Welcome to Math for Machine Learning!
Linear Algebra, Session 7

Disclaimer: Work in progress. Portions of these written materials are incomplete.

Copyright 2020 by Google LLC. This work is licensed under the Creative Commons CC BY 4.0 License.

Today's Agenda

Eigenvectors/values

The spectral theorem, PCA, SVD, and reconnecting these ideas with basic quantum probability theory.

Tensor Networks for ML

Loose Ends

Discuss affine spaces, manifolds, tensor products in data structures, and more!

Invariant Subspaces

To understand the behavior of a linear operator, it’s often helpful to look at subspaces that are “fixed” by it.

These are called *invariant subspaces*.

Disclaimer

The theory is rich and deep.

We'll just get a taste of it today.

Idea: Understand a linear operator by studying smaller “pieces” of it.

Example:

A vector space of polynomials.

Consider the set P_5 of polynomials of degree at most 5 with real coefficients.

$$p(x) = 3 + 2x^2 - x^5$$

$$q(x) = x - 7x^4$$

Consider the set P_5 of polynomials of degree at most 5 with real coefficients.

$$p(x) = 3 + 2x^2 - x^5$$

$$q(x) = x - 7x^4$$

This is a vector space.

$$p(x) + q(x) \in P_5 \quad \checkmark$$

Closed under addition.

a real number

$$k\,p(x) \in P_5 \quad \checkmark$$

Closed under scalar multiplication.

A **basis** for P_5 consists of **monomials** of degree 0 up to 5. These can be identified with standard basis vectors, so the dimension of the vector space P_5 is 6.

$$\{1, x, x^2, x^3, x^4, x^5\}$$

A "monomial" is a polynomial consisting of one term.

$$1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad x = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \ldots, \quad x^5 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

A **basis** for P_5 consists of **monomials** of degree 0 up to 5. These can be identified with standard basis vectors, so the dimension of the vector space P_5 is 6.

$$p(x) = 3 + 2x^2 - x^5 = 3\begin{bmatrix}1\\0\\0\\0\\0\\0\end{bmatrix} + 2\begin{bmatrix}0\\0\\1\\0\\0\\0\end{bmatrix} - \begin{bmatrix}0\\0\\0\\0\\0\\1\end{bmatrix} = \begin{bmatrix}3\\0\\2\\0\\0\\-1\end{bmatrix}$$

Example: Differentiation is a linear operator.

Example: Differentiation is a linear operator.

Consider the function M: $P_5 \to P_5$ defined by differentiation, $M(p(x)) = p'(x)$.

$$M(p(x)) = M(3 + 2x^2 - x^5) = 4x - 5x^4 = p'(x)$$

Example: Differentiation is a linear operator.

Consider the function $M: P_5 \to P_5$ defined by differentiation, $M(p(x)) = p'(x)$.

$$M(p(x)) = M(3 + 2x^2 - x^5) = 4x - 5x^4 = p'(x)$$

This is a linear operator.

Example: Differentiation is a linear operator.

Consider the function $M: P_5 \to P_5$ defined by differentiation, $M(p(x)) = p'(x)$.

$$M(p(x)) = M(3 + 2x^2 - x^5) = 4x - 5x^4 = p'(x)$$

This is a linear operator.

$$M(p(x) + q(x)) = p'(x) + q'(x) = M(p(x)) + M(q(x))$$

Follows from calculus: the derivative of the sum is the sum of the derivatives.

Example: Differentiation is a linear operator.

Consider the function $M: P_5 \to P_5$ defined by differentiation, $M(p(x)) = p'(x)$.

$$M(p(x)) = M(3 + 2x^2 - x^5) = 4x - 5x^4 = p'(x)$$

This is a linear operator.

the chain rule

$$M(kp(x)) = \cancel{k'p(x)} + kp'(x) = kM(p(x))$$

The derivative of a constant is 0.

Observation

Consider the subspace $P_3 \subseteq P_5$ consisting of polynomials of degree at most 3. It is **invariant** under the linear operator M.

Observation

Consider the subspace $P_3 \subseteq P_5$ consisting of polynomials of degree at most 3. It is **invariant** under the linear operator M.

If $p(x) \in P_3$ then $M(p(x)) \in P_3$, too.

"If the degree of p(x) is at most 3.."

"...then the degree of its derivative p'(x) is at most 3, too."

Given a linear operator $M: V \to V$,

a subspace $U \subseteq V$ is called <u>invariant</u>

<u>under M</u> if "$u \in U$" implies "$Mu \in U$".

"Whatever's in U *stays* in U."

Given a linear operator $M: V \to V$, a subspace $U \subseteq V$ is called <u>invariant under M</u> if "$u \in U$" implies "$Mu \in U$".

Example

P_3 is a 4-dimensional invariant subspace of the differentiation operator $M: P_5 \to P_5$.

That’s nice. Here’s something easier:

1-dimensional invariant subspaces

One-dimensional subspaces of a vector space V are easy to understand.

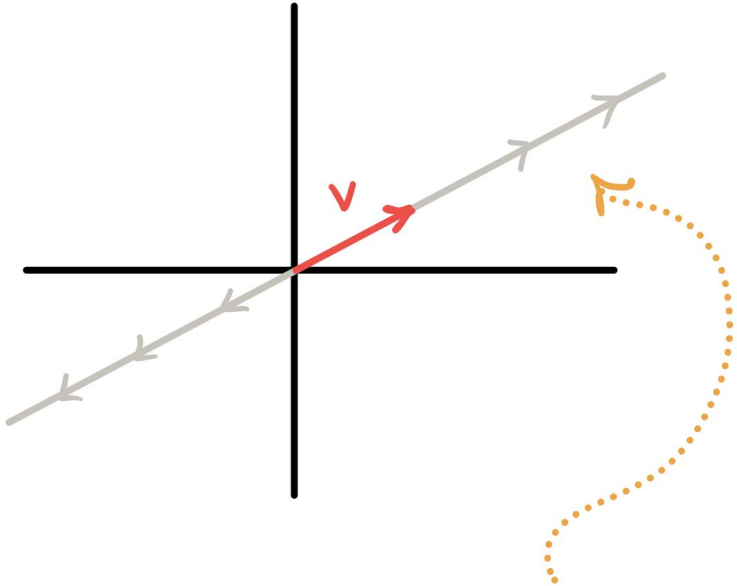
Pick any vector **v** in V and consider all scalar multiples of it. This is a 1-dimensional subspace of V.

One-dimensional subspaces of a vector space V are easy to understand.

Pick any vector **v** in V and consider all scalar multiples of it. This is a 1-dimensional subspace of V.

$$\mathcal{U} = \{\lambda v : \lambda \in F\}$$

F is the ground field of V.



This line is the set $\{\lambda v : \lambda \in F\}$

Now consider a linear operator M: V → V. *What does it mean for the 1-dimensional subspace U to be* ***invariant under M****?* Recall the definition:

Now consider a linear operator M: V → V. *What does it mean for the 1-dimensional subspace U to be* ***invariant under M****?* Recall the definition:

$\mathcal{U} = \{\lambda v \mid \lambda \in F\}$ is invariant under M if

"$v \in \mathcal{U}$" implies "$Mv \in \mathcal{U}$".

Now consider a linear operator M: V → V. *What does it mean for the 1-dimensional subspace U to be* ***invariant under M****?* Recall the definition:

$\mathcal{U} = \{\lambda v \mid \lambda \in F\}$ is invariant under M if

"$v \in \mathcal{U}$" implies $\boxed{\text{"}Mv \in \mathcal{U}\text{"}.}$

This means
$M\mathbf{v} = \lambda\mathbf{v}$
for some scalar λ.

A scalar λ is called an **eigenvalue** of a linear operator $M: V \to V$ if there is a *nonzero* vector **v** in V so that

$$\boxed{M\mathbf{v} = \lambda\mathbf{v}}$$

"eigenvalue"

"eigenvector"

in which case **v** is called the **eigenvector** of M corresponding to the eigenvalue λ.

We exclude the possibility that $\mathbf{v}=\mathbf{0}$, since otherwise $M\mathbf{0} = \lambda\mathbf{0}$ would be true for every operator M and every scalar λ, which isn't interesting.

Two perspectives:

Geometric vs. Algebraic

Geometric Perspective

As an example, consider a 2×2 matrix.

Consider the following 2×2 matrix, which represents a linear operator M: $\mathbb{R}^2 \to \mathbb{R}^2$. Does it have an eigenvector?

$$M = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

To find out, let’s see the transformation M “in action.”

“Eigenvectors and Eigenvalues” at 2:39 by Grant Sanderson (3Blue1Brown).

If $M = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$ then $v = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

is an eigenvector with

eigenvalue $\lambda = 2$:

$$Mv = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = 2 \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \lambda v$$

Algebraic Perspective

How to find eigenvalues/vectors?

$$Mv = \lambda v$$

$$Mv = \lambda v \quad \overset{\text{“if and only if”}}{\Longleftrightarrow} \quad Mv - \lambda v = 0$$

$$Mv = \lambda v \quad \overset{\text{“if and only if”}}{\Longleftrightarrow} \quad Mv - \lambda v = 0$$

$$\Longleftrightarrow \quad (M - \lambda I)v = 0$$

$$Mv = \lambda v \quad \overset{\text{“if and only if”}}{\Longleftrightarrow} \quad Mv - \lambda v = 0$$

$$\Longleftrightarrow \quad (M - \lambda I)v = 0$$

$$\lambda I = \lambda \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix}$$

$$= \begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{bmatrix}$$

I is notation for the n×n identity matrix.

$$Mv = \lambda v \quad \overset{\text{“if and only if”}}{\Longleftrightarrow} \quad Mv - \lambda v = 0$$

$$\Longleftrightarrow \quad (M - \lambda I)v = 0$$

$$\lambda I = \lambda \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{bmatrix}$$

I is notation for the n×n identity matrix.

$$\Longleftrightarrow \quad M - \lambda I \text{ is not injective}$$

A linear map is “injective” iff the only vector it sends to 0 is the zero vector.

$$Mv = \lambda v \quad \overset{\text{“if and only if”}}{\Longleftrightarrow} \quad Mv - \lambda v = 0$$

$$\Longleftrightarrow \quad (M - \lambda I)v = 0$$

$$\lambda I = \lambda \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{bmatrix}$$

I is notation for the n×n identity matrix.

$$\Longleftrightarrow \quad M - \lambda I \text{ is not injective}$$

A linear map is “injective” iff the only vector it sends to 0 is the zero vector.

$$\Longleftrightarrow \quad M - \lambda I \text{ is not invertible}$$

$$Mv = \lambda v \quad \overset{\text{“if and only if”}}{\Longleftrightarrow} \quad Mv - \lambda v = 0$$

$$\Longleftrightarrow \quad (M - \lambda I)v = 0$$

$$\lambda I = \lambda \begin{bmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{bmatrix} = \begin{bmatrix} \lambda & & & \\ & \lambda & & \\ & & \ddots & \\ & & & \lambda \end{bmatrix}$$

I is notation for the n×n identity matrix.

$$\Longleftrightarrow \quad M - \lambda I \text{ is not injective}$$

A linear map is “injective” iff the only vector it sends to 0 is the zero vector.

$$\Longleftrightarrow \quad M - \lambda I \text{ is not invertible}$$

$$\Longleftrightarrow \quad \det(M - \lambda I) = 0$$

Characteristic equation

An n×n matrix M has an eigenvalue λ if and only if

$$\det(M - \lambda I) = 0$$

Characteristic equation

An n×n matrix M has an eigenvalue λ if and only if

$$\underbrace{\det(M - \lambda I)}_{\text{"characteristic polynomial of M"}} = 0$$

Example

$\boxed{\text{Example}}$ Suppose $M = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$. consider the matrix $M - \lambda I$.

$\boxed{\text{Example}}$ Suppose $M = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$. consider the matrix $M - \lambda I$.

$$M - \lambda I = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}$$

$\boxed{\text{Example}}$ Suppose $M = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$. consider the matrix $M - \lambda I$.

$$M - \lambda I = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}$$

Now find the determinant & set = 0

$\boxed{\text{Example}}$ Suppose $M = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$. consider the matrix $M - \lambda I$.

$$M - \lambda I = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}$$

Now find the determinant & set = 0

$$\det\left(\begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}\right) = \underbrace{(3-\lambda)(2-\lambda) - 0}_{\lambda^2 - 5\lambda + 6} = 0$$

$\boxed{\text{Example}}$ Suppose $M = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}$. consider the matrix $M - \lambda I$.

$$M - \lambda I = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} = \begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}$$

Now find the determinant & set = 0

$$\det\left(\begin{bmatrix} 3-\lambda & 1 \\ 0 & 2-\lambda \end{bmatrix}\right) = \underbrace{(3-\lambda)(2-\lambda) - 0}_{\lambda^2 - 5\lambda + 6} = 0$$

iff $\boxed{\lambda = 3,\ \lambda = 2}$ ← eigenvalues of M

$\boxed{\text{Example}}$ How to find the eigen<u>vector</u> corresponding to $\lambda = 2$?

$\boxed{\text{Example}}$ How to find the eigenvector corresponding to $\lambda = 2$?

$$(M - \lambda I)v = 0 \xrightarrow{\text{"plug and chug"}} \left(\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right) \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{let } v = \begin{bmatrix} x \\ y \end{bmatrix}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\boxed{\text{Example}}$ How to find the eigenvector corresponding to $\lambda = 2$?

$$(M - \lambda I)v = 0 \xrightarrow{\text{"plug and chug"}} \left(\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right) \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{let } v = \begin{bmatrix} x \\ y \end{bmatrix}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Longleftrightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$\boxed{\text{Example}}$ How to find the eigenvector corresponding to $\lambda = 2$?

$$(M - \lambda I)v = 0 \xrightarrow{\text{"plug and chug"}} \left(\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right) \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{let } v = \begin{bmatrix} x \\ y \end{bmatrix}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow x + y = 0$$

$$\Leftrightarrow x = -y$$

$\boxed{\text{Example}}$ How to find the eigenvector corresponding to $\lambda = 2$?

$$(M - \lambda I)v = 0 \xrightarrow{\text{"plug and chug"}} \left(\begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix} - \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \right) \underbrace{\begin{bmatrix} x \\ y \end{bmatrix}}_{\text{let } v = \begin{bmatrix} x \\ y \end{bmatrix}} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\Leftrightarrow x + y = 0$$

$$\Leftrightarrow x = -y$$

$$\Leftrightarrow v = \begin{bmatrix} -y \\ y \end{bmatrix} \longrightarrow \boxed{v = \begin{bmatrix} -1 \\ 1 \end{bmatrix}} \checkmark$$

“Eigenvectors and Eigenvalues” at 2:39 by Grant Sanderson (3Blue1Brown).

There’s lots more we could say.

Here one thing to know.

Do all square matrices have eigenvalues?

Do all square matrices have eigenvalues?

NO. As an example, consider the matrix representing CCW rotation by 90° in $\mathbb{R}^2$.

Do all square matrices have eigenvalues?

NO. As an example, consider the matrix representing CCW rotation by 90° in $\mathbb{R}^2$.

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

Check:

The characteristic polynomial of this matrix is $p(\lambda)=\lambda^2+1$ which does not have solutions in $\mathbb{R}$. *No coincidence.*

So, *which* matrices?

So, *which* matrices?

Towards answering this question, we first need a definition.

Def. (lite): The adjoint of a linear operator $M : V \to V$ is the linear operator $M^{\dagger} : V \to V$ whose matrix representation is the conjugate transpose of the matrix representation of M.

This lite definition holds if the matrices are with respect to an orthonormal basis on V.

The "real" definition involves inner products. Look it up when you have a chance!

"conjugate

transpose"

"conjugate transpose"

The complex conjugate of a complex number $a+bi$ is the number $a-bi$.

The transpose of a matrix M is another matrix whose i^{th} column (row) is the ith row (column) of M.

"conjugate transpose"

The complex conjugate of a complex number $a+bi$ is the number $a-bi$.

$3-7i$ is the complex conjugate of $3+7i$

The transpose of a matrix M is another matrix whose i^{th} column (row) is the ith row (column) of M.

$$\begin{bmatrix} 3 & 7 \\ 1 & 0 \end{bmatrix}$$

is the transpose of

$$\begin{bmatrix} 3 & 1 \\ 7 & 0 \end{bmatrix}$$

The conjugate transpose of a matrix M is another matrix M^{+} obtained by first transposing M, then taking the complex conjugate of each entry (or vice versa).

$$M = \begin{bmatrix} 2+5i & 0 \\ 3i & 7 \end{bmatrix} \rightsquigarrow M^{+} = \begin{bmatrix} 2-5i & -3i \\ 0 & 7 \end{bmatrix}$$

Self-Adjoint Operators

A matrix is called **self-adjoint** if it is equal to its conjugate transpose.

$$M = M^{\dagger}$$

If the entries of $M=M^{\dagger}$ are real, then M is called a **symmetric matrix.**

$$M = \begin{bmatrix} 3 & -2i \\ 2i & 0 \end{bmatrix}$$

$$M^{\dagger} = \begin{bmatrix} 3 & -2i \\ 2i & 0 \end{bmatrix}$$

Self-adjoint matrices are guaranteed to have eigenvalues/eigenvectors. In fact...

Spectral Theorem

Let M be a **self-adjoint** operator on a complex Euclidean space V of dimension n. Then M has an *orthonormal* set of n eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ with eigenvalues $\{\lambda_1, \lambda_2, ..., \lambda_n\}$.

Moreover,

Spectral Theorem

Let M be a **self-adjoint** operator on a complex Euclidean space V of dimension n. Then M has an *orthonormal* set of n eigenvectors $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n\}$ with eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$.

Moreover,

$$M = \lambda_1 u_1 u_1^* + \lambda_2 u_2 u_2^* + \cdots + \lambda_n u_n u_n^*$$

Each $\mathbf{u}_i\mathbf{u}_i^*$ is the outer product of the ith eigenvector of M with itself. In other words, $\mathbf{u}_i\mathbf{u}_i^*$ is the rank 1 projection operator onto the line spanned by $\mathbf{u}_i$.

Flashback

	Usual	Bra-Ket	Arrays
vector	v	$\lvert v\rangle$	$\begin{bmatrix} \vdots \end{bmatrix}$
co-vector	v^*	$\langle v\rvert$	$\begin{bmatrix} \cdots \end{bmatrix}$
tensor product	$v \otimes w$	$\lvert v\rangle \otimes \lvert w\rangle = \lvert v\rangle\lvert w\rangle = \lvert vw\rangle$	$\begin{bmatrix} \vdots \end{bmatrix} \otimes \begin{bmatrix} \vdots \end{bmatrix}$
inner product	$\langle v, w\rangle$	$\langle v \rvert w\rangle$	$\underset{v}{\begin{bmatrix} \cdots \end{bmatrix}} \underset{w}{\begin{bmatrix} \vdots \end{bmatrix}}$
linear map V → W	$w \otimes v^*$	$\lvert w\rangle\langle v\rvert$	$\underset{w}{\begin{bmatrix} \vdots \end{bmatrix}} \underset{v}{\begin{bmatrix} \cdots \end{bmatrix}}$

Spectral Theorem

$$\boxed{M = \lambda_1 u_1 u_1^* + \lambda_2 u_2 u_2^* + \cdots + \lambda_n u_n u_n^*}$$

the **spectral decomposition** of M

*Why? The set of eigenvalues $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ is called the **spectrum** of M.*

More succinctly...

There exist operators U and D so that

$$M = UDU^{-1}$$

where U is invertible and $U^{\dagger}=U^{-1}$ and D is a **diagonal operator**.

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$

The columns of U
are the **eigenvectors** of M.

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

The diagonal entries of D
are the **eigenvalues** of M.

More succinctly...

If U and D are the matrices below,

$$U = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}$$

Because the $\mathbf{u}_i$ are orthonormal, the matrix U is invertible and $U^{-1}=U^{\dagger}$.

$$D = \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix}$$

The columns of U are the **eigenvectors** of M.

The diagonal entries of D are the **eigenvalues** of M.

then M can be written as the following product:

$$M = UDU^{-1}$$

$$M = UDU^{-1}$$

If a matrix U is invertible and its inverse is equal to its conjugate transpose, then U is called a <u>unitary matrix</u>.

$$UU^{+} = U^{+}U = I$$

$$\Rightarrow \quad U^{-1} = U^{+}$$

If the entries of U are real numbers, then U is called an **orthogonal matrix**.

$$M = UDU^{-1}$$

If a matrix U is invertible and its inverse is equal to its conjugate transpose, then U is called a <u>unitary matrix</u>.

$$UU^{+} = U^{+}U = I$$

$$\Rightarrow \quad U^{-1} = U^{+}$$

Proof of Spectral Theorem

Proof of Spectral Theorem

Omit.

Proof of Spectral Theorem

Omit.

The theorem holds more generally for normal operators, of which self-adjoint operators are an example.

See Axler's *Linear Algebra Done Right*.

Example

$\boxed{\text{Example}}$

The matrix $M = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ has eigenvectors/values:

$\boxed{\text{Example}}$

The matrix $M = \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}$ has eigenvectors/values:

$$\begin{cases} \lambda_1 = 3 \\ e_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{cases} \quad \& \quad \begin{cases} \lambda_2 = 2 \\ e_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \end{cases}$$

By the Spectral Theorem, M can be written as

$$M = 3e_1e_1^* + 2e_2e_2^*$$

By the Spectral Theorem, M can be written as

$$M = 3e_1e_1^* + 2e_2e_2^*$$

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

By the Spectral Theorem, M can be written as

$$M = 3e_1e_1^* + 2e_2e_2^*$$

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

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

$$= \begin{bmatrix}3 & 0\\0 & 2\end{bmatrix} \checkmark$$

or more succinctly ...

first column is $\mathbf{e}_1$

second column is $\mathbf{e}_2$

Diagonals are eigenvalues

$$M = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$= U \quad D \quad U^{-1}$$

or more succinctly ...

first column is $\mathbf{e}_1$

second column is $\mathbf{e}_2$

Diagonals are eigenvalues

BORING.

$$= U D U^{-1}$$

more interesting

$\boxed{\text{Example}}$

more interesting

Example

The matrix $M = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ has

three eigenvalues & eigenvectors:

more interesting

Example

The matrix $M = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ has

three eigenvalues & eigenvectors:

$\lambda_1 = 2$, $e_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$ $\lambda_2 = 1$, $e_1 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$ $\lambda_3 = 0$, $e_3 = \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix}$

By the Spectral Theorem, M may be written as

$$M = \lambda_1 e_1 e_1^* + \lambda_2 e_2 e_2^* + \lambda_3 e_3 e_3^*$$

By the Spectral Theorem, M may be written as

$$M = \lambda_1 e_1 e_1^* + \lambda_2 e_2 e_2^* + \lambda_3 e_3 e_3^*$$

$$= 2 \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix} \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} + 0$$

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

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

or more succinctly ...

$$M = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}}_{U^{-1}}$$

or more succinctly ...

first column is $\mathbf{e}_1$

second column is $\mathbf{e}_2$

third column is $\mathbf{e}_3$

diagonals are eigenvalues

$$M = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & 0 & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ 0 & 1 & 0 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 1 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix}}_{U^{-1}}$$

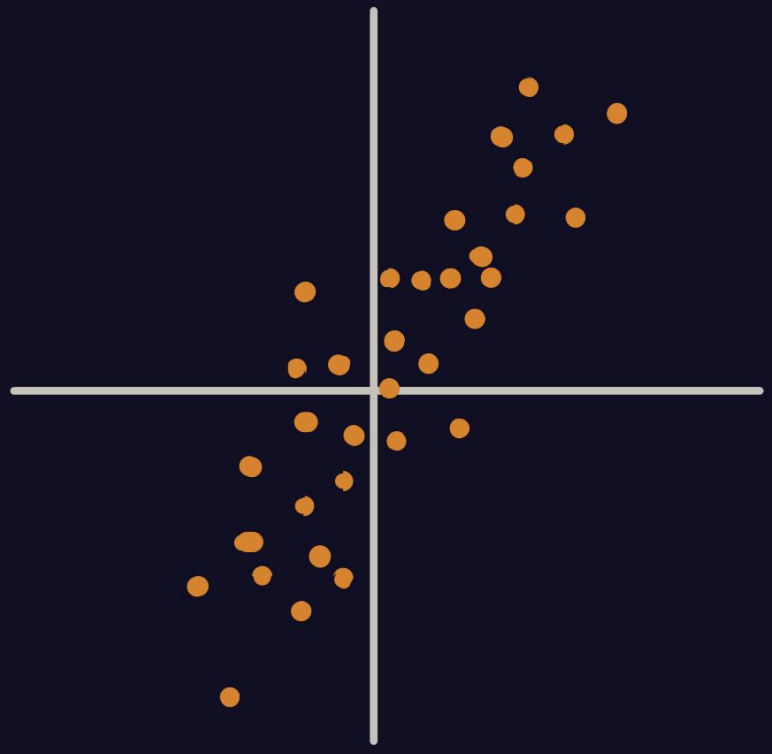
Okay. So what?

One answer: PCA

Suppose we have n data points in m-dimensional space.

Oftentimes, these points may be **clustered** along a line or other low-dimensional subspace.

What is that line or subspace?



To find this line, first organize the data points into an $m \times n$ matrix A, and then compute the eigenvectors of $AA^T/n-1$.

Exercise: Convince yourself that AA^T is self-adjoint!

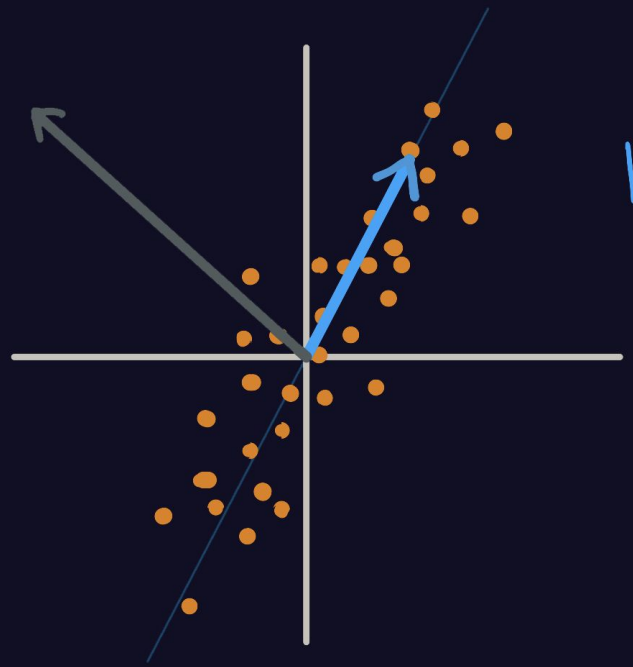
$$
\begin{array}{cc}
 & \begin{array}{cccccc} \text{student 1} & \text{student 2} & \text{student 3} & \text{student 4} & \text{student 5} & \text{student 6} \end{array} \\
A = & \begin{bmatrix} 3 & -4 & 7 & 1 & -4 & -3 \\ 7 & -6 & 8 & -1 & -1 & -7 \end{bmatrix} \begin{array}{l} \text{math} \\ \text{history} \end{array}
\end{array}
$$

There's a little more to this. We first center the data points by subtracting the *mean* of each row. The matrix $AA^T/n-1$ is called a sample covariance matrix.

In this example, the matrix $AA^T/5$ has two eigenvectors, shown below. The eigenvector with the largest eigenvalue picks out the **principal direction** of the data.

$$|e_2\rangle = \begin{bmatrix} -1.4 \\ 1 \end{bmatrix}$$

"scale factor" is 3



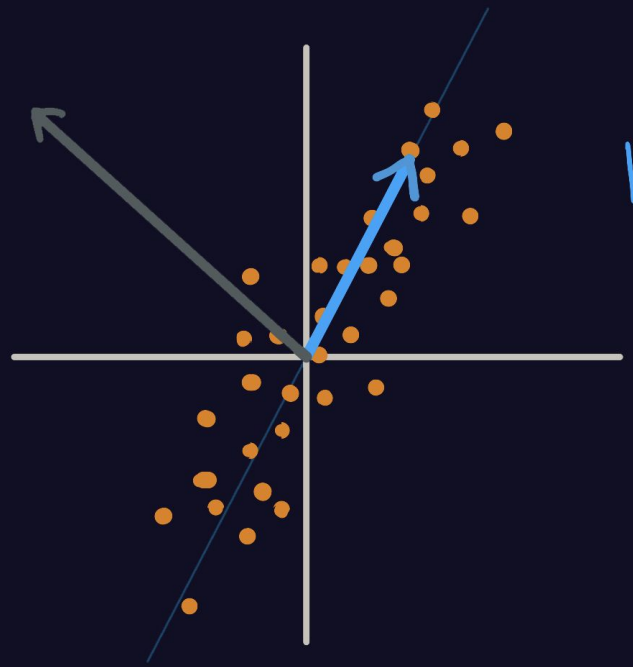
$$|e_1\rangle = \begin{bmatrix} .6 \\ .8 \end{bmatrix}$$

"scale factor" is 57

This is the idea behind **principal component analysis** (PCA). For more on this example, see ch. 7 of Gilbert Strang's *Introduction to Linear Algebra*, 5th ed.

$$|e_2\rangle = \begin{bmatrix} -1.4 \\ 1 \end{bmatrix}$$

"scale factor" is 3



$$|e_1\rangle = \begin{bmatrix} .6 \\ .8 \end{bmatrix}$$

"scale factor" is 57

Okay. So what?

Another answer: Quantum Probability

Think back to quantum computing.

(time permitting)

A pure **quantum state** of a quantum many body system is a unit vector in a **tensor product** of spaces.

$$|\psi\rangle \in V \otimes V$$

A **quantum gate** is a **linear transformation** satisfying a special property.

$$\mathcal{U} : V \otimes V \longrightarrow V \otimes V$$

A **quantum circuit** is a composition of **tensor products** of quantum gates.

$$\mathcal{U}' \circ (\mathcal{U} \otimes \mathrm{id}_V) : V^{\otimes 3} \longrightarrow V^{\otimes 4}$$

?

Think back to quantum computing.

(time permitting)

A pure **quantum state** of a quantum many body system is a unit vector in a **tensor product** of spaces.

$$|\psi\rangle \in V \otimes V$$

A **quantum gate** is a **linear transformation** satisfying a special property.

$$\mathcal{U} : V \otimes V \longrightarrow V \otimes V$$

A **quantum circuit** is a composition of **tensor products** of quantum gates.

$$\mathcal{U}' \circ (\mathcal{U} \otimes \mathrm{id}_V) : V^{\otimes 3} \longrightarrow V^{\otimes 4}$$

Behind the scenes

From a linear algebraic perspective, a “quantum state” is a special kind of self-adjoint operator ρ, called a **density operator.**

Behind the scenes

From a linear algebraic perspective, a “quantum state” is a special kind of self-adjoint operator ρ, called a **density operator.**

Because ρ is self-adjoint, it has a spectral decomposition:

$$\rho = \lambda_1 |u_1\rangle\langle u_1| + \lambda_2 |u_2\rangle\langle u_2| + \cdots + \lambda_n |u_n\rangle\langle u_n|$$

Punchline: The state ρ is described with a certain adjective, depending on the number of **nonzero eigenvalues** it has.

Pure Quantum State

If all eigenvalues are 0 except for one, then ρ is a called **pure state** and it can be associated with a unit vector.

$$\rho = \lambda_1 |u_1\rangle\langle u_1| + \overbrace{\lambda_2 |u_2\rangle\langle u_2| + \cdots + \lambda_n |u_n\rangle\langle u_n|}^{\text{ZERO}}$$

ρ is projection onto the unit vector $|u_1\rangle$, and λ_1 must equal 1 because density operators are (by definition) required to have trace 1.

Mixed Quantum State

Otherwise, ρ is called a **mixed state**. Think of it as a "mixture" of pure states (i.e. rank 1 density operators).

a bunch of pure states

$$\rho = \lambda_1 |u_1\rangle\langle u_1| + \lambda_2 |u_2\rangle\langle u_2| + \cdots + \lambda_n |u_n\rangle\langle u_n|$$

Want to learn more?

More info in the following slides.

A quantum state is the quantum version of a probability distribution.

The quantum version of a probability distribution on X is a particular **linear operator** on $\mathbb{C}^X$.

$$\pi : X \to \mathbb{R} \quad \rightsquigarrow \quad \rho : \mathbb{C}^X \to \mathbb{C}^X$$

must satisfy conditions analogous to probabilities

A **probability distribution** on a finite set X is a function:

$$\pi : X \to \mathbb{R}$$

real-valued function

$$\pi(x) \geq 0$$

values are nonnegative

$$\sum_{x \in X} \pi(x) = 1$$

values sum to 1

Example

Let $X = \{H, T\}$ denote the outcomes of a fair coin toss.

$$\pi(H) = \frac{1}{2} \qquad \pi(T) = \frac{1}{2}$$

The quantum version of a probability distribution on a finite set X is a **quantum state** on $\mathbb{C}^X$.

<u>Answer (v.0):</u>

A quantum state is a **unit vector** in $\mathbb{C}^X$.

$$|\psi\rangle \in \mathbb{C}^X$$

(last time)

<u>Answer (v.1):</u>

A quantum state is a **density operator** on $\mathbb{C}^X$.

$$\rho : \mathbb{C}^X \longrightarrow \mathbb{C}^X$$

(better)

$\rho : \mathbb{C}^x \to \mathbb{C}^x$ is a <u>density operator</u> if:

It's **positive semidefinite**.

$\langle v|\rho|v\rangle \geq 0 \quad \forall |v\rangle$

It's **self-adjoint**.

$\rho = \rho^\dagger$

Its **trace is 1**.

$\mathrm{tr}(\rho) = 1$

$\rho : \mathbb{C}^x \to \mathbb{C}^x$ is a <u>density operator</u> if:

It's **positive semidefinite.**

$$\langle v|\rho|v\rangle \geq 0 \quad \forall |v\rangle$$

This guarantees its eigenvalues are **nonnegative**.

It's **self-adjoint.**

$$\rho = \rho^\dagger$$

This guarantees its eigenvalues are **real**.

Its **trace is 1.**

$$\mathrm{tr}(\rho) = 1$$

This guarantees its eigenvalues sum to 1.

Example

Every probability distribution defines a density operator. (And in more than one way!)

1. A Diagonal Operator

Consider a probability distribution on a 3-element set $\pi = (1/7, 2/7, 4/7)$.

1. A Diagonal Operator

Consider a probability distribution on a 3-element set π = (1/7, 2/7, 4/7).

$$\rho = \begin{bmatrix} 1/7 & 0 & 0 \\ 0 & 2/7 & 0 \\ 0 & 0 & 4/7 \end{bmatrix} : \mathbb{C}^3 \longrightarrow \mathbb{C}^3$$

positive semidefinite
(entries are nonnegative)
✓

self-adjoint
(matrix is symmetric)
✓

unit trace
(sum of diagonals is 1)
✓

2. A Rank 1 Projection Operator

Again, suppose $\pi = (1/7, 3/7, 4/7)$ and consider the orthogonal projection onto the following unit vector:

2. A Rank 1 Projection Operator

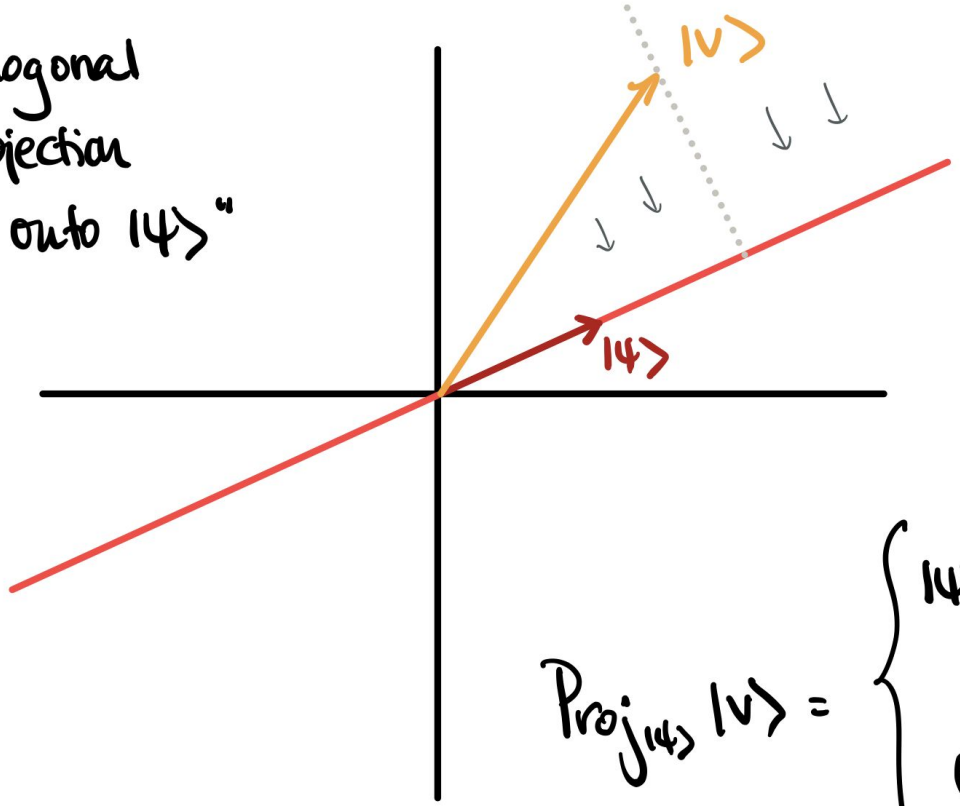
Again, suppose π = (1/7, 3/7, 4/7) and consider the orthogonal projection onto the following unit vector:

$$|\psi\rangle = \sqrt{\tfrac{1}{7}}\,|x\rangle + \sqrt{\tfrac{2}{7}}\,|y\rangle + \sqrt{\tfrac{4}{7}}\,|z\rangle \in \mathbb{C}^X$$

$$= \begin{bmatrix} \sqrt{1/7} \\ \sqrt{2/7} \\ \sqrt{4/7} \end{bmatrix}$$

sum of elements in X weighted by the square roots of their probabilities.

"orthogonal projection onto $|\psi\rangle$"



$$\mathrm{Proj}_{|\psi\rangle}|v\rangle = \begin{cases} |\psi\rangle & \text{if } |v\rangle = |\psi\rangle \\ 0 & \text{if } |v\rangle \perp |\psi\rangle \end{cases}$$

$$\rho = \mathrm{Proj}_{|\psi\rangle}$$

$$= |\psi\rangle\langle\psi|$$

the outer product of $|\psi\rangle$ with itself

$$= \begin{bmatrix} \sqrt{1/7} \\ \sqrt{2/7} \\ \sqrt{4/7} \end{bmatrix} \begin{bmatrix} \sqrt{\frac{1}{7}} & \sqrt{\frac{2}{7}} & \sqrt{\frac{4}{7}} \end{bmatrix}$$

a rank 1 operator

$$= \begin{bmatrix} \frac{1}{7} & \frac{\sqrt{2}}{7} & \frac{2}{7} \\ \frac{\sqrt{2}}{7} & \frac{2}{7} & \frac{\sqrt{8}}{7} \\ \frac{2}{7} & \frac{\sqrt{8}}{7} & \frac{4}{7} \end{bmatrix} : \mathbb{C}^3 \to \mathbb{C}^3$$

✓ **positive semidefinite**
(inner product with any vector $|v\rangle$ is $|\langle v|\psi\rangle|^2 \geq 0$.)

✓ **self-adjoint**
(matrix is symmetric)

✓ **unit trace**
(sum of diagonals is 1)

$\mathrm{Proj}_{|\psi\rangle}$ has $\underbrace{\underline{\underline{\text{one}}}\ \text{eigenvector}}_{|\psi\rangle}$ and $\underline{\underline{\text{one}}}$ nonzero $\underbrace{\text{eigenvalue}}_{\lambda = 1}$

$$\mathrm{Proj}_{|\psi\rangle}|\psi\rangle = |\psi\rangle\langle\psi|\psi\rangle$$

$$= |\psi\rangle \cdot 1$$

$$= |\psi\rangle$$

Take a look at their spectral decompositions.

<u>option #1</u>

$$\rho = \begin{bmatrix} \frac{1}{7} & 0 & 0 \\ 0 & \frac{2}{7} & 0 \\ 0 & 0 & \frac{4}{7} \end{bmatrix}$$

A **rank 3** density operator with 3 nonzero eigenvalues

$$= \frac{1}{7}|e_1\rangle\langle e_1| + \frac{2}{7}|e_2\rangle\langle e_2| + \frac{4}{7}|e_3\rangle\langle e_3|$$

Here, the $|e_i\rangle$ are the standard basis vectors in $\mathbb{R}^3$, i.e. “one hot encodings.”

Take a look at their spectral decompositions.

<u>option #2</u>

$$\rho = \begin{bmatrix} \frac{1}{7} & \frac{\sqrt{2}}{7} & \frac{2}{7} \\ \frac{\sqrt{2}}{7} & \frac{2}{7} & \frac{\sqrt{8}}{7} \\ \frac{2}{7} & \frac{\sqrt{8}}{7} & \frac{4}{7} \end{bmatrix}$$

$$= |\psi\rangle\langle\psi|$$

A **rank 1** density operator with 1 nonzero eigenvalue

These have names.

option # 1	option # 2
a density with rank > 1	a density with rank = 1

These have names.

mixed

quantum state

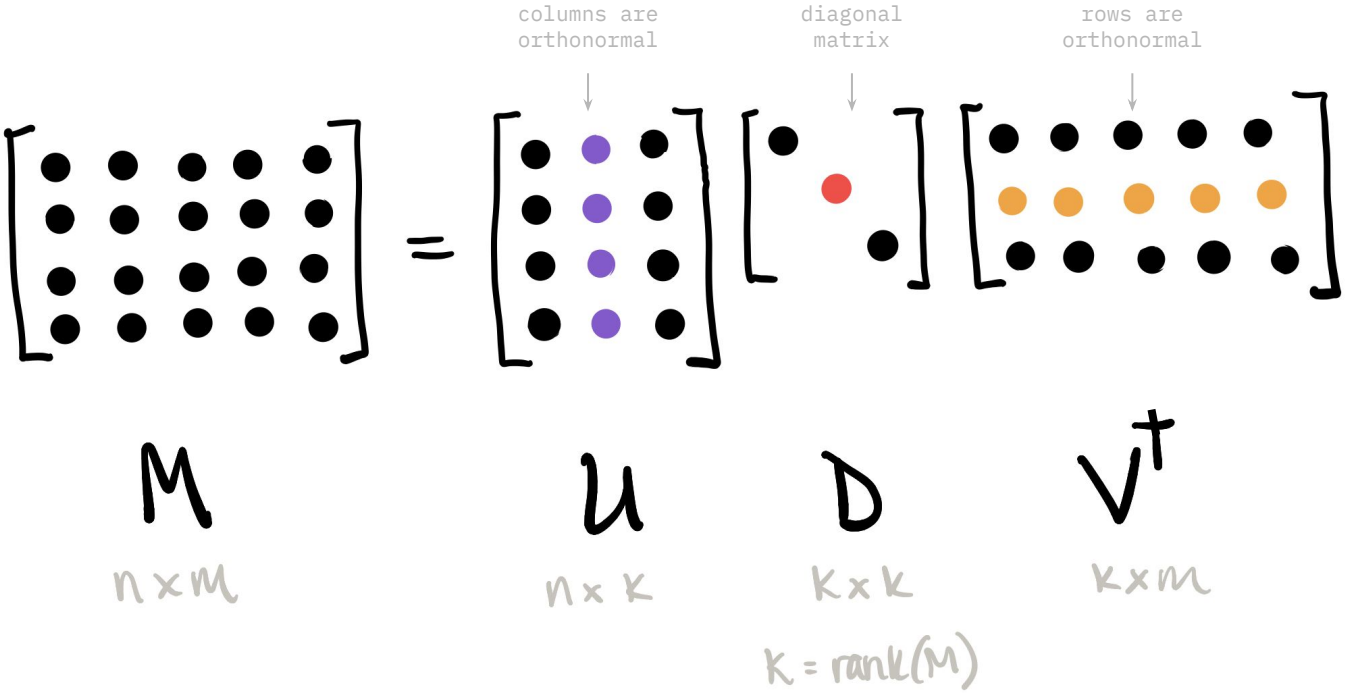
pure

quantum state

The Spectral Theorem is powerful. But what about matrices that aren’t square?

Singular Value Decomposition

Theorem: Every n×m matrix M can be written as



Singular Value Decomposition

Theorem: Every n×m matrix M can be written as

"Every"!

Don't think of SVD as a second-rate theorem.

It's a powerful statement.

In general, factorization is **hard**. Modern cryptography is based on this fact.

And yet in linear algebra, **every** matrix—not just those with special properties—can be factored in this very nice way.

Singular Value Decomposition

Theorem: Every n×m matrix M can be written as

$$M = UDV^T$$

Columns of U are called the **left singular vectors** of M.

Entries along the diagonal of D are called the **singular values** of M.

Columns of V are called the **right singular vectors** of M.

Collaborative Filtering