much about the eigenvalues of $A + B$, but we can use the property $\mathbf{x}^TA\mathbf{x} > 0$ and $\mathbf{x}^TB\mathbf{x} > 0$ to show that $\mathbf{x}^T(A+B)\mathbf{x} > 0$ for $\mathbf{x} \neq 0$ and so $A + B$ is also positive definite.

Now suppose $A$ is a rectangular ($m$ by $n$) matrix. $A$ is almost certainly not symmetric, but $A^TA$ is square and symmetric. Is $A^TA$ positive definite? We'd rather not try to find the eigenvalues or the pivots of this matrix, so we ask when $\mathbf{x}^TA^TA\mathbf{x}$ is positive.

Simplifying $\mathbf{x}^TA^TA\mathbf{x}$ is just a matter of moving parentheses:

$$\mathbf{x}^T(A^TA)\mathbf{x} = (A\mathbf{x})^T(A\mathbf{x}) = |A\mathbf{x}|^2 \geq 0.$$

The only remaining question is whether $A\mathbf{x} = \mathbf{0}$. If $A$ has rank $n$ (independent columns), then $\mathbf{x}^T(A^TA)\mathbf{x} = \mathbf{0}$ only if $\mathbf{x} = \mathbf{0}$ and $A$ is positive definite.

Another nice feature of positive definite matrices is that you never have to do row exchanges when row reducing – there are never 0's or unsuitably small numbers in their pivot positions.

## Similar matrices $A$ and $B = M^{-1}AM$

Two square matrices $A$ and $B$ are *similar* if $B = M^{-1}AM$ for some matrix $M$. This allows us to put matrices into families in which all the matrices in a family are similar to each other. Then each family can be represented by a diagonal (or nearly diagonal) matrix.

### Distinct eigenvalues

If $A$ has a full set of eigenvectors we can create its eigenvector matrix $S$ and write $S^{-1}AS = \Lambda$. So $A$ is similar to $\Lambda$ (choosing $M$ to be this $S$).

If $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ then $\Lambda = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$ and so $A$ is similar to $\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$. But $A$ is also similar to:

$$\begin{array}{ccccc} M^{-1} & A & M & & B \\ \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix} & = & \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}\begin{bmatrix} 2 & 9 \\ 1 & 6 \end{bmatrix} \\ & & & = & \begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}. \end{array}$$

In addition, $B$ is similar to $\Lambda$. All these similar matrices have the same eigenvalues, 3 and 1; we can check this by computing the trace and determinant of $A$ and $B$.

Similar matrices have the same eigenvalues!

In fact, the matrices similar to $A$ are all the 2 by 2 matrices with eigenvalues 3 and 1. Some other members of this family are $\begin{bmatrix} 3 & 7 \\ 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} 1 & 7 \\ 0 & 3 \end{bmatrix}$. To prove that similar matrices have the same eigenvalues, suppose $A\mathbf{x} = \lambda\mathbf{x}$. We modify this equation to include $B = M^{-1}AM$:

$$\begin{aligned} AMM^{-1}\mathbf{x} &= \lambda\mathbf{x} \\ M^{-1}AMM^{-1}\mathbf{x} &= \lambda M^{-1}\mathbf{x} \\ BM^{-1}\mathbf{x} &= \lambda M^{-1}\mathbf{x}. \end{aligned}$$

The matrix $B$ has the same $\lambda$ as an eigenvalue. $M^{-1}\mathbf{x}$ is the eigenvector.

If two matrices are similar, they have the same eigenvalues and the same number of independent eigenvectors (but probably not the same eigenvectors).

When we diagonalize $A$, we're finding a diagonal matrix $\Lambda$ that is similar to $A$. If two matrices have the same $n$ distinct eigenvalues, they'll be similar to the same diagonal matrix.

### Repeated eigenvalues

If two eigenvalues of $A$ are the same, it may not be possible to diagonalize $A$. Suppose $\lambda_1 = \lambda_2 = 4$. One family of matrices with eigenvalues 4 and 4 contains only the matrix $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$. The matrix $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ is not in this family.

There are two families of similar matrices with eigenvalues 4 and 4. The larger family includes $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$. Each of the members of this family has only one eigenvector.

The matrix $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ is the only member of the other family, because:

$$M^{-1}\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}M = 4M^{-1}M = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$$

for any invertible matrix $M$.

## Jordan form

Camille Jordan found a way to choose a "most diagonal" representative from each family of similar matrices; this representative is said to be in *Jordan normal form*. For example, both $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ and $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ are in Jordan form. This form used to be the climax of linear algebra, but not any more. Numerical applications rarely need it.

We can find more members of the family represented by $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ by choosing diagonal entries to get a trace of 4, then choosing off-diagonal entries to get a determinant of 16:

$$\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}, \begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 17 & 4 \end{bmatrix}, \begin{bmatrix} a & b \\ (8a - a^2 - 16)/b & 8 - a \end{bmatrix}.$$

(None of these are diagonalizable, because if they were they would be similar to $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$. That matrix is only similar to itself.) What about this one?

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Its eigenvalues are four zeros. Its rank is 2 so the dimension of its nullspace is $4 - 2 = 2$. It will have two independent eigenvectors and two "missing" eigenvectors. When we look instead at

$$\begin{bmatrix} 0 & 1 & 7 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

its rank and the dimension of its nullspace are still 2, but it's not as nice as $A$. $B$ *is* similar to $A$, which is the Jordan normal form representative of this family. $A$ has a 1 above the diagonal for every missing eigenvector and the rest of its entries are 0.

Now consider:

$$C = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Again it has rank 2 and its nullspace has dimension 2. Its four eigenvalues are 0. Surprisingly, it is not similar to $A$. We can see this by breaking the matrices into their *Jordan blocks*:

$$A = \left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right], \qquad C = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right].$$

A Jordan block $J_i$ has a repeated eigenvalue $\lambda_i$ on the diagonal, zeros below the diagonal and in the upper right hand corner, and ones above the diagonal:

$$J_i = \begin{bmatrix} \lambda_i & 1 & 0 & \cdots & 0 \\ 0 & \lambda_i & 1 & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & \lambda_i & 1 \\ 0 & 0 & \cdots & 0 & \lambda_i \end{bmatrix}.$$

Two matrices may have the same eigenvalues and the same number of eigenvectors, but if their Jordan blocks are different sizes those matrices can not be similar.

Jordan's theorem says that every square matrix $A$ is similar to a Jordan matrix $J$, with Jordan blocks on the diagonal:

$$J = \begin{bmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & J_d \end{bmatrix}.$$

In a Jordan matrix, the eigenvalues are on the diagonal and there may be ones above the diagonal; the rest of the entries are zero. The number of blocks is the number of eigenvectors – there is one eigenvector per block.

To summarize:

- If $A$ has $n$ distinct eigenvalues, it is diagonalizable and its Jordan matrix is the diagonal matrix $J = \Lambda$.
- If $A$ has repeated eigenvalues and "missing" eigenvectors, then its Jordan matrix will have $n - d$ ones above the diagonal.

We have not learned to compute the Jordan matrix of a matrix which is missing eigenvectors, but we do know how to diagonalize a matrix which has $n$ distinct eigenvalues.

MIT OpenCourseWare
http://ocw.mit.edu

18.06SC Linear Algebra
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

L#28:

$A^TA$ is positive definite!

Similar Matrices $A, B$ / Jordan Form

$B = M^{-1}AM$

Positive definite means
$x^TAx > 0$ (except for $x = 0$)

Positive definite matrices comes from least squares. When we say positive definite matrix, we mean that it is symmetric as well.

Now if $A, B$ are PD, Is $A+B$ a pd matrix.

If $x^TAx > 0$
and $x^TBx > 0$

so $\boxed{x^T(A+B)x > 0}$

If $A$ is PD, its inverse($A^{-1}$) will also be PD b/c $eig(A^{-1}) = \lambda_1^{-1}, \lambda_2^{-1} \cdots \lambda_M^{-1}$ if $\lambda_1, \lambda_2 \cdots \lambda_M = eig(A)$

Suppose $A = m \times n$ (Rectangular Matrix) $\neq$ PD as it is not symmetric

Now $A^TA$ = square, symmetric Matrix

Is $A^TA$ Positive definite or at least PSD?

Now have a look over $x^TA^TAx$.

Now $x^TA^TAx = (Ax)^T(Ax)$

$= \|Ax\|^2 \geq 0$ ($\|Ax\| = 0$ when $x$ = Null Vector

We want $\|Ax\| > 0$, This mean "A" doesnt have nullspace. This implies $\text{Rank}(A) = n$ As $A = m \times n$

Hence rectangular matrix $A = m \times n$ with rank $= n$, $A^TA$ is always P.D. With PD matrices, we dont need row exchanges & we never run into numerical issues.

## SIMILAR MATRICES

No longer Symmetric matrices

Let A and B are $\overset{A,B}{(n \times n)}$ similar means:

for some $M$: $B = M^{-1}AM$

Example: A is similar to $\Lambda$ when

$$S^{-1}AS = \Lambda$$

S = contains eigenvectors for A
A = has full set of eigenvector
$\Lambda$ = diagonal matrix with eigenvalues

Let $A = \overbrace{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}}^{\lambda = 3,1}$ $\quad \Lambda = \underbrace{\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}}_{\text{eigenvalue matrix for } A}$ $\quad \boxed{A \text{ \& } \Lambda \text{ are similar}}$

Also

$$\overset{M^{-1}}{\begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix}} \overset{A}{\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}} \overset{M}{\begin{bmatrix} 1 & 4 \\ 0 & 1 \end{bmatrix}} = \begin{bmatrix} 1 & -4 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 9 \\ 1 & 6 \end{bmatrix} = \underbrace{\begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}}_{B}$$

A, B are similar matrices. They have something in common. Because they have same eigenvalues

i.e $\lambda_A = 3,1 \qquad \lambda_B = 3,1$

Similar matrices have the same eigenvalues $\lambda$'s.

Other examples of matrices having eigen value 3, 1 are

$$\underbrace{\begin{bmatrix} 3 & 7 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 7 \\ 0 & 3 \end{bmatrix}}_{\text{triangular matrix}}$$

. This means that they are connected to one $A$ $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$ by some $M$

---

MATRICES similar to $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \quad \lambda = 1, 3$$

similar, similar, similar, similar

$$\begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 7 \\ 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 7 \\ 0 & 3 \end{bmatrix} \quad \begin{bmatrix} -2 & -15 \\ 1 & 6 \end{bmatrix}$$

$\lambda = 1, 3$ $\quad$ $\lambda = 1, 3$ $\quad$ $\lambda = 1, 3$ $\quad$ $\lambda = 1, 3$

---

Now why is that?

Let $Ax = \lambda x$ $\qquad (B = M^{-1} A M)$

Eigen vector of $B$ is $M^{-1}$(eigenvector of $A$)

$$A M M^{-1} x = \lambda x$$

$$\Rightarrow \underbrace{M^{-1} A M}_{B} M^{-1} x = \lambda M^{-1} x$$

$$\Rightarrow \boxed{B M^{-1} x = \lambda M^{-1} x}$$

Thus $B$ has ~~sam~~ same eigenvalue as that of $A$ but different eigenvectors.

---

Eigen vectors of diagonal $(2 \times 2)$ = $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

---

BAD CASE

If $\lambda_1 = \lambda_2$ (Repeated eigenvalues) then matrix might not be diagonalised

Let $\lambda_1 = \lambda_2 = 4$

One family has $\overbrace{\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}}^{\lambda = 4, 4}$ $\qquad$ what about $\overbrace{\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}}^{\lambda = 4, 4}$ ?

It doesnot belong to $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$

The big family includes $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ & many others.

If we have $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$:

$$\left.\begin{aligned} \text{Then } M^{-1}\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}M &= 4\,M^{-1}IM \\ &= 4I \\ &= \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix} \end{aligned}\right\}$$ Thus it has only one person i.e. $\begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$

Best Guy = $\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$ ← Jordan form (Most close to diagonal if it was)

More Members of the family with $\lambda = 4, 4$ (trace = 8, det = 16)

$$\begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}, \underbrace{\begin{bmatrix} 5 & 1 \\ -1 & 3 \end{bmatrix}}_{\text{Its not diagonalisable}}, \begin{bmatrix} 4 & 0 \\ 17 & 4 \end{bmatrix} \ldots$$

General $\begin{bmatrix} a & m \\ n & 8-a \end{bmatrix}$ $\left.\begin{matrix} \text{trace} = 8 \\ \text{det} = 16 \end{matrix}\right\}$ I get similar matrices & they all have only one eigen vector.

---

Thus Similar matrices have same $\lambda$'s, same number of eigenvectors

---

If $A = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$ If $A_1 = \underbrace{\begin{bmatrix} 0 & 1 & 7 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}}_{\text{similar to } A}$

$\lambda = 0, 0, 0, 0$ → hence eigenvectors lies in $N(A)$

# of Eigen Vectors : $Ax = 0x$

Dimension of Nullspace $= 2 = 4 - 2$

Rank $= 2$

Thus 2 eigen vectors are missing here

Let

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

Eig Values = 0, 0, 0, 0

Rank(B) = 2

# of Eigenvectors = 2

**Jordan Block** (It has one eigen vector)

$$J_i = \begin{bmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & 1 & \\ & & \ddots & \\ 0 & & & \lambda_i \end{bmatrix}$$

$$\left[\begin{array}{ccc|c} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \end{array}\right] \underbrace{\text{is not similar to}}_{\text{B/C block size is different}} \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

**Jordan's theorem:**

Every square $A$ is similar to Jordan matrix $J$

$$J = \begin{bmatrix} \boxed{J_1} & & & \\ & \boxed{J_2} & & \\ & & \ddots & \\ & & & \boxed{J_d} \end{bmatrix}$$

# of blocks = # of eigen-vectors

Good case: $J$ is $\Lambda$ i.e when $A$ is diagonilizable

i.e $A$ has distinct eigen values.

$\boxed{\text{Recitation ??}}$

Unit III: Lecture 5

1

"If $A$ is symmetric & positive definite, there is an orthogonal matrix $Q$ for which $A = Q\Lambda Q^T$. Here $\Lambda$ is the matrix of eigenvalues. SVD lets us write any matrix $A$ as a product of $U\Sigma V^T$ where $U$ and $V$ are orthogonal & $\Sigma$ is a diagonal matrix whose non-zero entries are square roots of the eigenvalues of $A^TA$. The columns of $U$ & $V$ gives bases for the four fundamental subspaces".

# Singular value decomposition

The *singular value decomposition* of a matrix is usually referred to as the *SVD*. This is the final and best factorization of a matrix:

$$A = U\Sigma V^T$$

where $U$ is orthogonal, $\Sigma$ is diagonal, and $V$ is orthogonal.

In the decomoposition $A = U\Sigma V^T$, $A$ can be *any* matrix. We know that if $A$ is symmetric positive definite its eigenvectors are orthogonal and we can write $A = Q\Lambda Q^T$. This is a special case of a SVD, with $U = V = Q$. For more general $A$, the SVD requires two different matrices $U$ and $V$.

We've also learned how to write $A = S\Lambda S^{-1}$, where $S$ is the matrix of $n$ distinct eigenvectors of $A$. However, $S$ may not be orthogonal; the matrices $U$ and $V$ in the SVD will be.

## How it works

We can think of $A$ as a linear transformation taking a vector $\mathbf{v}_1$ in its row space to a vector $\mathbf{u}_1 = A\mathbf{v}_1$ in its column space. The SVD arises from finding an orthogonal basis for the row space that gets transformed into an orthogonal basis for the column space: $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.

It's not hard to find an orthogonal basis for the row space – the Gram-Schmidt process gives us one right away. But in general, there's no reason to expect $A$ to transform that basis to another orthogonal basis.

You may be wondering about the vectors in the nullspaces of $A$ and $A^T$. These are no problem – zeros on the diagonal of $\Sigma$ will take care of them.

## Matrix language

The heart of the problem is to find an orthonormal basis $\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_r$ for the row space of $A$ for which

$$\begin{aligned} A\begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix} &= \begin{bmatrix} \sigma_1\mathbf{u}_1 & \sigma_2\mathbf{u}_2 & \cdots & \sigma_r\mathbf{u}_r \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_r \end{bmatrix} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}, \end{aligned}$$

with $\mathbf{u}_1, \mathbf{u}_2, ...\mathbf{u}_r$ an orthonormal basis for the column space of $A$. Once we add in the nullspaces, this equation will become $AV = U\Sigma$. (We can complete the orthonormal bases $\mathbf{v}_1, ...\mathbf{v}_r$ and $\mathbf{u}_1, ...\mathbf{u}_r$ to orthonormal bases for the entire space any way we want. Since $\mathbf{v}_{r+1}, ...\mathbf{v}_n$ will be in the nullspace of $A$, the diagonal entries $\sigma_{r+1}, ...\sigma_n$ will be 0.)

The columns of $U$ and $V$ are bases for the ~~row~~ column and ~~column~~ row spaces, respectively. Usually $U \neq V$, but if $A$ is positive definite we can use the *same* basis for its row and column space!

2

## Calculation

Suppose $A$ is the invertible matrix $\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We want to find vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ in the row space $\mathbb{R}^2$, $\mathbf{u}_1$ and $\mathbf{u}_2$ in the column space $\mathbb{R}^2$, and positive numbers $\sigma_1$ and $\sigma_2$ so that the $\mathbf{v}_i$ are orthonormal, the $\mathbf{u}_i$ are orthonormal, and the $\sigma_i$ are the scaling factors for which $A\mathbf{v}_i = \sigma_i u_i$.

This is a big step toward finding orthonormal matrices $V$ and $U$ and a diagonal matrix $\Sigma$ for which:

$$AV = U\Sigma.$$

Since $V$ is orthogonal, we can multiply both sides by $V^{-1} = V^T$ to get:

$$A = U\Sigma V^T.$$

Rather than solving for $U$, $V$ and $\Sigma$ simultaneously, we multiply both sides by $A^T = V\Sigma^T U^T$ to get:

$$\begin{aligned} A^TA &= V\Sigma U^{-1}U\Sigma V^T \\ &= V\Sigma^2 V^T \\ &= V \begin{bmatrix} \sigma_1^2 & & & \\ & \sigma_2^2 & & \\ & & \ddots & \\ & & & \sigma_n^2 \end{bmatrix} V^T. \end{aligned}$$

This is in the form $Q\Lambda Q^T$; we can now find $V$ by diagonalizing the symmetric positive definite (or semidefinite) matrix $A^TA$. The columns of $V$ are eigenvectors of $A^TA$ and the eigenvalues of $A^TA$ are the values $\sigma_i^2$. (We choose $\sigma_i$ to be the positive square root of $\lambda_i$.)

To find $U$, we do the same thing with $AA^T$.

## SVD example

We return to our matrix $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$. We start by computing

$$\begin{aligned} A^TA &= \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}. \end{aligned}$$

The eigenvectors of this matrix will give us the vectors $\mathbf{v}_i$, and the eigenvalues will gives us the numbers $\sigma_i$.

$\begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix} \to [$

3

$\begin{bmatrix} -1 \\ -1 \end{bmatrix}$ $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ } It seems that answer matches by these vectors.

Two orthogonal eigenvectors of $A^TA$ are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$. To get an orthonormal basis, let $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$. These have eigenvalues $\sigma_1^2 = 32$ and $\sigma_2^2 = 18$. We now have:

$$\begin{array}{cccc} A & U & \Sigma & V^T \\ \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = & \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} & \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix} & \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}. \end{array}$$

We could solve this for $U$, but for practice we'll find $U$ by finding orthonormal eigenvectors $\mathbf{u}_1$ and $\mathbf{u}_2$ for $AA^T = U\Sigma^2U^T$.

$$A = U\Sigma V^T$$
$$AA^T = U\Sigma V^T (U\Sigma V^T)^T$$
$$AA^T = U\Sigma V^T V \Sigma^T U^T$$
$$AA^T = U\Sigma\Sigma^T U^T$$
$$AA^T = U\Sigma^2 U^T$$

$$\begin{aligned} AA^T &= \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} \\ &= \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}. \end{aligned}$$

Luckily, $AA^T$ happens to be diagonal. It's tempting to let $\mathbf{u}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{u}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, as Professor Strang did in the lecture, but because $A\mathbf{v}_2 = \begin{bmatrix} 0 \\ -3\sqrt{2} \end{bmatrix}$ we instead have $\mathbf{u}_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$ and $U = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. Note that this also gives us a chance to double check our calculation of $\sigma_1$ and $\sigma_2$.

$$Av_2 = \sigma_2 U_2$$
$$AV_2 = 3\sqrt{2}\begin{bmatrix} 0 \\ 1 \end{bmatrix}$$
$$AV_2 = \begin{bmatrix} 0 \\ 3\sqrt{2} \end{bmatrix}$$

$$\begin{aligned} AV_2 &= \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \\ &= \tfrac{1}{\sqrt{2}}\begin{bmatrix} 4 \\ -3 \end{bmatrix} + -\tfrac{1}{\sqrt{2}}\begin{bmatrix} 4 \\ 3 \end{bmatrix} \\ &= \begin{bmatrix} 4/\sqrt{2} \\ -3/\sqrt{2} \end{bmatrix} + \begin{bmatrix} -4/\sqrt{2} \\ -3/\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -6/\sqrt{2} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ \frac{-6\sqrt{2}}{2} \end{bmatrix} \\ &= \begin{bmatrix} 0 \\ -3\sqrt{2} \end{bmatrix} \end{aligned}$$

Thus, the SVD of $A$ is:

$$\begin{array}{cccc} A & U & \Sigma & V^T \\ \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = & \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} & \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix} & \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}. \\ X \;= & U & S & V^T \end{array}$$

## Example with a nullspace

Now let $A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$. This has a one dimensional nullspace and one dimensional row and column spaces.

The row space of $A$ consists of the multiples of $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$. The column space of $A$ is made up of multiples of $\begin{bmatrix} 4 \\ 8 \end{bmatrix}$. The nullspace and left nullspace are perpendicular to the row and column spaces, respectively.

Unit basis vectors of the row and column spaces are $\mathbf{v}_1 = \begin{bmatrix} .8 \\ .6 \end{bmatrix}$ and $\mathbf{u}_1 =$

$$\hat{v}_1 = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix}$$
$$\hat{U}_1 = \begin{bmatrix} 4/\sqrt{80} \\ 8/\sqrt{80} \end{bmatrix}$$

4

$\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$. To compute $\sigma_1$ we find the nonzero eigenvalue of $A^TA$.

$$A^TA = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}.$$

Because this is a rank 1 matrix, one eigenvalue must be 0. The other must equal the trace, so $\sigma_1^2 = 125$. After finding unit vectors perpendicular to $\mathbf{u}_1$ and $\mathbf{v}_1$ (basis vectors for the left nullspace and nullspace, respectively) we see that the SVD of $A$ is:

$$\underbrace{\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}}_{A} = \underbrace{\frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix}}_{V^T}.$$

The singular value decomposition combines topics in linear algebra ranging from positive definite matrices to the four fundamental subspaces.

| | |
|---|---|
| $\mathbf{v}_1, \mathbf{v}_2, \ldots \mathbf{v}_r$ | is an orthonormal basis for the row space. |
| $\mathbf{u}_1, \mathbf{u}_2, \ldots \mathbf{u}_r$ | is an orthonormal basis for the column space. |
| $\mathbf{v}_{r+1}, \ldots \mathbf{v}_n$ | is an orthonormal basis for the nullspace. |
| $\mathbf{u}_{r+1}, \ldots \mathbf{u}_m$ | is an orthonormal basis for the left nullspace. |

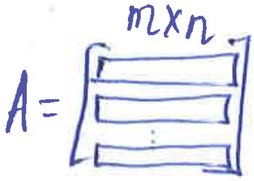
These are the "right" bases to use, because $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$.

$A v_1 = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} \begin{bmatrix} 0.8 \\ 0.6 \end{bmatrix}$ | $\sigma_1 u_1 = \frac{\sqrt{125}}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$A v_1 = \begin{bmatrix} 5 \\ 10 \end{bmatrix}$ | $\sigma_1 u_1 = \frac{5\sqrt{5}}{\sqrt{5}} \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$A v_1 = 5 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ | $\sigma_1 u_1 = 5 \begin{bmatrix} 1 \\ 2 \end{bmatrix}$

5




MIT OpenCourseWare
http://ocw.mit.edu

18.06SC Linear Algebra
Fall 2011

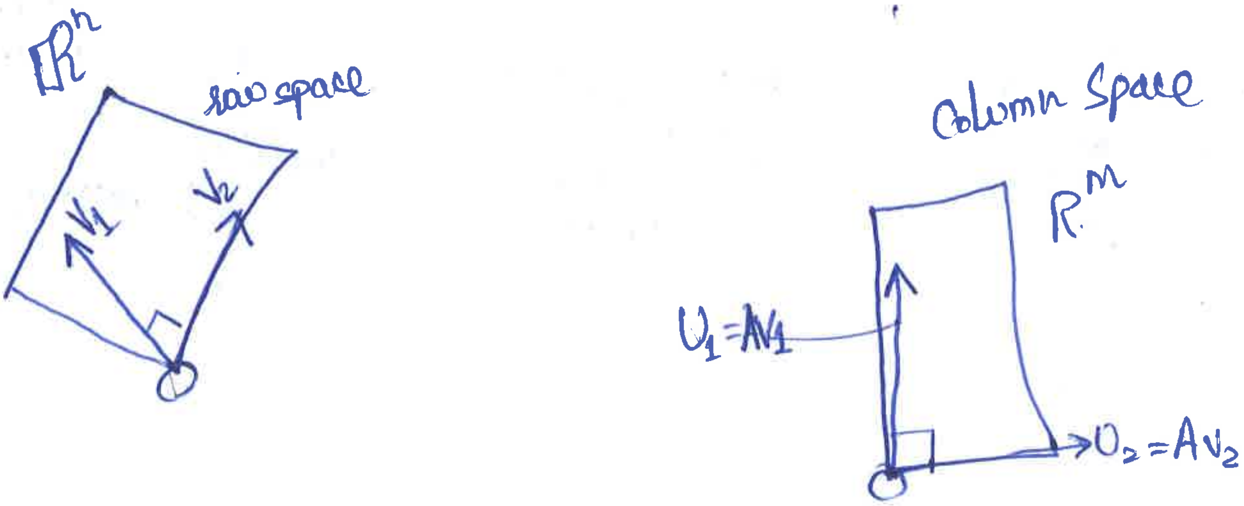
For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

LA lec 29

Singular Value Decomposition = SVD

$A = U\Sigma V^T$ / $\Sigma \rightarrow$ diagonal matrix
U,V are orthogonal

If A is Symmetric Pos definite then $A = Q\Lambda Q^T$ (Its SVD in case A is PD)
we dont want to look at $A = S\Lambda S^{-1}$




~~Spe~~

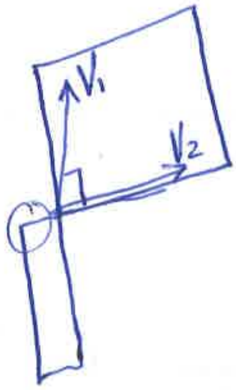
Special case SVD when A is PD: we will have $A = Q\Lambda Q^T$ i.e $U = V = Q$
b/c if A is PD, we can use same basis for row & column space.

Now in Above fig, we have vector $V_1$ in rowspace, we want to find $U_1$ in column space such that $AV_1 = U_1$ or we want to knock $V_1$ into column space.

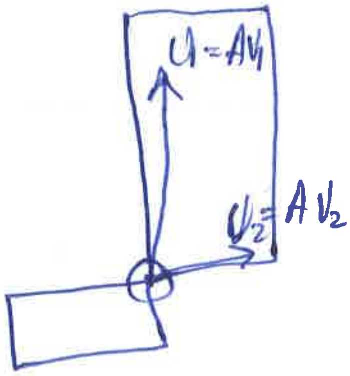
Goal of SVD: We want to find orthogonal basis in rowspace that gets knock over in orthogonal basis in columnspace.

Orthogonal basis in rowspace can be found by Gram-Schmidt
Now we can include nullspace as well.

$\mathbb{R}^n$ rowspace (Dimension of rowspace = r)



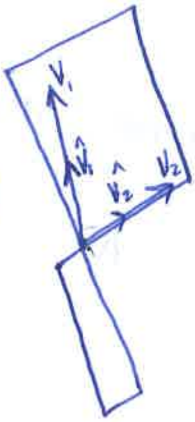

$\mathbb{R}^m$ columnspace




This means $A\begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix} =$

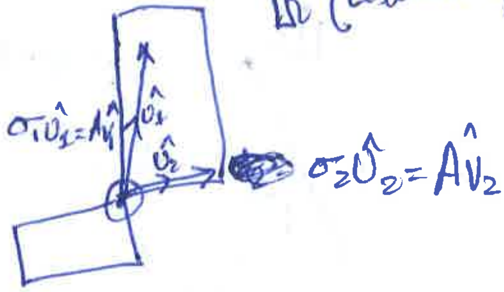
Now we want orthonormal vectors. So lets make vectors unit vectors.

$\mathbb{R}^n$ (rowspace)




let $u_1, u_2, v_1, v_2$ be unit vectors
thus $\sigma_1 u_1 = Av_1$
$\sigma_2 u_2 = Av_2$

$\mathbb{R}^m$ (column-space)




$$\Rightarrow A\underbrace{\begin{bmatrix} v_1 & v_2 & \cdots & v_r \end{bmatrix}}_{\text{Basis vectors in rowspace}} = \underbrace{\begin{bmatrix} u_1 & u_2 & \cdots & u_r \end{bmatrix}}_{\text{Basis vectors in column space}} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix}$$

Now $AV = U\Sigma$

Again our goal is to find orthogonal basis vector $V$ & $U$ such that $A$ is diagonalised in $\Sigma$. If $A$ is PD, then its easy b/c $AQ = Q\Sigma$ but now $A$ is any matrix.

If there is some Nullspace, then we want to stick them in our picture. If our rowspace is $r$ dimensional, then nullspace $= n-r$ dim

Sticking in orthonormal basis for null space of $A$.

$$\Rightarrow A\Big[ v_1 \; v_2 \ldots v_r \; \overbrace{\underbrace{v_{r+1} \; v_{r+2} \ldots v_n}_{\text{orthonormal basis for nullspace}}}^{(n-r)\text{ orthonormal basis}} \Big]$$

$$= \Big[ u_1 \; u_2 \ldots u_r \; u_{r+1} \; u_{r+2} \ldots u_m \Big] \begin{bmatrix} \sigma_1 & & & & \\ & \sigma_2 & & & \\ & & \ddots & & \\ & & & \sigma_r & \\ & & & & 0 \; \ddots \; 0 \end{bmatrix}$$

$n-r$ zeroes

Thus we will fill rowspace $R^n$ & column space $R^m$ with zeroes in $\Sigma$ matrix, which is diagonal.

For instance: $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$ — * $A$ is invertible * Rank = 2

we have to find orthonormal $v_1, v_2$ in rowspace $R^2$, $u_1$ & $u_2$ in column space $R^2$ & $\underbrace{\sigma_1 > 0 \quad \sigma_2 > 0}_{\text{scaling factor}}$

* $A$ is not symmetric, thus its eigenvalues are not orthogonal

So we want $\left.\begin{matrix} Av_1 = \sigma u_1 \\ Av_2 = \sigma u_2 \end{matrix}\right\}$ our goal

Now going back to $AV = U\Sigma$ — $V, U$ = orthogonal matrix (square orthogonal matrices)

$$\Rightarrow A = U\Sigma V^{-1}$$

$$\Rightarrow \boxed{A = U\Sigma V^T}$$

Now, we will eliminate $U$.

As $A^T = V\Sigma^T U^T$

Thus $A^TA = V\Sigma^T U^T U \Sigma V$

$= V\Sigma^T I \Sigma V$

$= V\underbrace{\Sigma^T\Sigma}_{\text{diagonal}} V$

$$A^TA = \underbrace{V}_{Q} \underbrace{\begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \end{bmatrix}}_{\lambda} \underbrace{V^T}_{Q^T}$$

eigenvalues of $A^TA$

$Q \quad \lambda \quad Q^T$ for $A^TA$

Here $V$ is matrix of eigenvectors of $A^TA$.

Now To get $U$, lets take $AA^T$:

$AA^T = U\Sigma V (U\Sigma V)^T$

$= U\Sigma V V^T \Sigma^T U^T$

$= U \Sigma\Sigma^T U^T$

$= U \begin{bmatrix} \sigma_1^2 & & \\ & \sigma_2^2 & \\ & & \ddots \end{bmatrix} U^T$

As $A^{-1}A = AA^{-1} = I$ If $A$ is orthogonal $A^T = A^{-1}$

$\Rightarrow V^TV = VV^T = I$

Here $U$ is is matrix of eigenvectors of $AA^T$ which is different from $A^TA$. But eigenvalues$(A^TA)$ = eigenvalues$(AA^T)$

Now $A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$

$\Rightarrow A^TA = \begin{bmatrix} 25 & 7 \\ 7 & 25 \end{bmatrix}$ Its eigenvectors will be $V$'s & eigenvalues will be $\sigma$'s.

Eigenvectors are $\begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \end{bmatrix}$ b/c $(A^TA)\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 32\begin{bmatrix} 1 \\ 1 \end{bmatrix} \xrightarrow{\text{normalise}} 32\begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$

$A^TA\begin{bmatrix} 1 \\ -1 \end{bmatrix} = 18\begin{bmatrix} 1 \\ -1 \end{bmatrix} \xrightarrow{\text{normalise}} 18\begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

Now: $\sigma_1^2 = 32$ $\quad \sigma_1 = \sqrt{32}$ $\qquad V = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$ $\qquad$ 9

$\sigma_2^2 = 18$ $\quad \sigma_2 = \sqrt{18}$

$$\begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \underbrace{\begin{bmatrix} & \\ & \end{bmatrix}}_{U??} \begin{bmatrix} \sqrt{32} & 0 \\ 0 & \sqrt{18} \end{bmatrix} \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}}_{V^T}$$

Finding Us using $AA^T$:

$$AA^T = U \Sigma \underbrace{V^T V}_{I} \Sigma^T U^T$$

$\cancel{AA^T}$.

symmetric Positive definite $AA^T = U \Sigma \Sigma^T U^T$ $\qquad$ eigenvalues of U will go in $\Sigma$

$$AA^T = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 4 & -3 \\ 4 & 3 \end{bmatrix} = \begin{bmatrix} 32 & 0 \\ 0 & 18 \end{bmatrix}$$

Eigenvectors $(AA^T)$: $\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ b/c $AA^T$ comes out to be diagonal

$AA^T \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 32 \begin{bmatrix} 1 \\ 0 \end{bmatrix}$

$AA^T \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 18 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

We are seeing eigenvalues of 32, 18 again. It's not surprising because $\boxed{\text{eigenvalues of } \underset{(A^TA)}{AB} = \text{eigenvalues of } \underset{(AA^T)}{BA}}$.

Now we can have $U = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ $\qquad u_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ $\quad u_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$

1st eig vector $\quad$ 2nd vector $\qquad v_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ $\quad v_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

$\sigma_1 = 4\sqrt{2}$

$\sigma_2 = 3\sqrt{2}$

As $Av_2 = \sigma_2 u_2$ $\qquad A = \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix}$

$\Rightarrow u_2 = \frac{1}{\sigma_2} A v_2$

$u_2 = \frac{1}{3\sqrt{2}} \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$

$u_2 = \begin{bmatrix} 0 \\ -1 \end{bmatrix}$

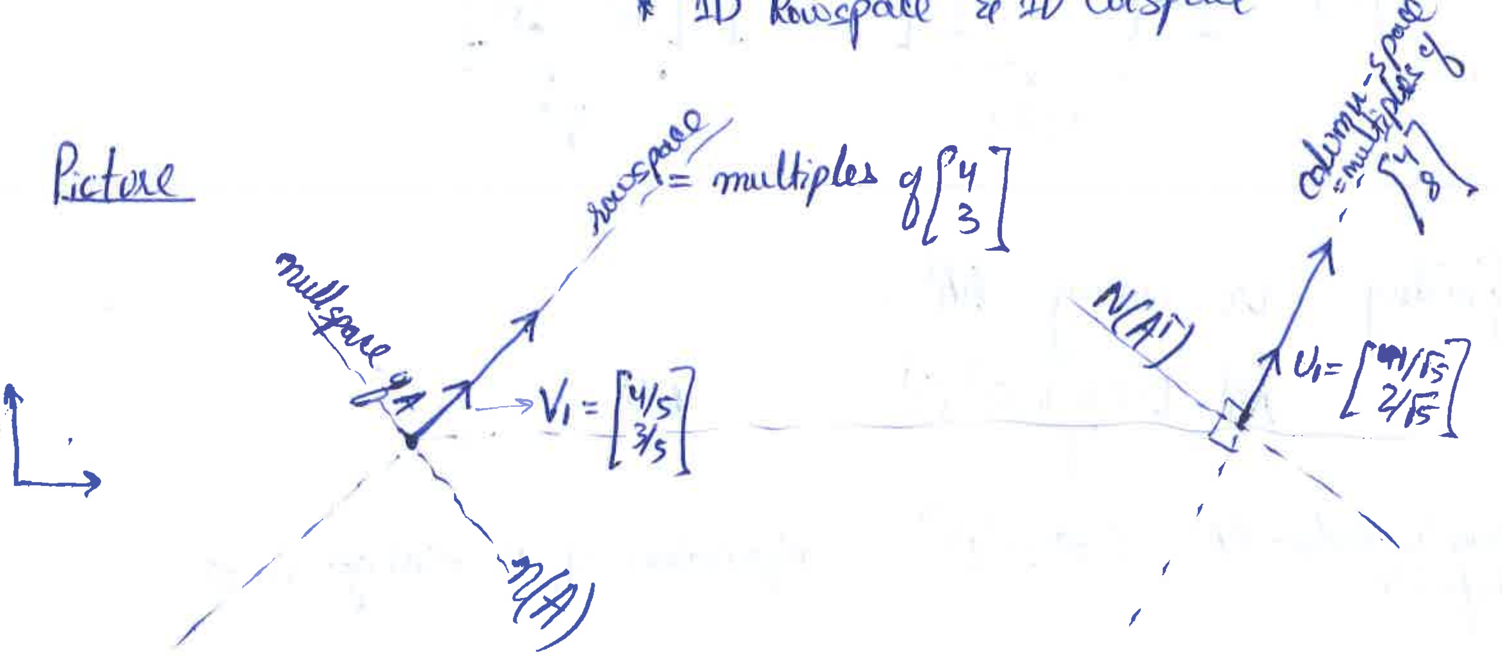
$$\Rightarrow \begin{bmatrix} 4 & 4 \\ -3 & 3 \end{bmatrix} = \begin{array}{cc} u_1 & u_2 \end{array}\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 4\sqrt{2} & 0 \\ 0 & 3\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Another Example:

$$A = \begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}$$

* Rank-1 matrix
* It has nullspace
* 1D Rowspace & 1D Colspace

Picture




Choosing orthogonal basis for above A is not a problem since have 1D $C(A^T)$ & 1D $C(A)$. Thus for col space, we can choose $\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 4\begin{bmatrix} 1 \\ 2 \end{bmatrix}$ as its basis. Now we will make it unit basis i.e $U_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$

$$\Rightarrow \underset{A}{\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}} = \underset{U}{\begin{bmatrix} 1/\sqrt{5} & \\ 2/\sqrt{5} & \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} \sigma_1 & 0 \\ 0 & 0 \end{bmatrix}} \overset{V^T}{\begin{bmatrix} & \\ & \end{bmatrix}}$$

$\boxed{\sigma_2 = 0 \text{ b/c } A \text{ is rank-1}}$

$A^TA = \begin{bmatrix} 4 & 8 \\ 3 & 6 \end{bmatrix}\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \begin{bmatrix} 80 & 60 \\ 60 & 45 \end{bmatrix}$

As $A$ = rank-1
Then $A^TA$ = rank-1
Thus we can say
$\sigma_1 = 0$ (as rank-1)
$\sigma_2$ = Trace = 80 + 45
= 125

$$= \begin{bmatrix} 1/\sqrt{5} & \\ 2/\sqrt{5} & \end{bmatrix} \begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} & \\ & \end{bmatrix}$$

As $V_1 = \begin{bmatrix} 4/5 \\ 3/5 \end{bmatrix} = \begin{bmatrix} .8 \\ .6 \end{bmatrix}$ & $V_2 \perp V_1$ (so to find orthonormal basis, we know its in N(A) & $\langle V_1, V_2 \rangle = 0$)

$$\Rightarrow V_2 = \begin{bmatrix} 3/5 \\ -4/5 \end{bmatrix} = \begin{bmatrix} .6 \\ -.8 \end{bmatrix}$$

So $A = U \Sigma V^T$

$$\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix} = \overset{U}{\begin{bmatrix} 1/\sqrt{5} & \cdot \\ 2/\sqrt{5} & \end{bmatrix}} \overset{\Sigma}{\begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix}} \overset{V^T}{\begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix}}$$

As $U_1 = \begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$. To get $U_2$ so that $U_2 \perp U_1$ (orthogonal basis for full column space with null space will be calculated in the same way we calculated $V_1$ & $V_2$.)

$$\boxed{\begin{aligned} \text{full column-space} &= \text{column space} + \text{nullspace} \\ \text{full row-space} &= \text{row space} + \text{nullspace} \end{aligned}}$$

$\Rightarrow U_2 = \begin{bmatrix} 2/\sqrt{5} \\ -1/\sqrt{5} \end{bmatrix}$

$$\Rightarrow \overset{A}{\begin{bmatrix} 4 & 3 \\ 8 & 6 \end{bmatrix}} = \frac{1}{\sqrt{5}} \overset{U}{\begin{bmatrix} 1 & 2 \\ 2 & -1 \end{bmatrix}} \overset{\Sigma}{\begin{bmatrix} \sqrt{125} & 0 \\ 0 & 0 \end{bmatrix}} \overset{V^T}{\begin{bmatrix} .8 & .6 \\ .6 & -.8 \end{bmatrix}}$$

are multiplied by 0s.

What we were doing? We were choosing right basis for four fundamental subspaces.

$V_1, \dots, V_r$ orthonormal basis for row space $\Big\}$ Rank $= r$

$U_1, \dots, U_r$ " " " column space

$V_{r+1}, \dots, V_n$ " " " nullspace $\}$ Dim $N(A) = n-r$

$U_{r+1}, \dots, U_m$ " " " $N(A^T)$ $\}$ Dim $N(A^T) = m-r$ = dimension of left nullspace

This means if we select these basis then
$$A V_i = \sigma_i U_i$$
there is no coupling between $V_i$ & $U_i$ & if we multiply $V_i$ by $A$, it will be in direction of $U_i$.

Recitation??

Unit III : Lecture 6 ①

"When we multiply a matrix by an input vector we get an output vector, often in a new space. We can ask what this "linear transformation" does to all vectors in a space. In fact, matrices were originally invented for the study of linear transformation."

# Linear transformations and their matrices

In older linear algebra courses, linear transformations were introduced before matrices. This geometric approach to linear algebra initially avoids the need for coordinates. But eventually there must be coordinates and matrices when the need for computation arises.

## Without coordinates (no matrix)

### Example 1: Projection

We can describe a projection as a *linear transformation* $T$ which takes every vector in $\mathbb{R}^2$ into another vector in $\mathbb{R}^2$. In other words,

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2.$$

The rule for this *mapping* is that every vector $\mathbf{v}$ is projected onto a vector $T(\mathbf{v})$ on the line of the projection. Projection is a linear transformation.

### Definition of *linear*

A transformation $T$ is *linear* if:

$$T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$$

and

$$T(c\mathbf{v}) = cT(\mathbf{v})$$

for all vectors $\mathbf{v}$ and $\mathbf{w}$ and for all scalars $c$. Equivalently,

$$T(c\mathbf{v} + d\mathbf{w}) = cT(\mathbf{v}) + dT(\mathbf{w})$$

for all vectors $\mathbf{v}$ and $\mathbf{w}$ and scalars $c$ and $d$. It's worth noticing that $T(\mathbf{0}) = \mathbf{0}$, because if not it couldn't be true that $T(c\mathbf{0}) = cT(\mathbf{0})$.

### Non-example 1: Shift the whole plane

Consider the transformation $T(\mathbf{v}) = \mathbf{v} + \mathbf{v}_0$ that shifts every vector in the plane by adding some fixed vector $\mathbf{v}_0$ to it. This is *not* a linear transformation because $T(2\mathbf{v}) = 2\mathbf{v} + \mathbf{v}_0 \neq 2T(\mathbf{v})$.

### Non-example 2: $T(\mathbf{v}) = ||\mathbf{v}||$

The transformation $T(\mathbf{v}) = ||\mathbf{v}||$ that takes any vector to its length is not a linear transformation because $T(c\mathbf{v}) \neq cT(\mathbf{v})$ if $c < 0$.

We're not going to study transformations that aren't linear. From here on, we'll only use $T$ to stand for linear transformations.

②

**Example 2: Rotation by 45°**

This transformation $T : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ takes an input vector $\mathbf{v}$ and outputs the vector $T(\mathbf{v})$ that comes from rotating $\mathbf{v}$ counterclockwise by 45° about the origin. Note that we can describe this and see that it's linear without using any coordinates.

**The big picture**

One advantage of describing transformations geometrically is that it helps us to see the big picture, as opposed to focusing on the transformation's effect on a single point. We can quickly see how rotation by 45° will transform a picture of a house in the plane. If the transformation was described in terms of a matrix rather than as a rotation, it would be harder to guess what the house would be mapped to.

Frequently, the best way to understand a linear transformation is to find the matrix that lies behind the transformation. To do this, we have to choose a basis and bring in coordinates.

## With coordinates (matrix!)

All of the linear transformations we've discussed above can be described in terms of matrices. In a sense, linear transformations are an abstract description of multiplication by a matrix, as in the following example.

**Example 3:** $T(\mathbf{v}) = A\mathbf{v}$

Given a matrix $A$, define $T(\mathbf{v}) = A\mathbf{v}$. This is a linear transformation:

$$A(\mathbf{v}+\mathbf{w}) = A(\mathbf{v}) + A(\mathbf{w})$$

and

$$A(c\mathbf{v}) = cA(\mathbf{v}).$$

**Example 4**

Suppose $A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. How would we describe the transformation $T(\mathbf{v}) = A\mathbf{v}$ geometrically?

When we multiply $A$ by a vector $\mathbf{v}$ in $\mathbb{R}^2$, the $x$ component of the vector is unchanged and the sign of the $y$ component of the vector is reversed. The transformation $\mathbf{v} \mapsto A\mathbf{v}$ reflects the $xy$-plane across the $x$ axis.

**Example 5**

How could we find a linear transformation $T : \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ that takes three dimensional space to two dimensional space? Choose any 2 by 3 matrix $A$ and define $T(\mathbf{v}) = A\mathbf{v}$.

②

### Describing $T(\mathbf{v})$

How much information do we need about $T$ to to determine $T(\mathbf{v})$ for all $\mathbf{v}$? If we know how $T$ transforms a single vector $\mathbf{v}_1$, we can use the fact that $T$ is a linear transformation to calculate $T(c\mathbf{v}_1)$ for any scalar $c$. If we know $T(\mathbf{v}_1)$ and $T(\mathbf{v}_2)$ for two independent vectors $\mathbf{v}_1$ and $\mathbf{v}_2$, we can predict how $T$ will transform any vector $c\mathbf{v}_1 + d\mathbf{v}_2$ in the plane spanned by $\mathbf{v}_1$ and $\mathbf{v}_2$. If we wish to know $T(\mathbf{v})$ for all vectors $\mathbf{v}$ in $\mathbb{R}^n$, we just need to know $T(\mathbf{v}_1), T(\mathbf{v}_2), ..., T(\mathbf{v}_n)$ for any basis $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ of the input space. This is because any $\mathbf{v}$ in the input space can be written as a linear combination of basis vectors, and we know that $T$ is linear:

$$\begin{aligned} \mathbf{v} &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_n\mathbf{v}_n \\ T(\mathbf{v}) &= c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_nT(\mathbf{v}_n). \end{aligned}$$

This is how we get from a (coordinate-free) linear transformation to a (coordinate based) matrix; the $c_i$ are our coordinates. Once we've chosen a basis, every vector $\mathbf{v}$ in the space can be written as a combination of basis vectors in exactly one way. The coefficients of those vectors are the *coordinates* of $\mathbf{v}$ in that basis.

Coordinates come from a basis; changing the basis changes the coordinates of vectors in the space. We may not use the standard basis all the time – we sometimes want to use a basis of eigenvectors or some other basis.

### The matrix of a linear transformation

Given a linear transformation $T$, how do we construct a matrix $A$ that represents it?

First, we have to choose two bases, say $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n$ of $\mathbb{R}^n$ to give coordinates to the input vectors and $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m$ of $\mathbb{R}^m$ to give coordinates to the output vectors. We want to find a matrix $A$ so that $T(\mathbf{v}) = A\mathbf{v}$, where $\mathbf{v}$ and $A\mathbf{v}$ get their coordinates from these bases.

The first column of $A$ consists of the coefficients $a_{11}, a_{21}, ..., a_{1m}$ of $T(\mathbf{v}_1) = a_{11}\mathbf{w}_1 + a_{21}\mathbf{w}_2 + \cdots + a_{1m}\mathbf{w}_m$. The entries of column $i$ of the matrix $A$ are determined by $T(\mathbf{v}_i) = a_{1i}\mathbf{w}_1 + a_{2i}\mathbf{w}_2 + \cdots + a_{1i}\mathbf{w}_m$. Because we've guaranteed that $T(\mathbf{v}_i) = A\mathbf{v}_i$ for each basis vector $\mathbf{v}_i$ and because $T$ is linear, we know that $T(\mathbf{v}) = A\mathbf{v}$ for all vectors $\mathbf{v}$ in the input space.

In the example of the projection matrix, $n = m = 2$. The transformation $T$ projects every vector in the plane onto a line. In this example, it makes sense to use the same basis for the input and the output. To make our calculations as simple as possible, we'll choose $\mathbf{v}_1$ to be a unit vector on the line of projection and $\mathbf{v}_2$ to be a unit vector perpendicular to $\mathbf{v}_1$. Then

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) = c_1\mathbf{v}_1 + \mathbf{0}$$

and the matrix of the projection transformation is just $A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

$$A\mathbf{v} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} c_1 \\ 0 \end{bmatrix}.$$

This is a nice matrix! If our chosen basis consists of eigenvectors then the matrix of the transformation will be the diagonal matrix $\Lambda$ with eigenvalues on the diagonal.

To see how important the choice of basis is, let's use the standard basis for the linear transformation that projects the plane onto a line at a 45° angle. If we choose $\mathbf{v}_1 = \mathbf{w}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\mathbf{v}_2 = \mathbf{w}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$, we get the projection matrix $P = \frac{\mathbf{a}\mathbf{a}^T}{\mathbf{a}^T\mathbf{a}} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$. We can check by graphing that this is the correct matrix, but calculating $P$ directly is more difficult for this basis than it was with a basis of eigenvectors.

**Example 6:** $T = \dfrac{d}{dx}$

Let $T$ be a transformation that takes the derivative:

$$T(c_1 + c_2x + c_3x^2) = c_2 + 2c_3x. \tag{1}$$

The input space is the three dimensional space of quadratic polynomials $c_1 + c_2x + c_3x^2$ with basis $\mathbf{v}_1 = 1$, $\mathbf{v}_2 = x$ and $\mathbf{v}_3 = x^2$. The output space is a two dimensional subspace of the input space with basis $\mathbf{w}_1 = \mathbf{v}_1 = 1$ and $\mathbf{w}_2 = \mathbf{v}_2 = x$.

This is a linear transformation! So we can find $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$ and write the transformation (1) as a matrix multiplication (2):

$$T\left(\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}\right) = A \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} c_2 \\ 2c_3 \end{bmatrix}. \tag{2}$$

## Conclusion

For any linear transformation $T$ we can find a matrix $A$ so that $T(\mathbf{v}) = A\mathbf{v}$. If the transformation is invertible, the inverse transformation has the matrix $A^{-1}$. The product of two transformations $T_1 : \mathbf{v} \mapsto A_1\mathbf{v}$ and $T_2 : \mathbf{w} \mapsto A_2\mathbf{w}$ corresponds to the product $A_2A_1$ of their matrices. This is where matrix multiplication came from!

MIT OpenCourseWare
http://ocw.mit.edu

18.06SC Linear Algebra
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

LA
Lecture 30:

Linear Transformation $T$
without coordinates: no matrices
with coordinates $\Rightarrow$ matrix

Rules for linear transformation

i) $T(v+w) = T(v) + T(w)$
ii) $T(cv) = cT(v)$

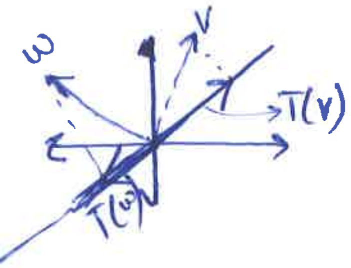
Also $T(0) = 0$ in linear transformation

$T(cv + dw) = cT(v) + dT(w)$

Example 1: Projection

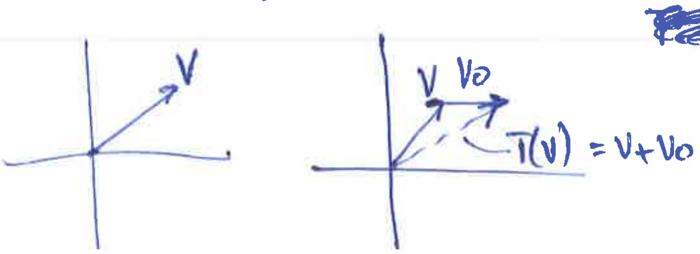
$T: \mathbb{R}^2 \to \mathbb{R}^2$ (mapping)

Projection is a linear transformation.




v is transformed into T(V). It's like a function (No Axis here)

Example 2 Shift the whole plane by $V_0$



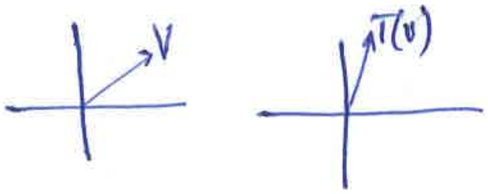

It's not a linear transformation b/c if $v = 2v$ then $T(2v) \neq 2T(v)$

Non-Example $T(v) = \|v\|$ $T: \mathbb{R}^3 \to \mathbb{R}$

Property (ii) is violated hence it is not linear transformation.

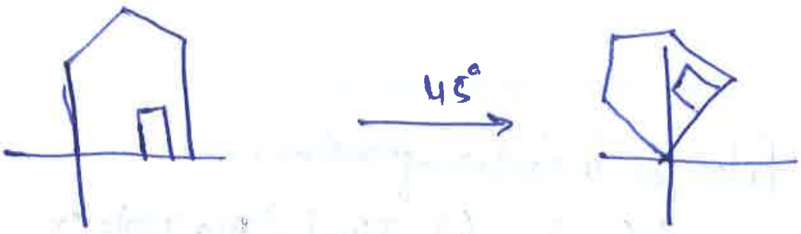
Example 2 Rotation by $45°$

$T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$




* It is linear transformation
* We don't need axes here.

OR




The idea of linear transformation is to see the effect of transformation all together.

Example 3 !! Matrix A

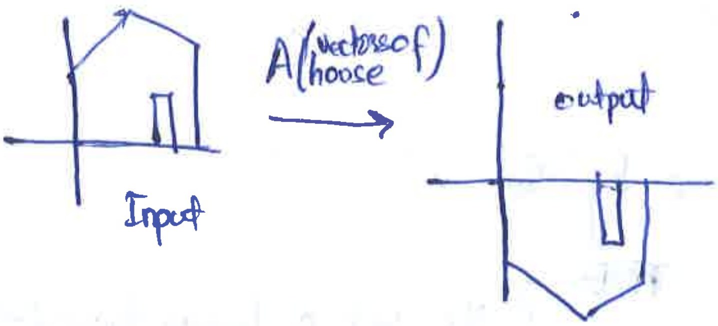
$T(v) = Av$

Its linear b/c:

i) $A(v+w) = Av + Aw$

ii) $A(cv) = cA(v)$

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



Start: $T$ (linear transformation)

$T: R^3 \longrightarrow R^2$ (Any matrix of right size will do it)

Examples: $T(V) = A\vec{V}$ ← $\boxed{2\times 3 \text{ matrix}}$

output in $R^2$    input in $R3$

we need coordinate system now.

<u>Information needed to know $T(v)$ for all inputs.</u>

$T(V_1), T(V_2), \ldots, T(V_n)$ for any input basis $V_1, \ldots V_n$

This is all we need to know. because

Every $V = C_1V_1 + \cdots + C_nV_n$ (b/c basis span the space)

we know $T(V) = C_1T(V_1) + \cdots + C_nT(V_n)$ } Because of linearity

Here $C_1, \ldots C_n$ are coordinates that comes from ~~of~~ basis $V_1, \ldots V_n$

<u>Coordinates</u> come from a basis

coordinates of $V = C_1V_1 + \cdots + C_nV_n$ are the amount $(C_1 \cdots C_n)$ of basis vectors.

<u>Example</u>

If $V = \begin{bmatrix}3\\2\\4\end{bmatrix}$ then in standard basis $\begin{bmatrix}1\\0\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\end{bmatrix}$

$$\begin{bmatrix}3\\2\\4\end{bmatrix} = 3\begin{bmatrix}1\\0\\0\end{bmatrix} + 2\begin{bmatrix}0\\1\\0\end{bmatrix} + 4\begin{bmatrix}0\\0\\1\end{bmatrix}$$

This $\begin{bmatrix}3\\2\\4\end{bmatrix}$ is located in standard basis. If we change the basis, for instance we may select eigenvectors of a matrix as basis, then these co-ordinate $C_1, \ldots, C_n$ will change ~~acc~~ accordingly.

8

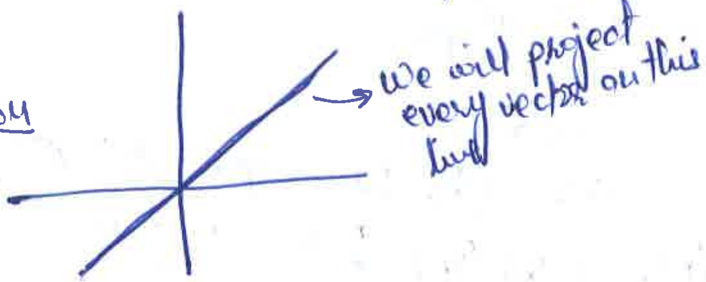
Construct matrix $A$ that represents linear transformation $T$.
$T: \mathbb{R}^n \to \mathbb{R}^m$

i) Choose a basis $V_1, \ldots, V_n$ for inputs $\mathbb{R}^n$ $T: \mathbb{R}^n \to \mathbb{R}^m$
ii) Choose " " $W_1, \ldots, W_m$ " output $\mathbb{R}^m$

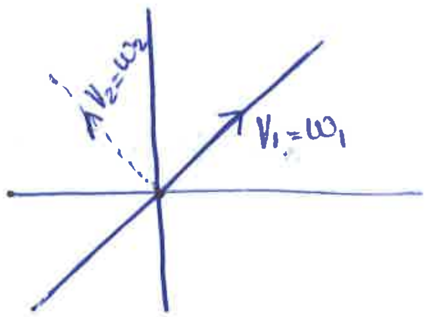
<u>We want</u>: A matrix "A" that does $T: \mathbb{R}^n \to \mathbb{R}^m$

Lets take an example of projection

<u>Projection</u>




Lets choose basis (a different basis than standard one)
$V_1, V_2 \to$ input basis
$W_1, W_2 \to$ output basis




Any $V$ can be described as $V = C_1 V_1 + C_2 V_2$
Thus $T(V) = C_1 V_1$

We know what projection $T(V)$ does to $\vec{V}$. It saves 1st basis $V_1$ while it kills component of $\vec{V}$ in the direction of $\vec{V_2}$.

thus $T(V) = C_1 V_1$
Input $= C_1, C_2$ (co-ordinates of input)
output $= C_1$ (coordinates of output)

$$\Rightarrow A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

$$\underset{A}{\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}} \underset{\text{input coords}}{\begin{bmatrix} C_1 \\ C_2 \end{bmatrix}} = \underset{\text{Output coords}}{\begin{bmatrix} C_1 \\ 0 \end{bmatrix}}$$

we choose same input basis & out basis ~~so~~ that lie along perpendicular lines
They are the eigen values of projection matrix.

Eigen Vector basis (good basis) : It leads to diagonal matrix $\Lambda$.

Say Proj onto 45° line & use standard basis $V_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \omega_1, V_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \omega_2$

Then Projection matrix will be :

$$P = \frac{aa^T}{a^Ta} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix}$$

$$PV_1 = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix} \qquad PV_2 = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

If we had choosen good basis, then we can see that Projection matrix is diagonal, & we can easily see that $A^2 = A$

Rule to find A : Given basis $V_1 \rightarrow V_n$, $\omega_1 \rightarrow \omega_m$

1st column of A : write $T(V_1) = a_{11}\omega_1 + a_{21}\omega_2 + \cdots + a_{m1}\omega_m$

2nd column of A write $T(V_2) = a_{12}\omega_1 + \cdots + a_{m2}\omega_m$

Outcome $A\begin{pmatrix} \text{input} \\ \text{co-ords} \end{pmatrix} = \begin{pmatrix} \text{output} \\ \text{co-ords} \end{pmatrix}$

Another linear transformation :

$T = \frac{d}{dx}$ Input : $c_1 + c_2x + c_3x^2$ basis : $1, x, x^2$

Output derivative : $c_2 + 2c_3x$ basis : $1, x$

linear (derivatives are linear)

$$\underset{2\times3}{A} \underset{3\times1}{\begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix}} = \underset{2\times1}{\begin{bmatrix} c_2 \\ 2c_3 \end{bmatrix}}$$

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix}$$

* ~~Matrix multiplication is linear operation~~ If $T(V) = AV$ & if $T(V)$ is invertible then $V = A^{-1}T(V)$. If $T_1 : V \rightarrow A_1V$, $T_2 : \omega \rightarrow A_2\omega$ And ~~then $T(V\omega) = T(A_1A_2)$~~

See last para

Video ~~data~~ camera record video data in a poor format for broadcasting video. To transmit video efficiently, linear algebra is used to change the basis. But which basis is best for video compression is an important question that has not been fully answered.

# Change of basis; image compression

We've learned that computations can be made easier by an appropriate choice of basis. One application of this principle is to image compression. Lecture videos, music, and other data sources contain a lot of information; that information can be efficiently stored and transmitted only after we change the basis used to record it.

## Compression of images

Suppose one frame of our lecture video is 512 by 512 pixels and that the video is recorded in black and white. The camera records a brightness level for each of the $(512)^2$ pixels; in this sense, each frame of video is a vector in a $(512)^2$ dimensional vector space.

The standard basis for this space has a vector for each pixel. Transmitting the values of all $(512)^2$ components of each frame using the standard basis would require far too much bandwidth, but if we change our basis according to the JPEG image compression standard we can transmit a fairly good copy the video very efficiently.

For example, if we're reporting light levels pixel by pixel, there's no efficient way to transmit the information "the entire frame is black". However, if one of our basis vectors corresponds to all pixels having the same light level (say 1), we can very efficiently transmit a recording of a blank blackboard.

Along with a vector of all 1's, we might choose a basis vector that alternates 1's and $-1$'s, or one that's half 1's and half $-1$'s corresponding to an image that's bright on the left and dark on the right. Our choice of basis will directly affect how much data we need to download to watch a video, and the best choice of basis for algebra lectures might differ from the best choice for action movies!

### Fourier basis vectors

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \omega^3 \\ \omega^4 \\ \omega^5 \\ \omega^6 \\ \omega^7 \end{bmatrix}, \begin{bmatrix} 1 \\ \omega^2 \\ \omega^4 \\ \omega^6 \\ \omega^8 \\ \omega^{10} \\ \omega^{12} \\ \omega^{14} \end{bmatrix}, \ldots, \begin{bmatrix} 1 \\ \omega^7 \\ \omega^{14} \\ \omega^{21} \\ \omega^{28} \\ \omega^{35} \\ \omega^{42} \\ \omega^{49} \end{bmatrix}$$

The best known basis is the Fourier basis, which is closely related to the Fourier matrices we studied earlier. The basis used by JPEG is made up of cosines – the real parts of $\omega^{jk}$.

This method breaks the 512 by 512 rectangle of pixels into blocks that are 8 pixels on a side, each block containing 64 pixels total. The brightness information for those pixels is then compressed, possibly by eliminating all coefficients below some threshold chosen so that we can hardly see the difference once they're gone.

$$\text{signal } \mathbf{x} \xrightarrow{\text{lossless}} 64 \text{ coefficients } c \xrightarrow{\text{lossy compression}} \hat{c} \text{ (many zeros)} \longrightarrow \hat{\mathbf{x}} = \sum \hat{c}_i \mathbf{v}_i$$

In video, not only should we consider compressing each frame, we can also consider compressing sequences of frames. There's very little difference between one frame and the next. If we do it right, we only need to encode and compress the differences between frames, not every frame in its entirety.

### The Haar wavelet basis

$$\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}$$

The closest competitor to the JPEG encoding method uses a wavelet basis. (JPEG2000 improves on the Haar wavelets above.) In Haar's wavelet basis for $\mathbb{R}^8$, the non-zero entries are half 1's and half $-1$'s (except for the vector of all 1's). However, half or even three quarters of a basis vector's entries may be 0. These vectors are chosen to be orthogonal and can be adjusted to be orthonormal.

### Compression and matrices

Linear algebra is used to find the coefficients $c_i$ in the change of basis from the standard basis (light levels for each pixel) to the Fourier or wavelet basis. For example, we might want to write:

$$\mathbf{x} = c_1\mathbf{w}_1 + \cdots + c_8\mathbf{w}_8.$$

But this is just a linear combination of the wavelet basis vectors. If $W$ is the matrix whose columns are the wavelet vectors, then our task is simply to solve for $\mathbf{c}$:

$$\mathbf{x} = W \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_8 \end{bmatrix}.$$

So $\mathbf{c} = W^{-1}\mathbf{x}$.

Our calculations will be faster and easier if we don't have to spend a lot of time inverting a matrix (e.g. if $W^{-1} = W^T$) or multiplying by the inverse. So in the field of image compression, the criteria for a good basis are:

3

- Multiplication by the basis matrix and its inverse is fast (as in the FFT or in the wavelet basis).
- Good compression – the image can be approximated using only a few basis vectors. Most components $c_i$ are small – safely set to zero.

## Change of basis

### Vectors

Let the columns of matrix $W$ be the basis vectors of the new basis. Then if $\mathbf{x}$ is a vector in the old basis, we can convert it to a vector $\mathbf{c}$ in the new basis using the relation:

$$\mathbf{x} = W\mathbf{c}.$$

### Transformation matrices

Suppose we have a linear transformation $T$. If $T$ has the matrix $A$ when working with the basis $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_8$ and $T$ has the matrix $B$ when working with the basis $\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_8$, it turns out that $A$ and $B$ must be similar matrices. In other words, $B = M^{-1}AM$ for some change of basis matrix $M$.

Reminder: If we have a basis $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_8$ and we know $T(\mathbf{v}_i)$ for each $i$, then we can use the fact that $T$ is a linear transformation to find $T(\mathbf{v}) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + \cdots + c_8T(\mathbf{v}_8)$ for any vector $\mathbf{v} = c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + \cdots + c_8\mathbf{v}_8$ in the space. The entries of column $i$ of the matrix $A$ are the coefficients of the output vector $T(\mathbf{v}_i)$.

If our basis consists of eigenvectors of our transformation, i.e. if $T(\mathbf{v}_i) = \lambda_i\mathbf{v}_i$, then $A = \Lambda$, the (diagonal) matrix of eigenvalues. It would be wonderful to use a basis of eigenvectors for image processing, but finding such a basis requires far more computation than simply using a Fourier or wavelet basis.

### Summary

When we change bases, the coefficients of our vectors change according to the rule $\mathbf{x} = W\mathbf{c}$. Matrix entries change according to a rule $B = M^{-1}AM$.

MIT OpenCourseWare
http://ocw.mit.edu

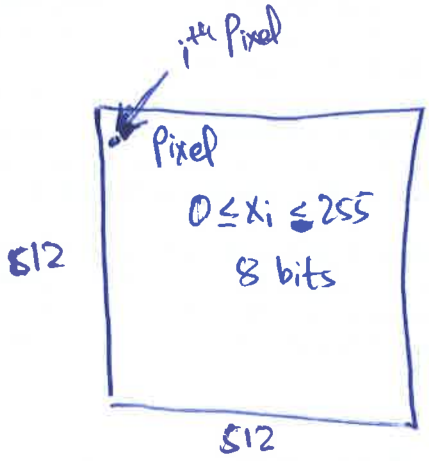
18.06SC Linear Algebra
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.



L#31

Change of basis
Compression of Images
Transformation $\leftrightarrow$ Matrix




$x \in \mathbb{R}^n$
$n = (512)^2$

$\underbrace{\text{STANDARD Compression : JPEG}}_{\text{It's all change of basis}}$

$\text{Image} = x = \begin{bmatrix} \updownarrow (512)^2 \end{bmatrix}$ values

Standard basis:

$$\underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \cdots \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{(512)^2}$$

Better basis.

$$\underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{\text{solid vector}} \quad \begin{bmatrix} 1 \\ \vdots \\ 1 \\ -1 \\ \vdots \\ -1 \end{bmatrix} \quad \underbrace{\begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ \vdots \end{bmatrix}}_{\text{checkerboard vector}}$$

$\boxed{\text{What basis to choose ??}}$

Fourier basis: (8x8) is a good choice

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \begin{bmatrix} 1 \\ \omega \\ \omega^2 \\ \vdots \\ \omega^{n-1} \end{bmatrix}$$

lossless → signal $x$ ($\mathbb{R}^{64}$)

↓ change basis

Coeffs $c$ (64 coeffs)

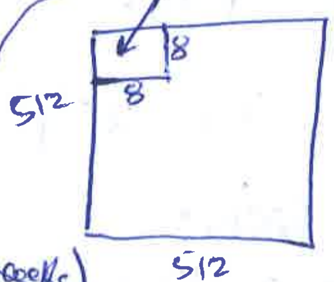
lossy → ↓ compression (1. throw small coeffs 2.)

$\hat{c}$ (many zeros)

↓

$$\hat{x} = \sum \hat{c}_i v_i$$

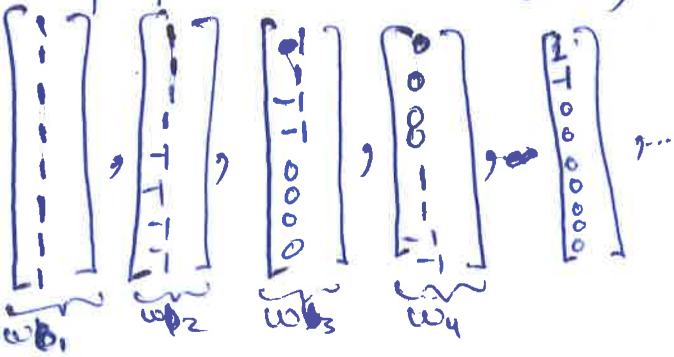
64 coefficients = 64 pixels




**Video**

Sequence of images are highly correlated & the problem in compression is to use this correlation

Competition for fourier: Wavelets ($\mathbb{R}^8$):




* We want to find $\hat{c}$ given basis & Pixels

$$P = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_8 \end{bmatrix} \text{(In standard basis)}$$

We want $P = c_1 w_1 + c_2 w_2 + \dots + c_8 w_8$

(we want to change from standard basis to wavelet basis)

$$P = \begin{bmatrix} w_1 & w_2 & w_3 & \cdots & w_8 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_8 \end{bmatrix}$$

$$\boxed{P = W c}$$



5

Solve $P = Wc$

$$\Rightarrow c = W^{-1} P$$

Good BASIS:

① FAST [FAST Multiplication] FFT, Fast wavelet transform FWT.

$W^{-1}$ should be easy to compute (If basis are orthogonal, it will be great Because then $W^{-1} = W^{T}$)

② Few basis vectors should be enough to reproduce image.

Change of basis:

Columns of $W$ = new basis vectors

$$[x]_{\text{old basis}} \longrightarrow [c]_{\text{new basis}} \qquad x = Wc$$

$T$ with respect to $v_1, \dots, v_8$ it has matrix $A$
with respect to $w_1, \dots, w_8$ it has matrix $B$

$\Bigg\}$ $T$ is transformation which is settled (linear transformation)

This means $A$ & $B$ are similar.

$$\Rightarrow B = M^{-1} A M$$

$M$ = change of basis matrix

What is $A$? Using a basis $V_1, \dots, V_8$

~~Fom~~ I know $T$ completely from $T(V_1), T(V_2), \dots, T(V_8)$ b/c

$T$ is linear transformation OR because

every $X = c_1 V_1 + c_2 V_2 + \dots + c_8 V_8$

Then $T(X) = c_1 T(V_1) + c_2 T(V_2) + \dots + c_8 T(V_8)$

Write $T(V_1) = a_{11} V_1 + a_{21} V_2 + \dots + a_{81} V_8$

$T(V_2) = a_{12} V_2 + a_{22} V_2 + \dots + a_{82} V_8$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \cdots \\ \vdots & & \vdots \\ a_{81} & & \cdots \end{bmatrix}$$

$$\begin{array}{c} \begin{array}{ccc} V_1 & V_2 \cdots & V_8 \end{array} \\ [A] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{18} \\ a_{21} & a_{22} & \cdots & a_{28} \\ a_{31} & a_{32} & \cdots & a_{38} \\ a_{41} & a_{42} & \cdots & a_{48} \\ a_{51} & a_{52} & \cdots & a_{58} \\ a_{81} & a_{812} & \cdots & a_{88} \end{bmatrix} \end{array}$$

Suppose we have an Eigenvector basis:

$T(V_i) = \lambda_i V_i$

What is $A$?

$$A = \underbrace{\begin{bmatrix} \lambda_1 & 0 & 0 & 0 \\ 0 & \lambda_2 & 0 \cdots & 0 \\ 0 & \vdots & \lambda_3 & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & & & \lambda_8 \end{bmatrix}}_{\text{Diagonal basis}}$$

1st input is $V_1$

Its output is $\lambda_1 V_1$:

2nd input is $V_2$:

Output is $\lambda_2 V_2$

We'd like to be able to "invert A" to solve $Ax=b$, but A may have only a left inverse or right inverse (or no inverse). This discussion of how and when matrices have inverses improves our understanding of the four fundamental subspaces and of many other key topics in the course

# Left and right inverses; pseudoinverse

Although pseudoinverses will not appear on the exam, this lecture will help us to prepare.

## Two sided inverse

A *2-sided inverse* of a matrix $A$ is a matrix $A^{-1}$ for which $AA^{-1} = I = A^{-1}A$. This is what we've called the *inverse* of $A$. Here $r = n = m$; the matrix $A$ has full rank.

## Left inverse

Recall that $A$ has full column rank if its columns are independent; i.e. if $r = n$. In this case the nullspace of $A$ contains just the zero vector. The equation $A\mathbf{x} = \mathbf{b}$ either has exactly one solution $\mathbf{x}$ or is not solvable.

The matrix $A^TA$ is an invertible $n$ by $n$ symmetric matrix, so $(A^TA)^{-1}A^TA = I$. We say $A^{-1}_{\text{left}} = (A^TA)^{-1}A^T$ is a *left inverse* of $A$. (There may be other left inverses as well, but this is our favorite.) The fact that $A^TA$ is invertible when $A$ has full column rank was central to our discussion of least squares.

Note that $AA^{-1}_{\text{left}}$ is an $m$ by $m$ matrix which only equals the identity if $m = n$. A rectangular matrix can't have a two sided inverse because either that matrix or its transpose has a nonzero nullspace.

## Right inverse

If $A$ has full row rank, then $r = m$. The nullspace of $A^T$ contains only the zero vector; the rows of $A$ are independent. The equation $A\mathbf{x} = \mathbf{b}$ always has at least one solution; the nullspace of $A$ has dimension $n - m$, so there will be $n - m$ free variables and (if $n > m$) infinitely many solutions!

Matrices with full row rank have right inverses $A^{-1}_{\text{right}}$ with $AA^{-1}_{\text{right}} = I$. The nicest one of these is $A^T(AA^T)^{-1}$. Check: $A$ times $A^T(AAT)^{-1}$ is $I$.

## Pseudoinverse

An invertible matrix ($r = m = n$) has only the zero vector in its nullspace and left nullspace. A matrix with full column rank $r = n$ has only the zero vector in its nullspace. A matrix with full row rank $r = m$ has only the zero vector in its left nullspace. The remaining case to consider is a matrix $A$ for which $r < n$ and $r < m$.

If $A$ has full column rank and $A^{-1}_{\text{left}} = (A^TA)^{-1}A^T$, then

$$AA^{-1}_{\text{left}} = A(A^TA)^{-1}A^T = P$$

is the matrix which projects $\mathbb{R}^m$ onto the column space of $A$. This is as close as we can get to the product $AM = I$.

②

Similarly, if $A$ has full row rank then $A^{-1}_{\text{right}}A = A^T(AA^T)^{-1}A$ is the matrix which projects $\mathbb{R}^n$ onto the row space of $A$.

It's nontrivial nullspaces that cause trouble when we try to invert matrices. If $A\mathbf{x} = \mathbf{0}$ for some nonzero $\mathbf{x}$, then there's no hope of finding a matrix $A^{-1}$ that will reverse this process to give $A^{-1}\mathbf{0} = \mathbf{x}$.

The vector $A\mathbf{x}$ is always in the column space of $A$. In fact, the correspondence between vectors $\mathbf{x}$ in the ($r$ dimensional) row space and vectors $A\mathbf{x}$ in the ($r$ dimensional) column space is one-to-one. In other words, if $\mathbf{x} \neq \mathbf{y}$ are vectors in the row space of $A$ then $A\mathbf{x} \neq A\mathbf{y}$ in the column space of $A$. (The proof of this would make a good exam question.)

### Proof that if $\mathbf{x} \neq \mathbf{y}$ then $A\mathbf{x} \neq A\mathbf{y}$

Suppose the statement is false. Then we can find $\mathbf{x} \neq \mathbf{y}$ in the row space of $A$ for which $A\mathbf{x} = A\mathbf{y}$. But then $A(\mathbf{x} - \mathbf{y}) = \mathbf{0}$, so $\mathbf{x} - \mathbf{y}$ is in the nullspace of $A$. But the row space of $A$ is closed under linear combinations (like subtraction), so $\mathbf{x} - \mathbf{y}$ is also in the row space. The only vector in both the nullspace and the row space is the zero vector, so $\mathbf{x} - \mathbf{y} = \mathbf{0}$. This contradicts our assumption that $\mathbf{x}$ and $\mathbf{y}$ are not equal to each other.

We conclude that the mapping $\mathbf{x} \mapsto A\mathbf{x}$ from row space to column space is invertible. The inverse of this operation is called the *pseudoinverse* and is very useful to statisticians in their work with linear regression – they might not be able to guarantee that their matrices have full column rank $r = n$.

### Finding the pseudoinverse $A^+$

The *pseudoinverse* $A^+$ of $A$ is the matrix for which $\mathbf{x} = A^+A\mathbf{x}$ for all $\mathbf{x}$ in the row space of $A$. The nullspace of $A^+$ is the nullspace of $A^T$.

We start from the singular value decomposition $A = U\Sigma V^T$. Recall that $\Sigma$ is a $m$ by $n$ matrix whose entries are zero except for the singular values $\sigma_1, \sigma_2, ..., \sigma_r$ which appear on the diagonal of its first $r$ rows. The matrices $U$ and $V$ are orthonormal and therefore easy to invert. We only need to find a pseudoinverse for $\Sigma$.

The closest we can get to an inverse for $\Sigma$ is an $n$ by $m$ matrix $\Sigma^+$ whose first $r$ rows have $1/\sigma_1, 1/\sigma_2, ..., 1/\sigma_r$ on the diagonal. If $r = n = m$ then $\Sigma^+ = \Sigma^{-1}$. Always, the product of $\Sigma$ and $\Sigma^+$ is a square matrix whose first $r$ diagonal entries are 1 and whose other entries are 0.

If $A = U\Sigma V^T$ then its pseudoinverse is $A^+ = V\Sigma^+\mathbf{U}^T$. (Recall that $Q^T = Q^{-1}$ for orthogonal matrices $U$, $V$ or $Q$.)

We would get a similar result if we included non-zero entries in the lower right corner of $\Sigma^+$, but we prefer not to have extra non-zero entries.

### Conclusion

Although pseudoinverses will not appear on the exam, many of the topics we covered while discussing them (the four subspaces, the SVD, orthogonal matrices) are likely to appear.

MIT OpenCourseWare
http://ocw.mit.edu

18.06SC Linear Algebra
Fall 2011

For information about citing these materials or our Terms of Use, visit: http://ocw.mit.edu/terms.

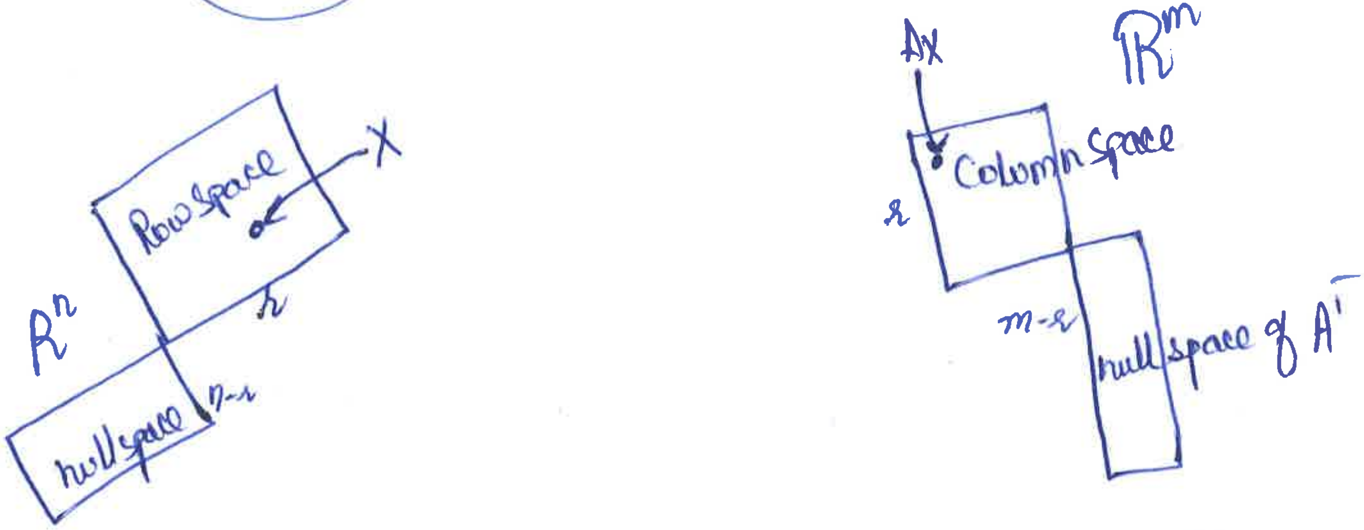
L#33 LA

4 subspaces

Left-inverses

Right-inverses

Pseudo inverses




2-sided inverse

$$AA^{-1} = I = A^{-1}A$$

$r = m = n$ (square-matrix)

$\boxed{\text{Full rank}}$

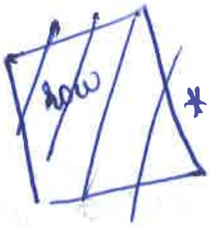
## Left inverse

full column rank $\overbrace{r=n}^{n<m}$ ($n$ - columns are independent)

Nullspace $= \{0\}$

Independent columns / Zero or One solution to $Ax=b$

(The particular solution is the solution)




* Whats deal with $A^TA$ here?

If $\text{rank}(A)=n$ then $A^TA$ will be invertible with rank $= n$.

$$(A^TA)^{-1}A^TA = I$$

$$\underbrace{(A^TA)^{-1}A^T}_{A^{-1}_{left}} A = I$$

Here $(AA^T)$ is bad

$$\Rightarrow \underset{n\times m}{A^{-1}_{left}}\ \underset{m\times n}{A} = I_{n\times n}$$

## Right Inverse

full row rank $r=m<n$

$N(A^T)=\{0\}$ We have independent rows

Solutions of $Ax=b$ ?? Since we have $n-m$ free variables thus we have $\boxed{\infty \text{ solutions to } Ax=b}$.

$$A A^{-1}_{right} = I$$

$$A \underbrace{A^T(AA^T)^{-1}}_{A^{-1}_{right}} = I$$

Here $A^TA$ is bad

Now from left inverse $A^{-1}_{left}A$, (6)

$$AA^{-1}_{left} = A(A^TA)^{-1}A^T = P \text{ (projection on column space of } A)$$

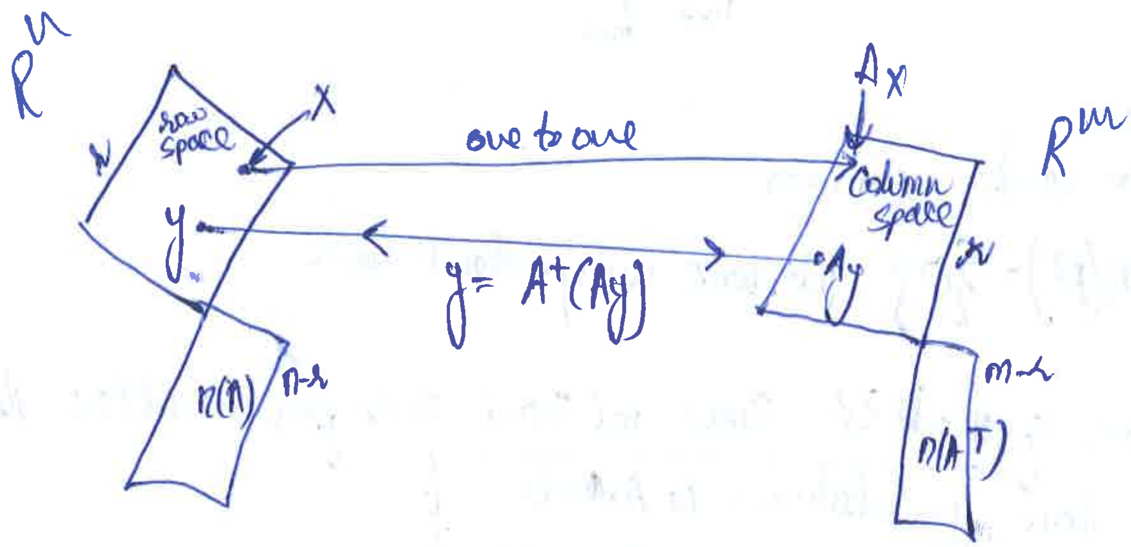
Now in right inverse:

$$AA^{-1}_{right} = I$$

$$A^{-1}_{right}A = A^T(AA^T)^{-1}A = \text{Projection onto row space of } A.$$

## General Case

Nullspace screw up inverses b/c they take every thing to zerovectors.




(

If $x \neq y$ BOTH in row space then

$$Ax \neq Ay$$

$\boxed{Ax, Ay \text{ in column space}}$

Proof: Suppose

$$Ax = Ay$$
$$A(x-y) = 0$$

Thus $(x-y)$ is in null space. If $x$ is in rowspace

& y is in rowspace so difference is also in rowspace
$\Rightarrow Ax, Ay$ are different vectors in columnspace.

Find pseudoinverse $A^+$.

① Start from SVD $A = U\Sigma V^T$

$$\Sigma = \underbrace{\begin{bmatrix} \sigma_1 & & 0 \\ & \ddots & \\ 0 & & \sigma_r \ddots \end{bmatrix}}_{n \text{ cols},\ \text{rank} = r} \begin{matrix} \} m \text{ rows} \\ m \times n \end{matrix}$$

$$\Sigma^+ = \begin{bmatrix} 1/\sigma_1 & & 0 \\ & \ddots & \\ 0 & 1/\sigma_r & 0 \end{bmatrix} n \times m$$

* Pseudoinverse is best close to inverse
* $$\Sigma\Sigma^+ = \underbrace{\begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots\ 1 & \\ 0 & & & 0 \end{bmatrix}}_{m \times m}$$ Projection onto columnspace

$$\underset{n\times m}{\Sigma^+}\ \underset{m \times n}{\Sigma} = \underset{n \times n}{\begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots\ 1 & \\ 0 & & & 0 \end{bmatrix}}$$ Projection onto rowspace

(Good pseudoinverse should be minimum matrix)

What PseudoInverse do? It projects on column space (if multiplied on right) & on row space (if multiplied on left) & gets rid of nullspaces.

As $A = U\Sigma V^T$

then $A^+ = V\Sigma^+ U^T$

SVD is best for finding pseudo inverse b/c we know how to find inverse for U & V in SVD. For $\Sigma$, we know that its diagonal thus we can calculate its pseudo inverse easily. thus pseudo inverse for A whose rank $r < (m \& n)$ can be found by
$A^+ = V\Sigma^+ U^T$.