Formal Concepts vs. Eigenvectors of Density Operators

Tai-Danae Bradley

CUNY Graduate Center

Outline[1]

1. example / theory
2. application
3. formal concepts
4. motivation

[1] T.-D. B., *At the Interface of Algebra and Statistics*, arxiv:2004.05631.

A joint probability distribution gives rise to marginal probability distributions.

Marginalizing loses information.

Consider a **joint probability distribution** on the Cartesian product of finite sets.

$$\pi\colon X \times Y \to \mathbb{R}, \qquad \sum_{x,y} \pi(x,y) = 1, \qquad \pi(x,y) \geq 0$$

This gives rise to a **marginal distribution**.

$$\pi_X\colon X \to \mathbb{R}$$

$$\pi_X(x) = \sum_y \pi(x,y)$$

orange fruit

green fruit

purple vegetable

Here is a joint distribution.

	orange	green	purple
fruit	$\frac{1}{3}$	$\frac{1}{3}$	0
vegetable	0	0	$\frac{1}{3}$

$X = \{\text{orange, green, purple}\}$ $\quad Y = \{\text{fruit, vegetable}\}$

Marginal probabilities are sums of rows and columns.

$$\pi_X = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right) \quad \leftrightsquigarrow \quad \begin{matrix} \text{orange} \\ \text{green} \\ \text{purple} \end{matrix}$$

$$\pi_Y = \left(\frac{2}{3}, \frac{1}{3}\right) \quad \leftrightsquigarrow \quad \begin{matrix} \text{fruit} \\ \text{fruit} \\ \text{vegetable} \end{matrix}$$

Marginal probability doesn't have memory.

- The marginal probability of **fruit** is 2/3, but that doesn't tell us that half the fruits are **orange** and half are **green**.
- The marginal probability of **vegetable** is 1/3, but that doesn't tell us that all of the vegetables are **purple**.

fruit fruit vegetable

There's another way.

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

$$M^\dagger M = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}$$

$$
\begin{array}{c} \\ \text{orange} \\ \text{green} \\ \text{purple} \end{array}
\begin{array}{c}
\begin{array}{ccc} \text{orange} & \text{green} & \text{purple} \end{array} \\
\begin{bmatrix} \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{3} \end{bmatrix}
\end{array}
\quad \leftrightsquigarrow \quad \pi_X = \left(\tfrac{1}{3}, \tfrac{1}{3}, \tfrac{1}{3}\right)
$$

The squares of the entries of the eigenvectors of $M^\dagger M$ define **conditional probability** distributions on X.

orange fruit

green fruit

$$\begin{bmatrix} \sqrt{\frac{1}{2}} \\ \sqrt{\frac{1}{2}} \\ 0 \end{bmatrix} \begin{matrix} \pi(\text{orange}|\text{fruit}) \\ \pi(\text{green}|\text{fruit}) \\ \pi(\text{purple}|\text{fruit}) \end{matrix}$$

$$\begin{matrix} \pi(\text{orange}|\text{vegetable}) \\ \pi(\text{green}|\text{vegetable}) \\ \pi(\text{purple}|\text{vegetable}) \end{matrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

purple vegetable

What's really going on?

Let π be any probability distribution on $X \times Y$. Define a matrix M by

$$M_{yx} := \sqrt{\pi(x, y)}$$

These two operators are special:

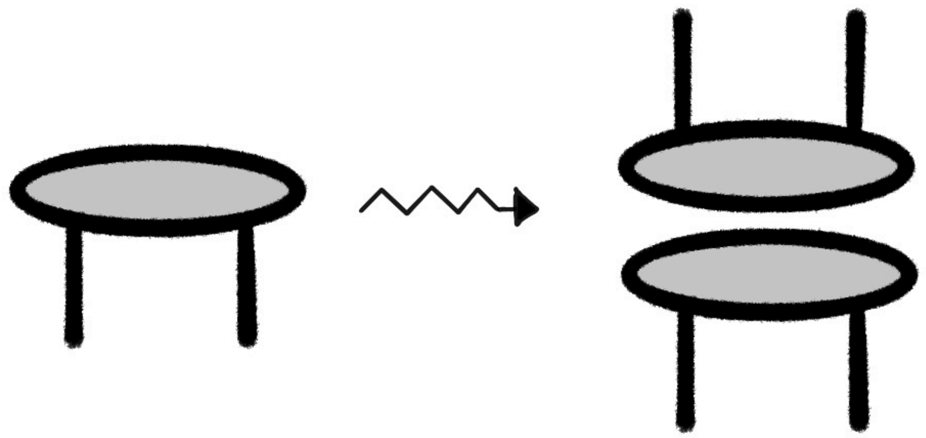
$$M^\dagger M \colon \mathbb{C}^X \to \mathbb{C}^X \qquad\qquad M M^\dagger \colon \mathbb{C}^Y \to \mathbb{C}^Y.$$

This is part of a larger story.

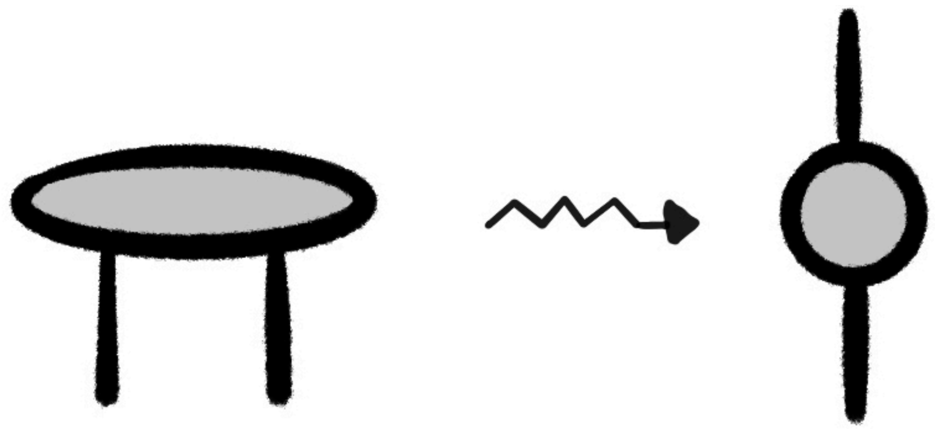
Every probability distribution on $X \times Y$ defines a particular **linear operator** on $\mathbb{C}^X \otimes \mathbb{C}^Y$, namely orthogonal projection onto this unit vector:

$$\psi := \begin{bmatrix} \sqrt{\pi(x_1, y_1)} & \cdots & \sqrt{\pi(x_n, y_m)} \end{bmatrix}^\top$$

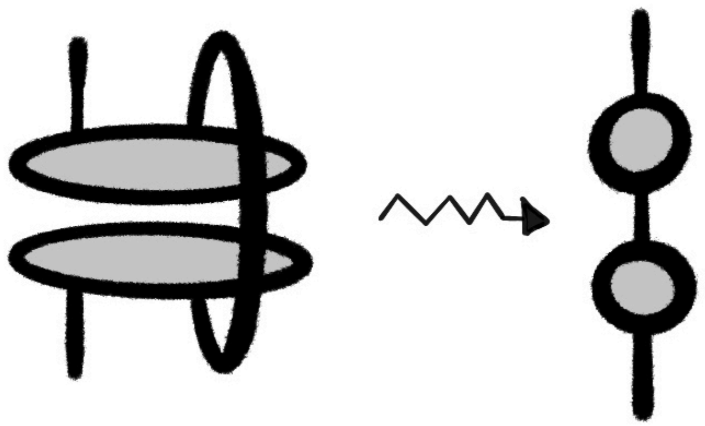
It is a **density operator.** As diagrams:



The linear map associated to the vector ψ is M.



The operators $M^\dagger M$ and $MM^\dagger$ are **reduced densities** associated to the projection onto ψ.



Think of reduced densities as the linear algebraic versions of **marginal probability** distributions.

Diagonals recover marginal probability distributions.

$$(M^\dagger M)_{xx} = \sum_y \sqrt{\pi(x,y)\pi(x,y)} = \pi_X(x)$$

Off-diagonals know about subsystem interactions.

$$(M^\dagger M)_{xx'} = \sum_y \sqrt{\pi(x,y)\pi(x',y)}$$

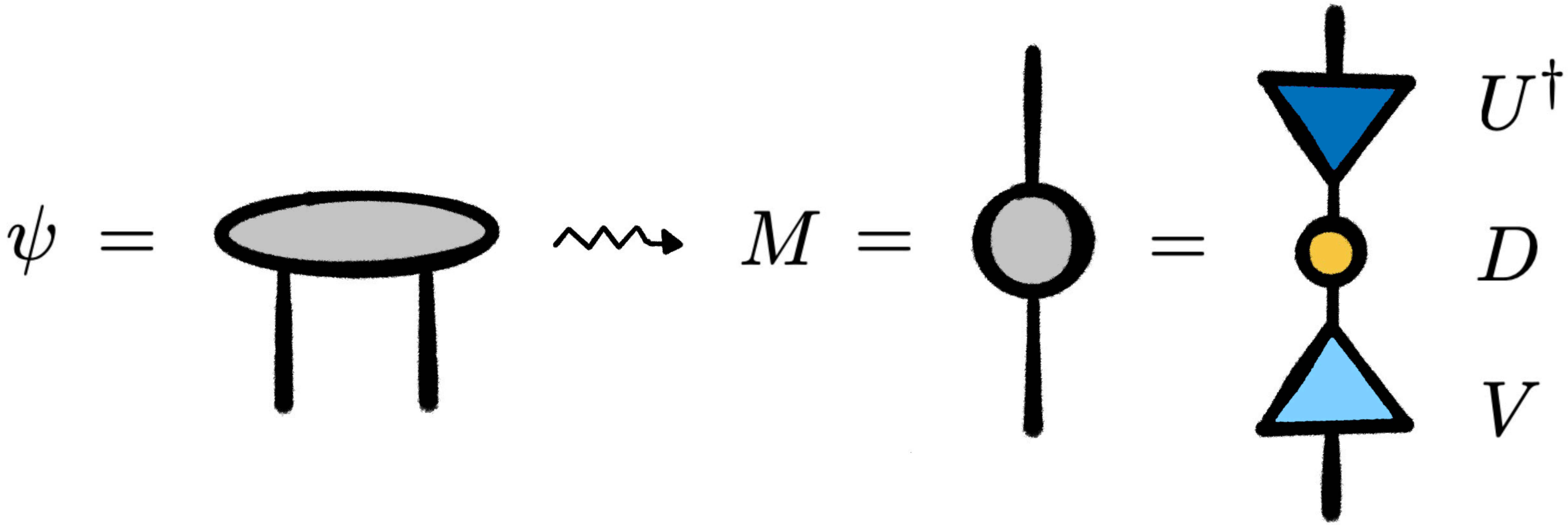
The extra information stored in the off-diagonals of $MM^\dagger$ and $M^\dagger M$ is akin to conditional probability.

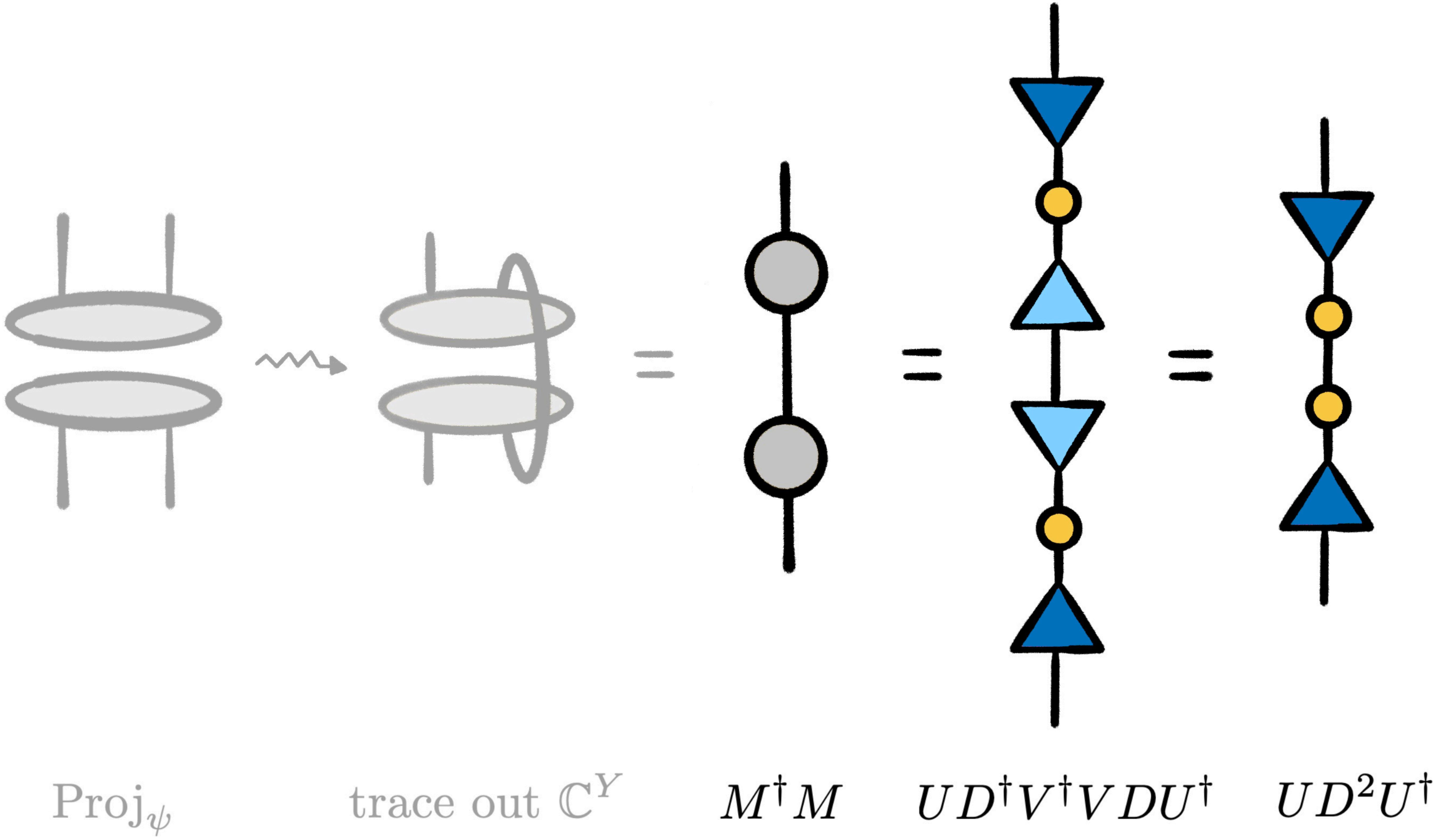
Proposition

Let $\psi \in \mathbb{C}^X \otimes \mathbb{C}^Y$ be any unit vector and let $M\colon \mathbb{C}^X \to \mathbb{C}^Y$ be the linear map associated to ψ.

- The operators $M^\dagger M$ and $MM^\dagger$ have the **same spectrum**.
- There is a **bijection** between their eigenvectors.

Proof





Proj_ψ
trace out $\mathbb{C}^Y$
$M^\dagger M$
$UD^\dagger V^\dagger VDU^\dagger$
$UD^2U^\dagger$

The two operators have the same spectrum,

$$M^\dagger M = U D^2 U^\dagger$$

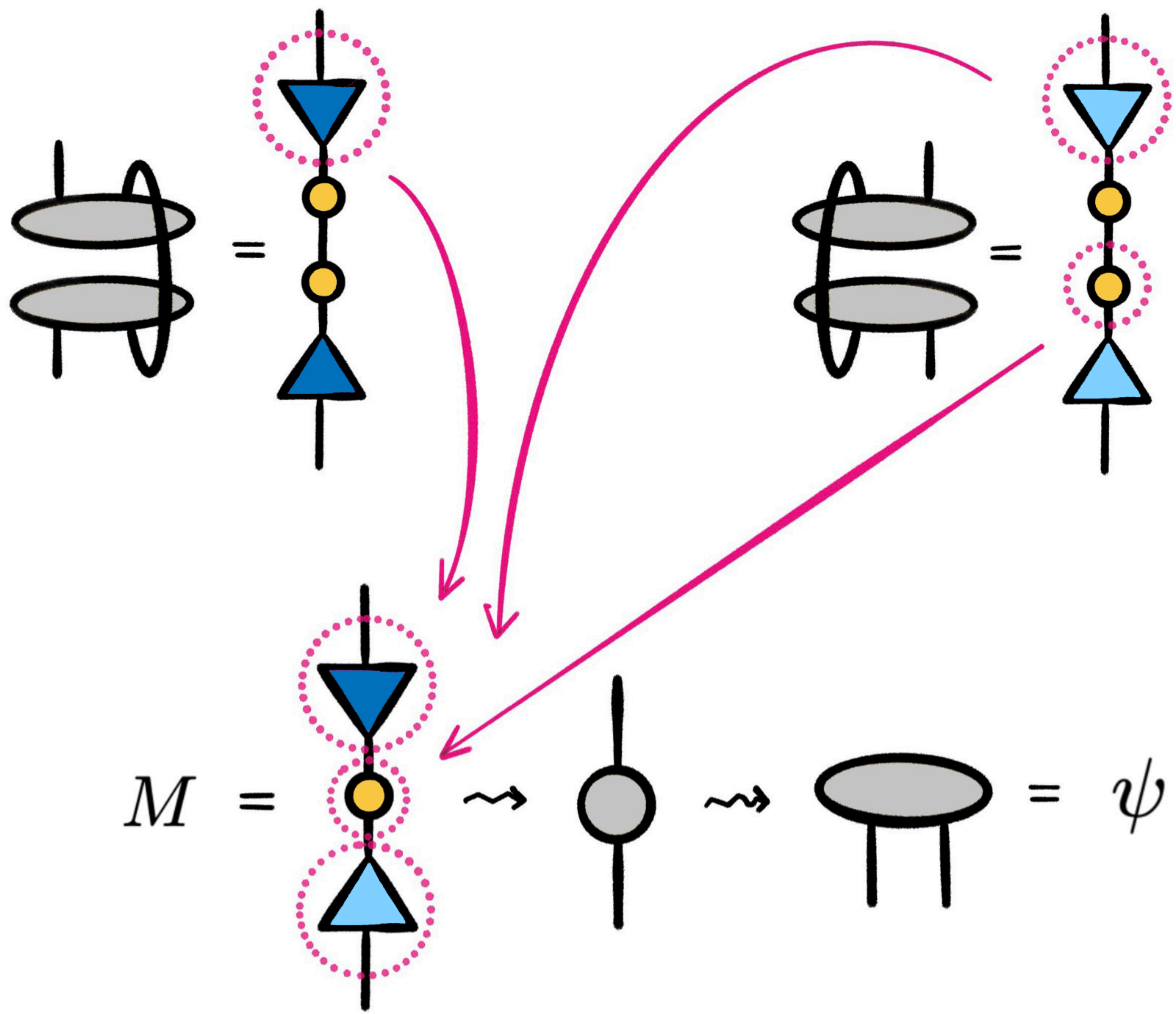
$$M M^\dagger = V D^2 V^\dagger$$

and there is a bijection between their eigenvectors.

$$u_i \quad \overset{M}{\longleftrightarrow} \quad v_i$$

The "extra information" is akin to conditional probability.

$$\pi(x, y) = \pi(y|x)\pi_X(x)$$


M
ψ

Why bother?

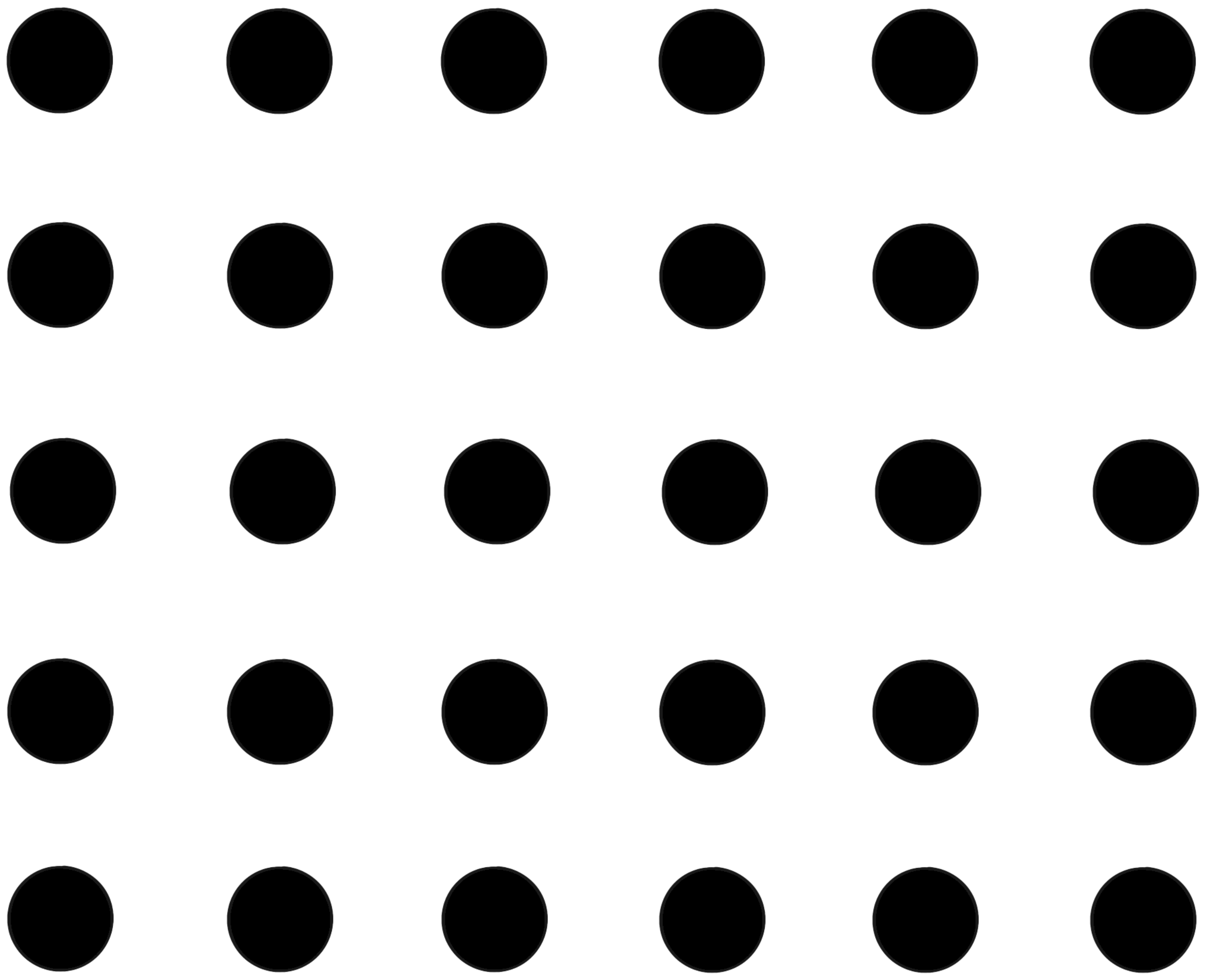
This suggests a new algorithm for reconstructing a joint probability distribution given some samples.

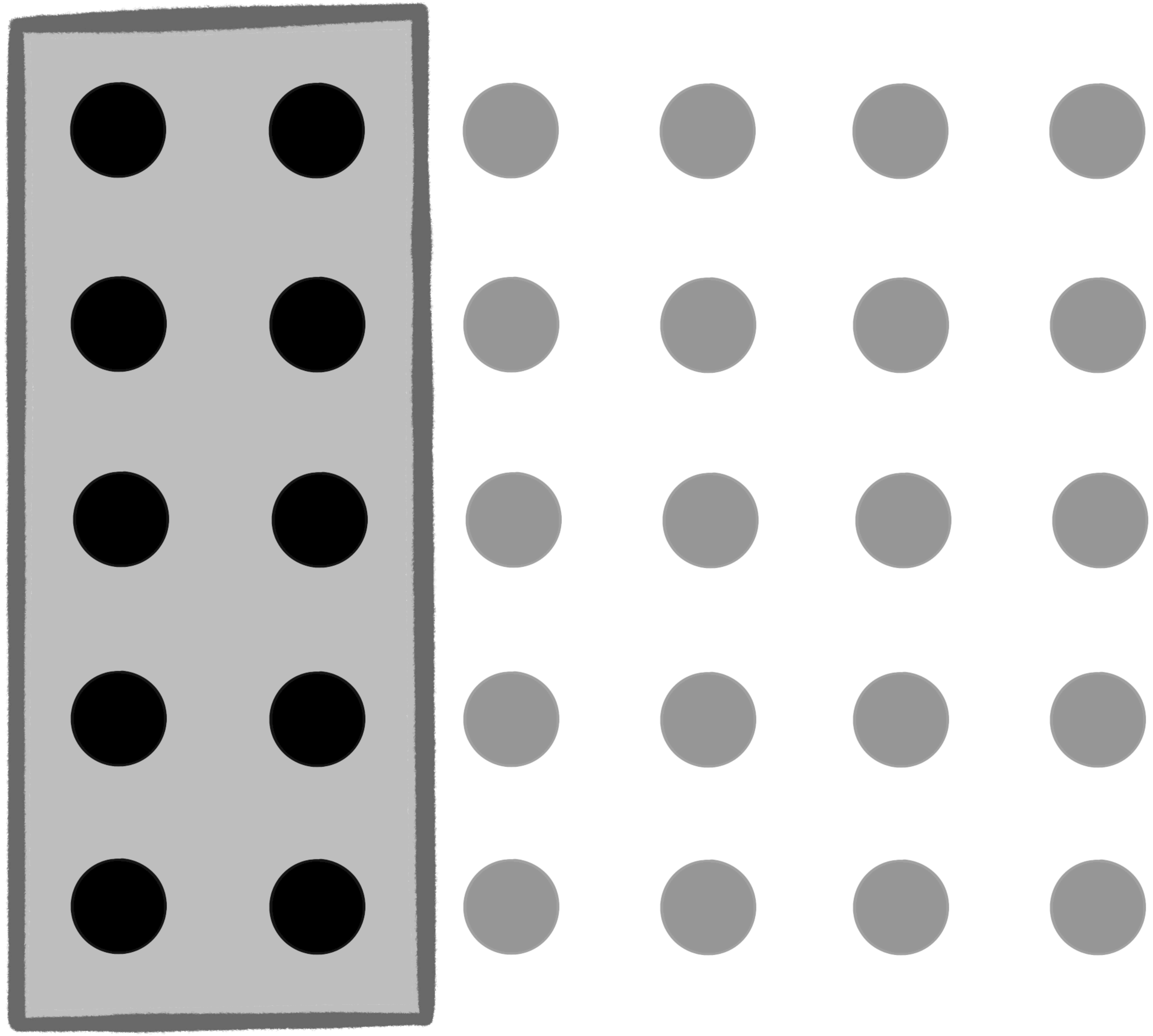
How?

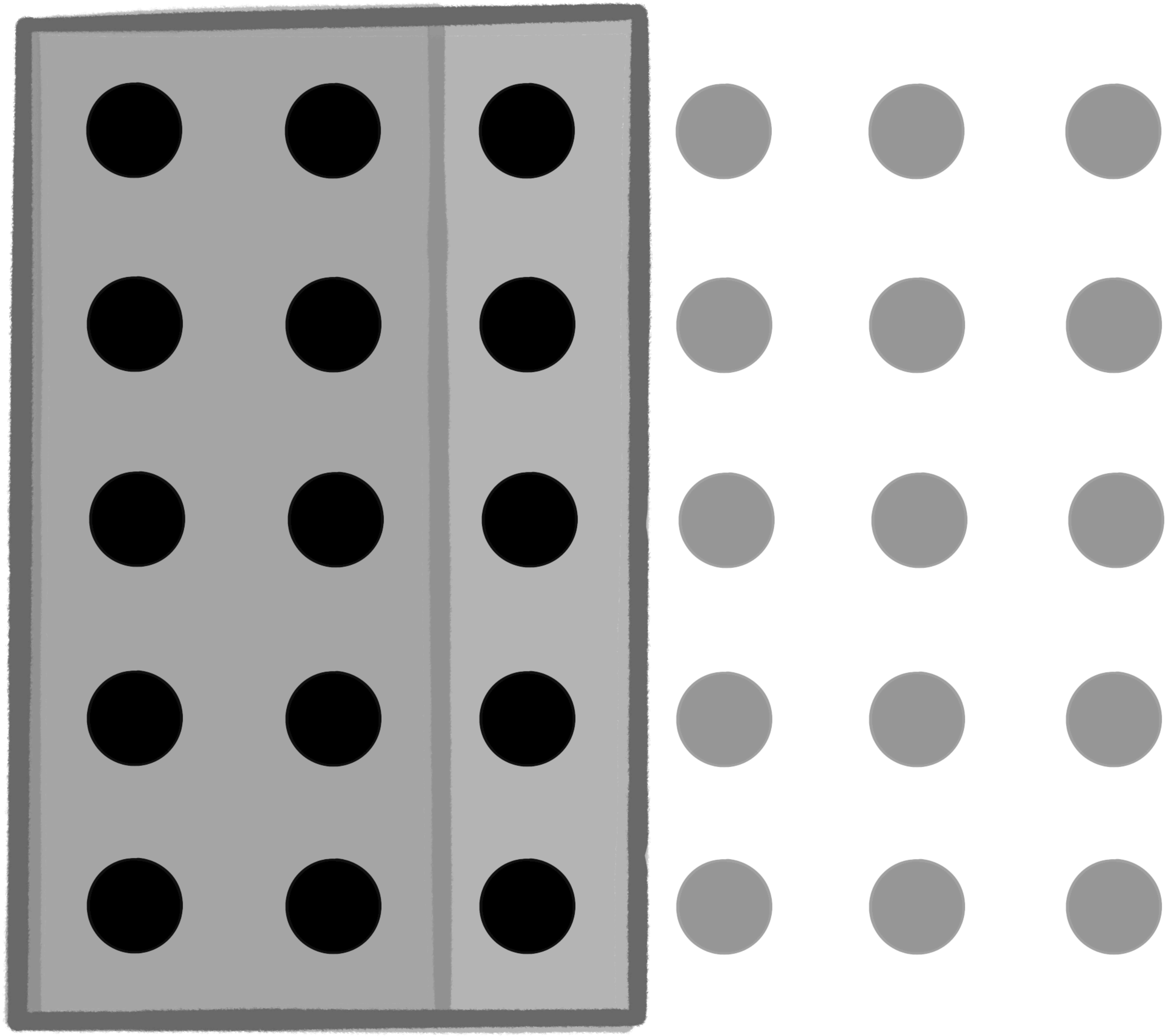
First, use samples to form the orthogonal projection operator (a rank 1 density).

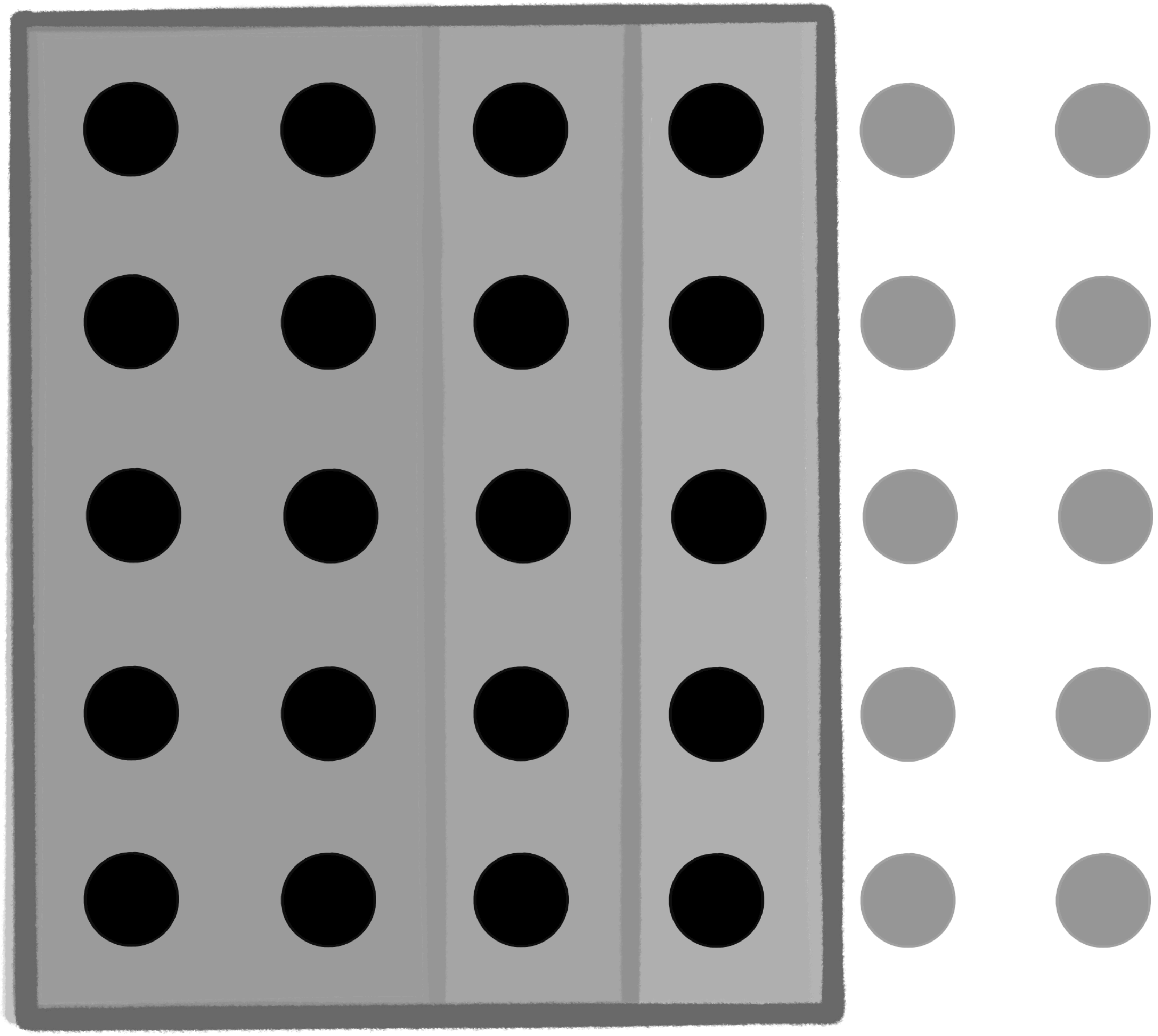
Then find reduced densities on *small* subsystems, and piece their eigenvectors together.

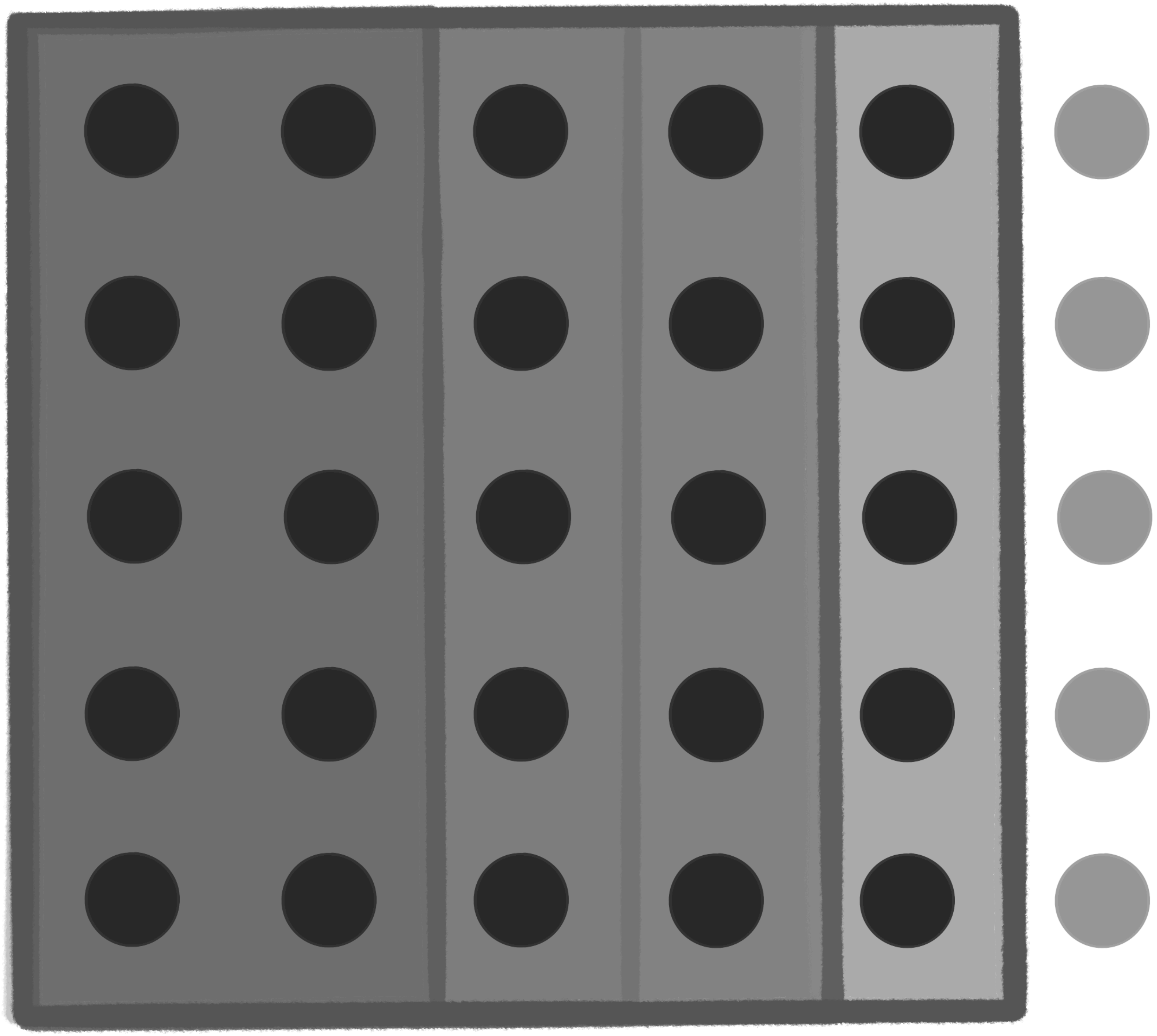
Here's the main idea....

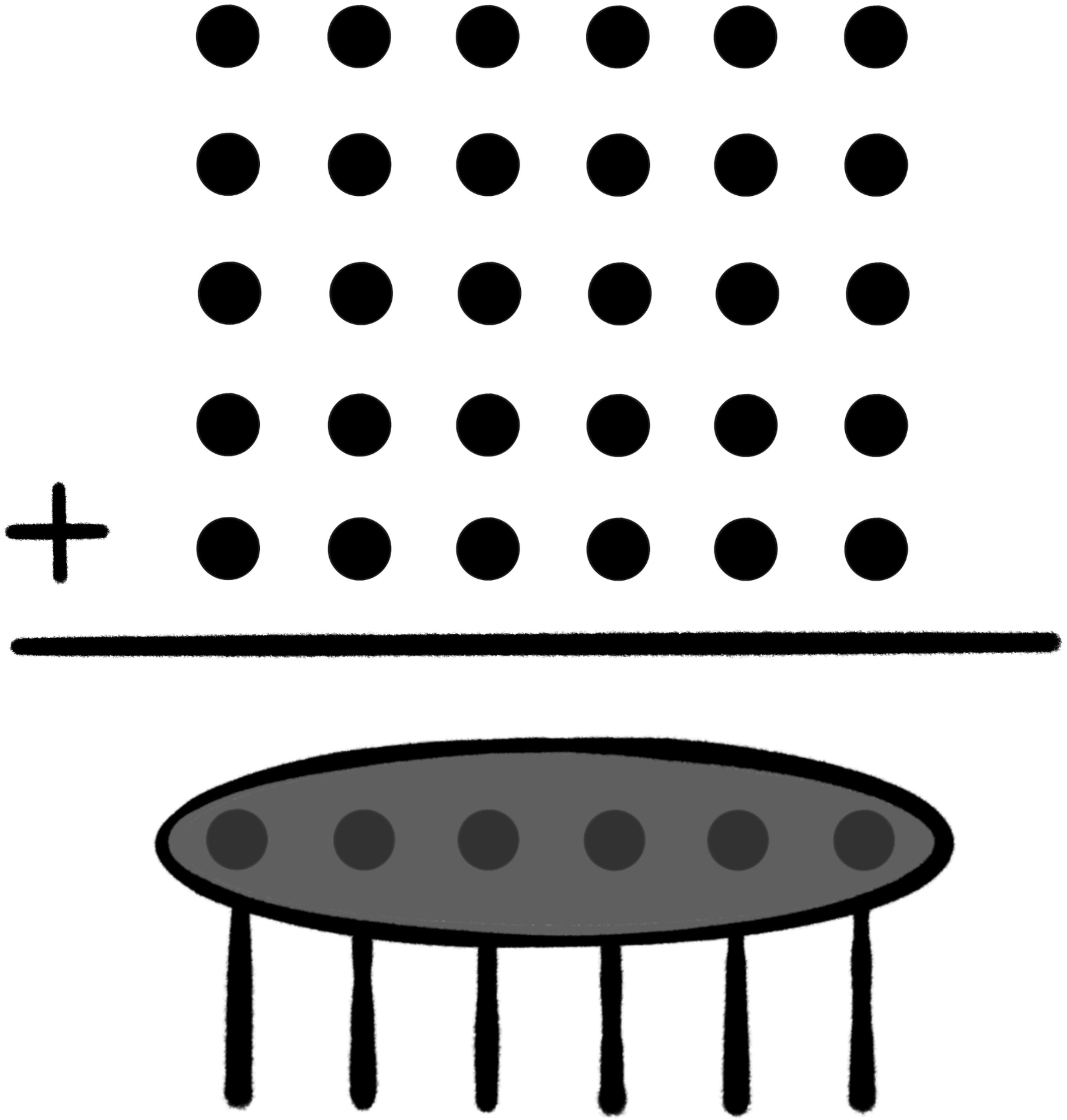


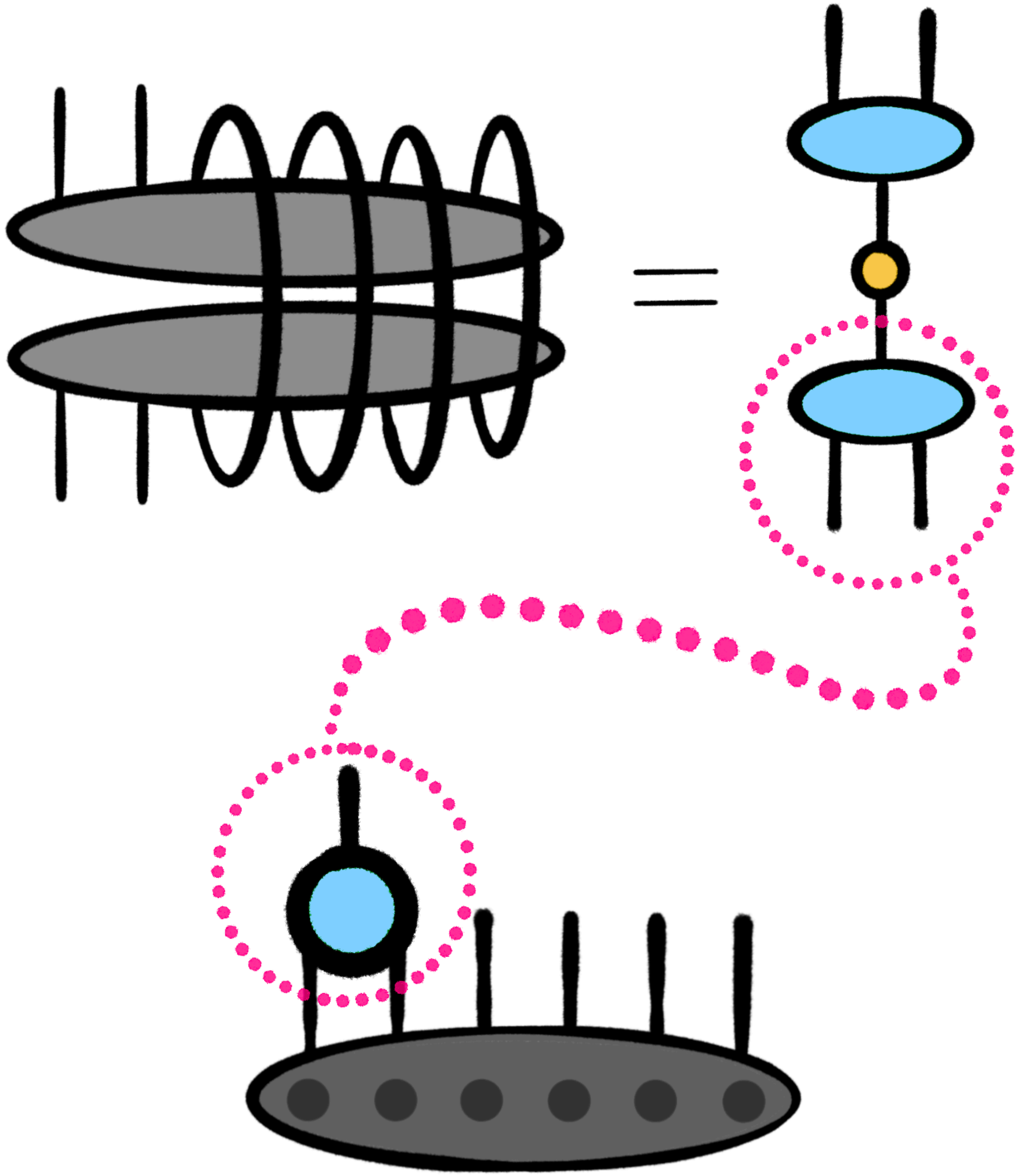


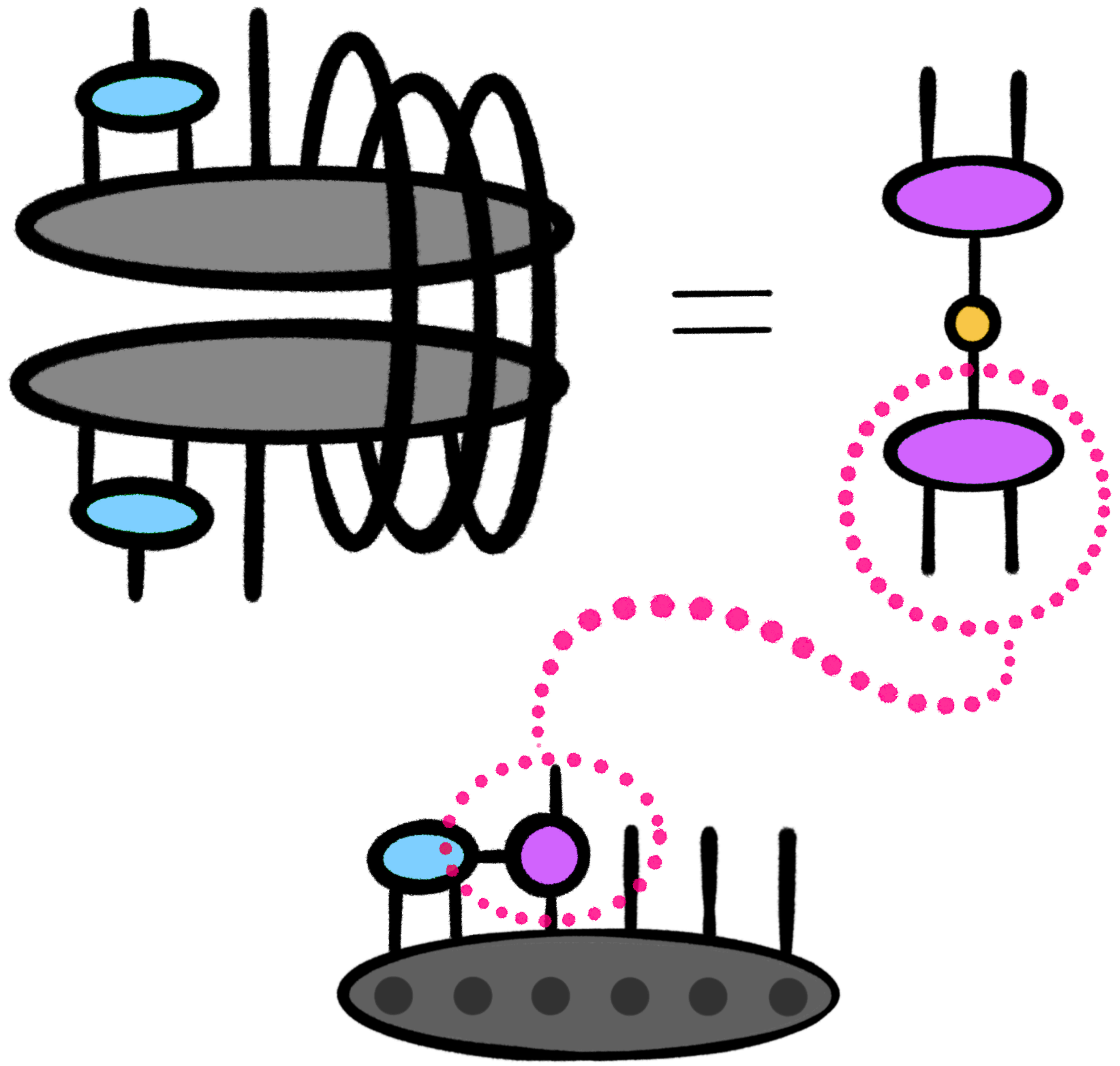






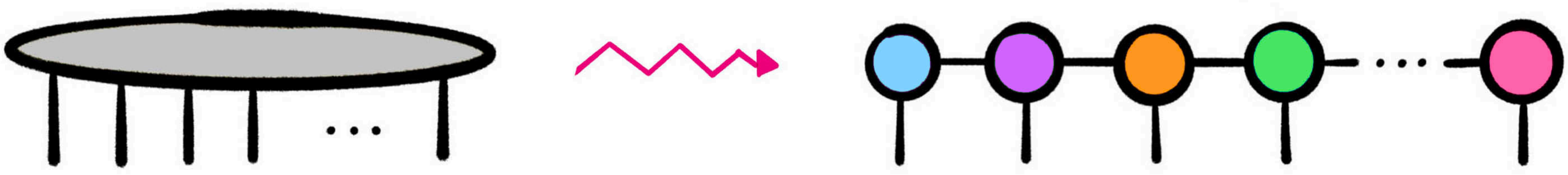






In the context of machine learning...

This procedure learns a famously difficult joint probability distribution *very efficiently*![2]



[2] T.-D. B., E. M. Stoudenmire, and J. Terilla. Modeling Sequences with Quantum States: A Look Under the Hood. *Machine Learning: Science and Technology*, 2020.

What if we replace *probabilities* with possibilities?

So far:

We started with a matrix

$$M\colon X \times Y \to \mathbb{C}$$

and considered the one-dimensional **invariant subspaces** of the linear maps $M^\dagger M$ and $MM^\dagger$.

$$M\colon \mathbb{C}^X \leftrightarrows \mathbb{C}^Y : M^\dagger$$

Instead, let's try this:

Start with a matrix

$$R\colon X \times Y \to \{0,1\}$$

and consider the **invariant subsets** of the poset maps gf and fg.

$$f\colon 2^X \leftrightarrows 2^Y : g$$

These maps form an adjunction.

For all $A \subseteq X$ and $B \subseteq Y$,

$$A \subseteq g(B) \text{ if and only if } B \subseteq f(A).$$

There are some "ops" involved in the details (omitted)...

Behind the scenes: free (co)completions.

$$\mathsf{Set}^{\mathsf{C}^{\mathrm{op}}} \leftrightarrows (\mathsf{Set}^{\mathsf{D}})^{\mathrm{op}} \quad \rightsquigarrow \quad 2^{X^{\mathrm{op}}} \leftrightarrows (2^{Y})^{\mathrm{op}}$$

Exchanging Set for *truth values* leads to formal concepts.

A **formal concept** is a pair $A \subseteq X$ and $B \subseteq Y$ such that

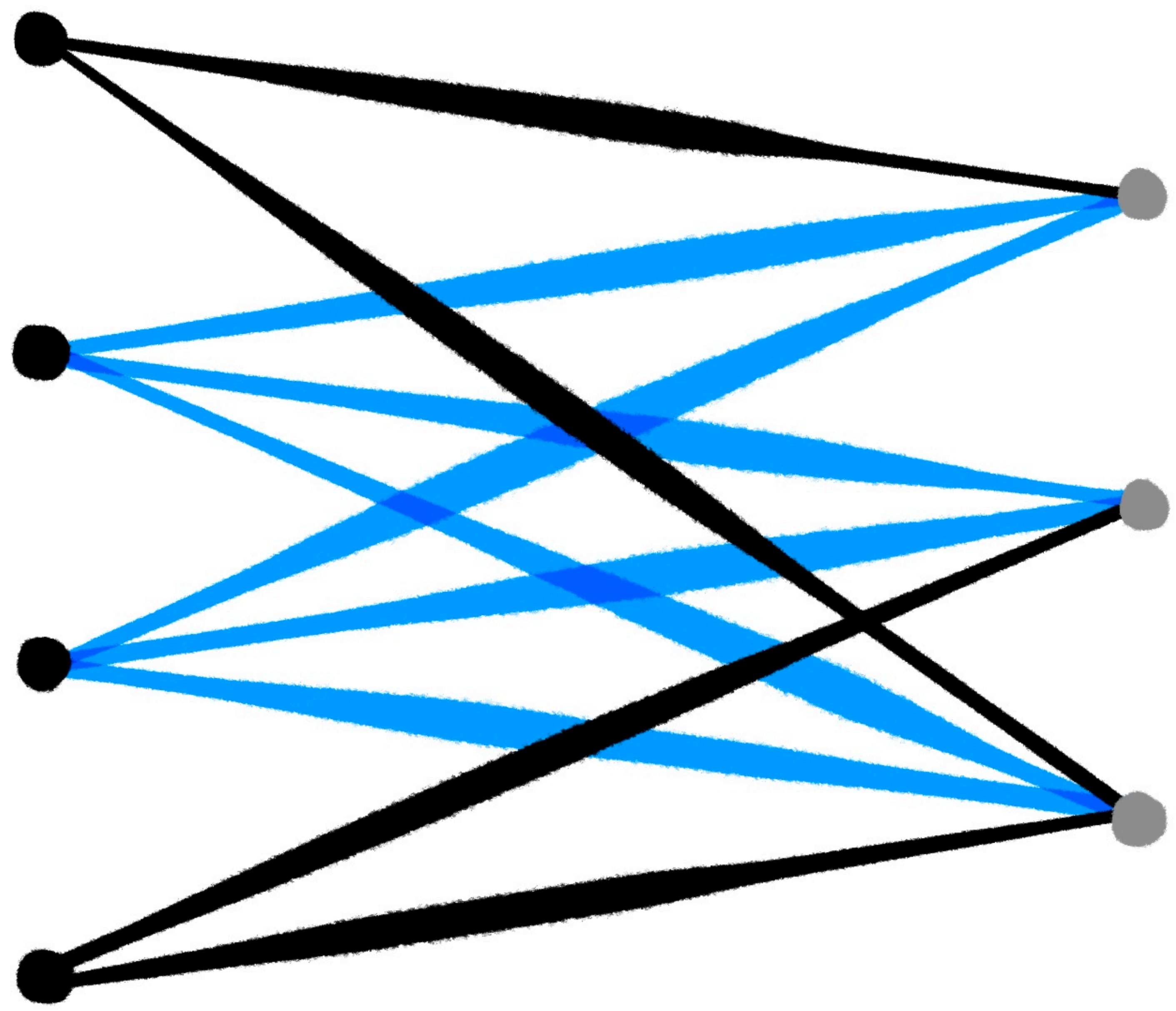
$$f(A) = B$$

$$g(B) = A$$

Formal concepts coincide with **invariant subsets** of the compositions gf and fg.

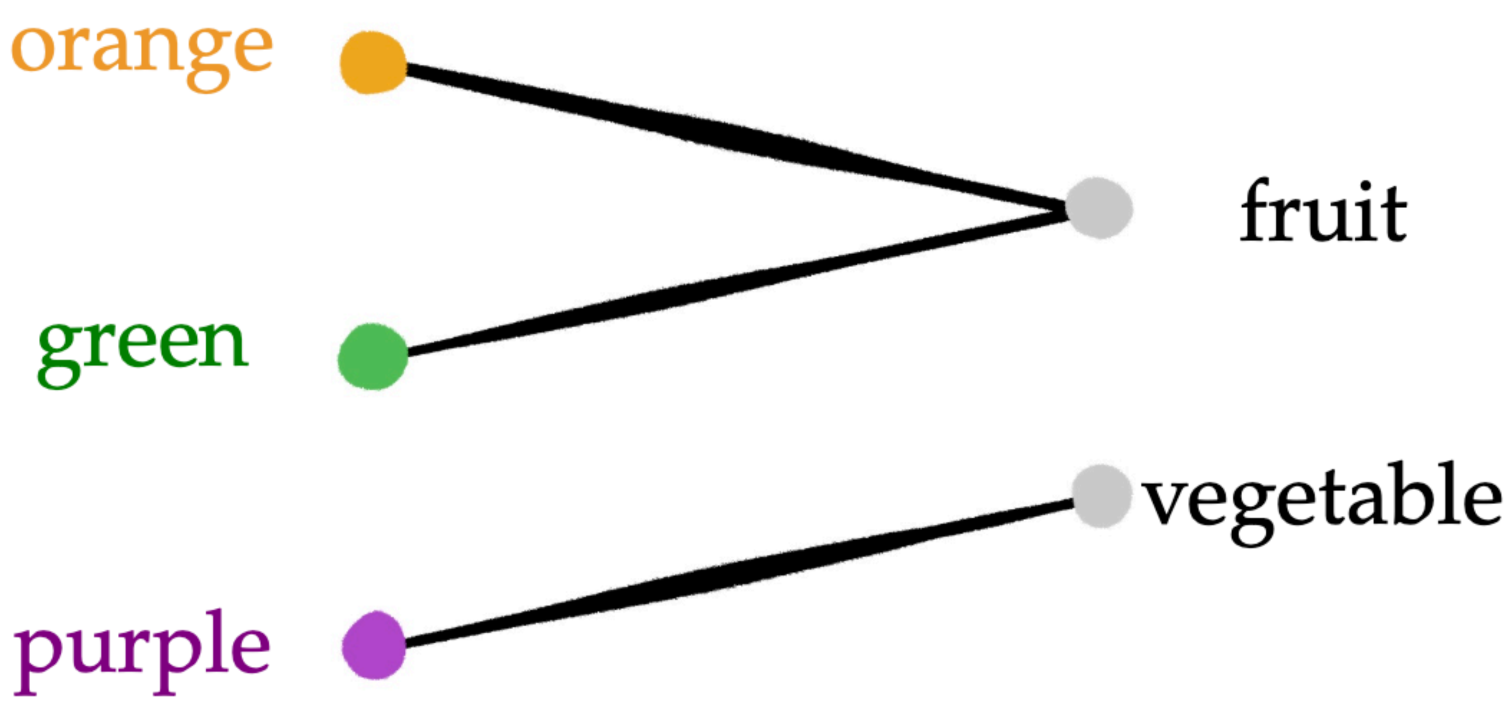
$$gf(A) = A \qquad fg(B) = B$$

Formal concepts also coincide with maximal **complete bipartite subgraphs** of a bipartite graph.



Example

This relation has two formal concepts, which coincide with the **two eigenvectors** of $M^\dagger M$ and $MM^\dagger$ in the opening example.



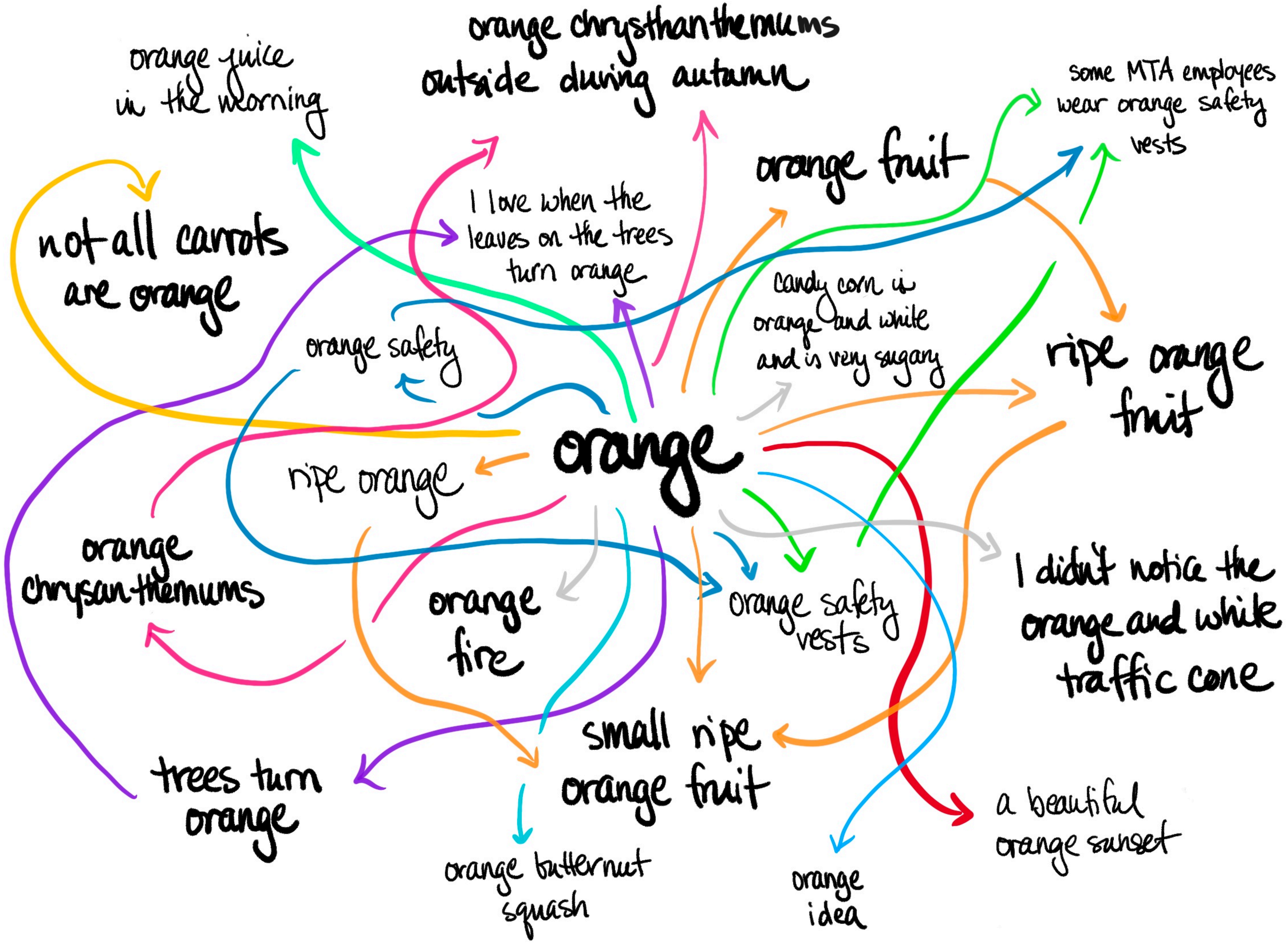
By way of analogy:

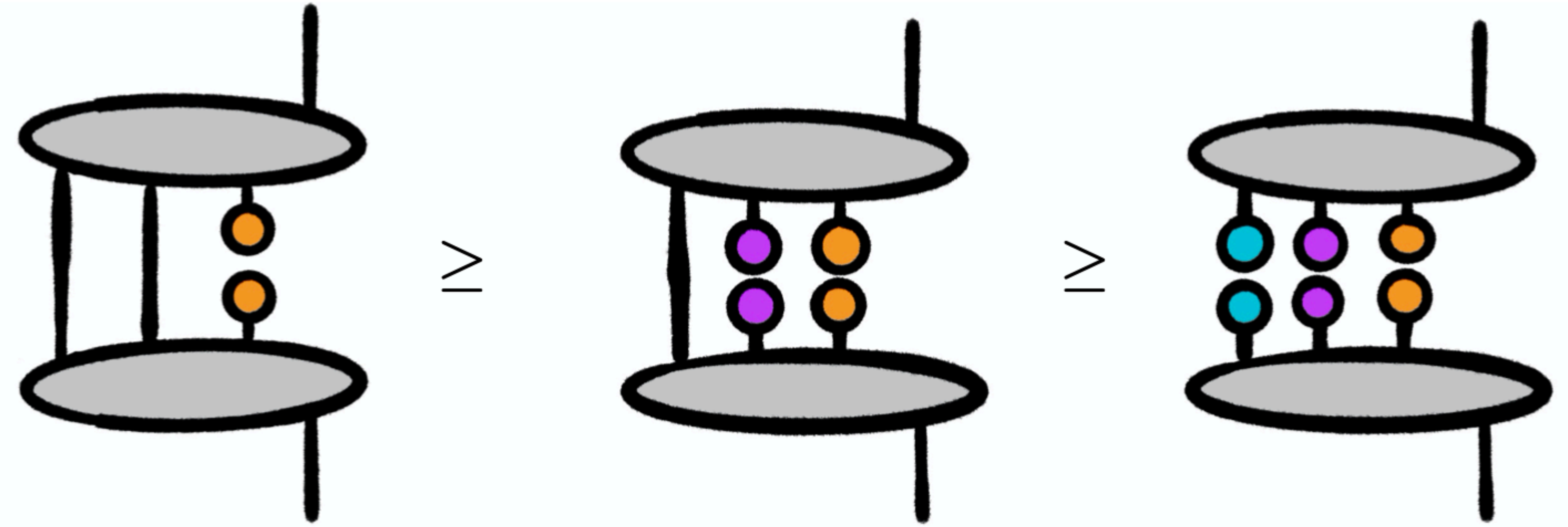
Exchanging *truth values* for (square roots of) probabilities leads to something new.

But wait! There's more.

Natural language exhibits mathematical structure that is both algebraic and statistical.

We can explore this with the ideas here.


orange
orange juice in the morning
orange chrysthanthemums outside during autumn
some MTA employees wear orange safety vests
not all carrots are orange
I love when the leaves on the trees turn orange
orange fruit
candy corn is orange and white and is very sugary
orange safety
ripe orange fruit
ripe orange
orange chrysanthemums
orange fire
orange safety vests
I didn't notice the orange and white traffic cone
small ripe orange fruit
trees turn orange
orange butternut squash
orange idea
a beautiful orange sunset



ρ_{orange} $\rho_{\text{ripe orange}}$ $\rho_{\text{small ripe orange}}$