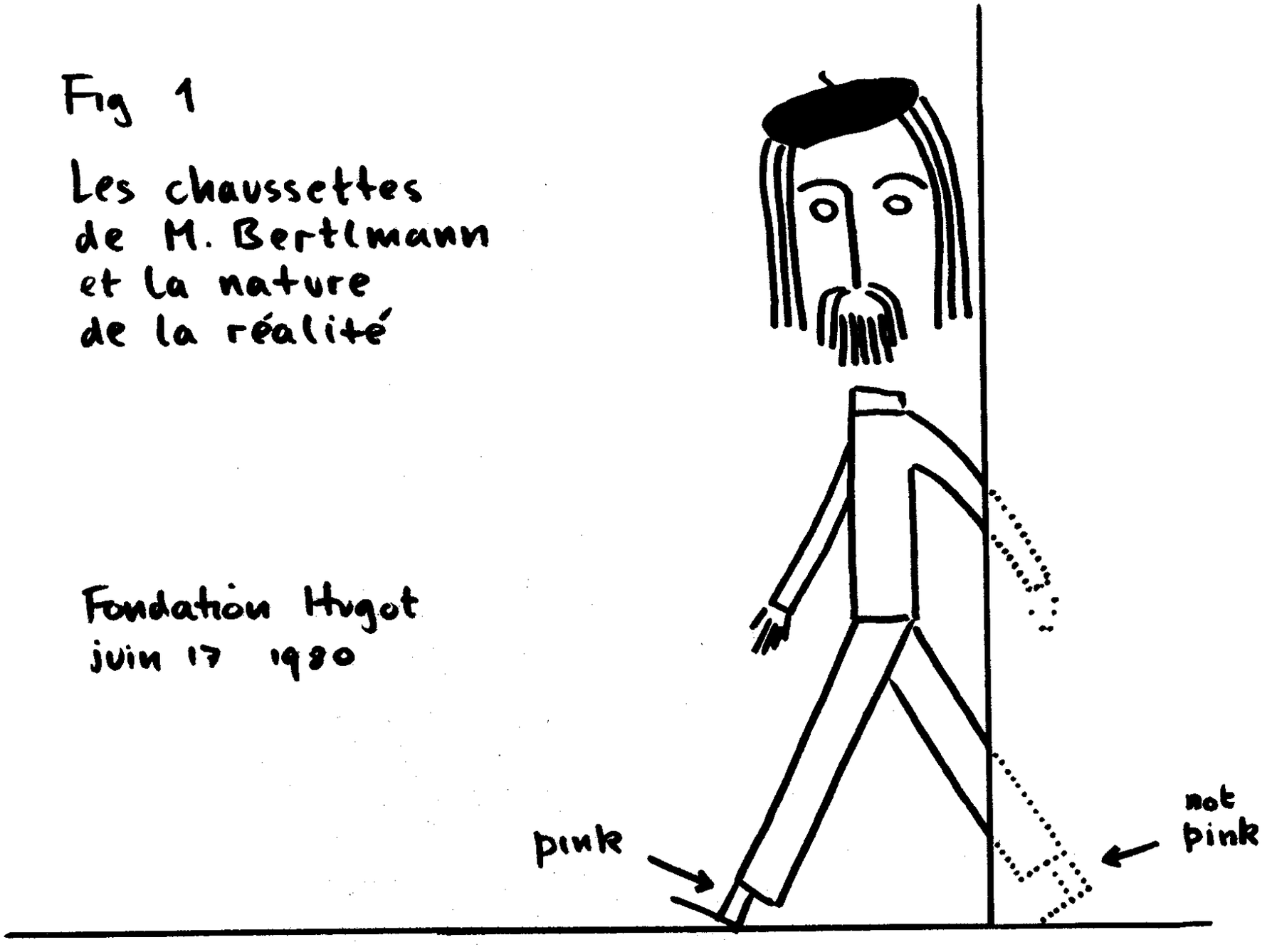
Fig 1

Les chaussettes
de M. Bertlmann
et la nature
de la réalité

Fondation Hugot
juin 17 1980



Quantum Probability

To study a physical system, we will think about

- the set of possible states of the system, and
- the set of possible observables for the system.

By thinking about the algebraic properties of the observables, we will be led to a formalism which includes both classical probability and the basic rules of quantum mechanics as special cases.

Let's suppose for some system we have a set A of observables ("things we might chose to observe"), each of which when measured produces a real number.

- How tall are you?
- How many classes are you taking this semester?
- Do you like chocolate cake? - Answer 1 if yes, 0 if no

What structure does the set have?

First, there are some very boring observables, which don't even look at the state of the system, and merely output some fixed real number.

Thus we have $\mathbb{R} \subset \mathcal{A}$.

Next, let's define a multiplication on $\mathcal{A}$.

Given two observables X and Y, you could measure X, write down (but don't tell us) the answer, subsequently measure Y, and then report the product of the answers.

Sometimes this gives silly answers —
when you enter a country town and see the sign "Nil Desperandum – Population 17 Elevation 28m" you're allowed to observe that it has 476 people metres.

We're also not insisting that $XY = YX$; measuring X first could have an effect on the system ("Do you like this chocolate cake?") that changes the results of measuring Y ("How much do you weigh?").

Because we've already got the scalars as the 'constant observables', this gives us a scalar multiplication. If $X \in A$, so is $2X$ and $7X$.

Finally, you should think about why this operation is associative: $(XY)Z = X(YZ)$.

Next we're going to define addition and more generally linear combinations, but we'll start with the special case of convex linear combinations $\sum_i \alpha_i X_i$, where $\sum \alpha_i = 1$, and in particular $\frac{X+Y}{2}$.

To observe $\frac{X+Y}{2}$, you should take a (fair) coin, toss it, and if it comes up heads, measure and report the value of X, while if it comes up tails, measure and report the value of Y.

(Without, notice, reporting to the outside world which you did!)

More generally to observe $\sum_{i \in I} \alpha_i X_i$ choose an element i from the set I with probability α_i, and then measure the corresponding X_i.

Finally, addition is defined using scalar multiplication and convex combination:

$$X + Y = 2 \cdot \left(\frac{X+Y}{2}\right)$$

(Exercise: this is identical to $\frac{2X+2Y}{2}$.) (Exercise: $X+Y = Y+X$)

Now we'd like to check two things about these operations.

① Multiplication distributes over addition

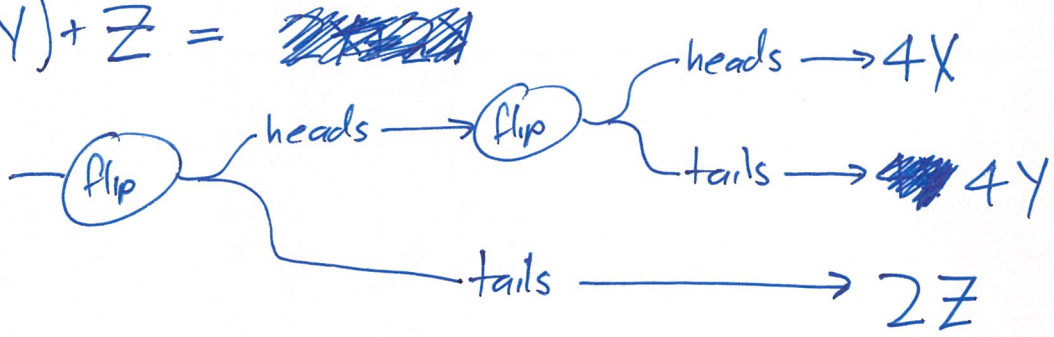
$$X(Y+Z) = XY + XZ$$

This is pretty straightforward.

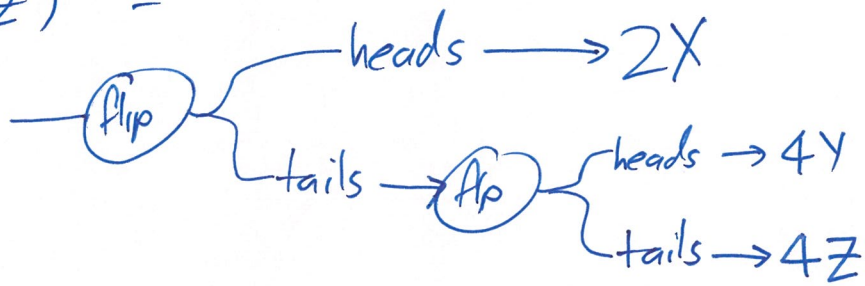
② Addition is associative!

This seems more subtle at first:

$(X+Y)+Z =$



$X+(Y+Z) =$



These seem different!

However, with the system in any given state, these measurements certainly have the same expected value.

It seems pretty reasonable to use this as the definition of two observables being equal!

(It certainly works out in the classical probability world...)

At this point we've shown that the set of observables A naturally has the structure of an <u>associative algebra</u>.

We'll soon turn to thinking about states, but let's first remark on some examples.

① For any set $\mathcal{X}$, $\mathbb{C}^{\mathcal{X}} = \{\text{functions from } \mathcal{X} \text{ to } \mathbb{C}\}$ is an algebra with pointwise multiplication and addition.

In particular we'll look at $\mathbb{C}^2 = \left\{\begin{pmatrix} z \\ w \end{pmatrix}\right\}$ in detail later.

② $M_n(\mathbb{C}) = \{n \times n \text{ complex matrices}\}$ with matrix multiplication and addition

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} w & x \\ y & z \end{pmatrix} = \begin{pmatrix} aw+by & ax+bz \\ cw+dy & cx+dz \end{pmatrix}$$

This will be our first example in quantum probability

③ $\mathbb{C}[x]/x^2=0$ is an associative algebra, but we'll later add a bit more structure that rules it out!

It might seem intuitive to think of observables as functions on some collection of configurations.

We'll prefer to think of this the other way round: states will be functions on the observables!

A state ρ of the system will be a linear function

$$\rho : \mathcal{A} \longrightarrow \mathbb{C}$$

such that $\rho(1) = 1$ (remember the observable 1 always reports 1!)

and $\rho(X) \geq 0$ for each "positive" observable X.

We'll interpret this function as giving the expected value of measuring X, if the system is in the state ρ.

(Aside: You might ask if we intended 'expected value' in the frequentist or Bayesian sense. Hopefully it will become clear that only the Bayesian interpretation is compatible with our formalism.)

What are positive observables?

Roughly they are observables that always give non-negative answers.

We will formalize this in the following way:

Our algebra A of observables should have a "$*$-operation", $*: A \to A$, such that

$(AB)^* = B^* A^*$, $A^{**} = A$

and $(zA)^* = \bar{z} A^*$.

A positive observable is then one of the form $X^* X$, for some other observable X.

(Notice this is just like in the complex numbers: a complex number z is a positive real number if and only if $z = \bar{w} w$ for some $w \in \mathbb{C}$.)

I don't know a good heuristic explanation for this - it's just something that seems to be true about the world.

Let's look at some examples.

① $A = \mathbb{C}^2$, with $\begin{pmatrix} z \\ w \end{pmatrix}^* = \begin{pmatrix} \bar{z} \\ \bar{w} \end{pmatrix}$

The identity of A is $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$.

A state $p : \mathbb{C}^2 \to \mathbb{C}$ must be of the form

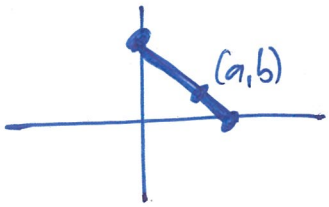
$$p\begin{pmatrix} z \\ w \end{pmatrix} = az + bw$$

and then $p\begin{pmatrix} 1 \\ 1 \end{pmatrix} = 1 \Rightarrow a + b = 1$.

Moreover, clearly $\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}^2$, $\begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}^2$ are positive ~~operators~~ observables, so

$$p\begin{pmatrix} 1 \\ 0 \end{pmatrix} = a \geq 0, \text{ and } p\begin{pmatrix} 0 \\ 1 \end{pmatrix} = b \geq 0.$$

These conditions completely characterise the states:

$$S = \left\{ \right.$$ 

$$\left. \right\}$$

What is this system?

A 2-configuration classical system,
for example a coin.

A "heads-up" coin is the state $\rho\begin{pmatrix} z \\ w \end{pmatrix} = z$,

a "tails-up" coin is the state $\rho\begin{pmatrix} z \\ w \end{pmatrix} = w$,

while a coin we've just flipped, and not looked at yet,
is the state $\rho\begin{pmatrix} z \\ w \end{pmatrix} = \frac{z+w}{2}$.

Notice that "state" here <u>does not</u> refer to some
absolute, real-world, how-the-coin-actually-is
description of the system,
but merely to <u>our knowledge</u> of the system.
It's important here we don't get too hung up on
assuming there really is a
"how-the-coin-actually-is" description available.

We'll return to this system in a moment to talk
about measurement.

(2) $A = M_2(\mathbb{C})$, with $\begin{pmatrix} w & x \\ y & z \end{pmatrix}^* = \begin{pmatrix} \bar{w} & \bar{y} \\ \bar{x} & \bar{z} \end{pmatrix}$.

The identity of A is $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

A state $\rho : M_2(\mathbb{C}) \to \mathbb{C}$ must be of the form

$$\rho\begin{pmatrix} w & x \\ y & z \end{pmatrix} = aw + bx + cy + dz$$

for some $a, b, c, d \in \mathbb{C}$.

$$\rho\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = 1 \Rightarrow a + d = 1.$$

Let's introduce α so $a = \frac{1+\alpha}{2}$, $d = \frac{1-\alpha}{2}$, with $\alpha \in [-1, 1]$.

(Exercise:) A general positive element of A is of the form $\begin{pmatrix} w & \chi \\ \bar{\chi} & \xi \end{pmatrix}$ with $w, \xi \in \mathbb{R}_{\geq 0}$, $\chi \in \mathbb{C}$,

and $w\xi \geq \chi\bar{\chi}$.

Then $\rho\begin{pmatrix} w & \chi \\ \bar{\chi} & \xi \end{pmatrix} = \frac{1+\alpha}{2} w + \frac{1-\alpha}{2} \xi + b\chi + c\bar{\chi}$.

This must be real for all $\chi \in \mathbb{C}$, so we must have

$$c = \bar{b}.$$

Now $\min_{\chi : |\chi|^2 \leq w\xi} b\chi + \overline{b\chi} = -2|b|\sqrt{w\xi}$,

so we must have $\frac{1+\alpha}{2} w + \frac{1-\alpha}{2} \xi - 2|b|\sqrt{w\xi} \geq 0$

Minimising this with respect to $s = \omega^{\frac{1}{2}} \xi^{-\frac{1}{2}}$, we find

the minimum occurs at $s = \frac{2|b|}{1+\alpha}$.

The minimum still has to be positive, and this says

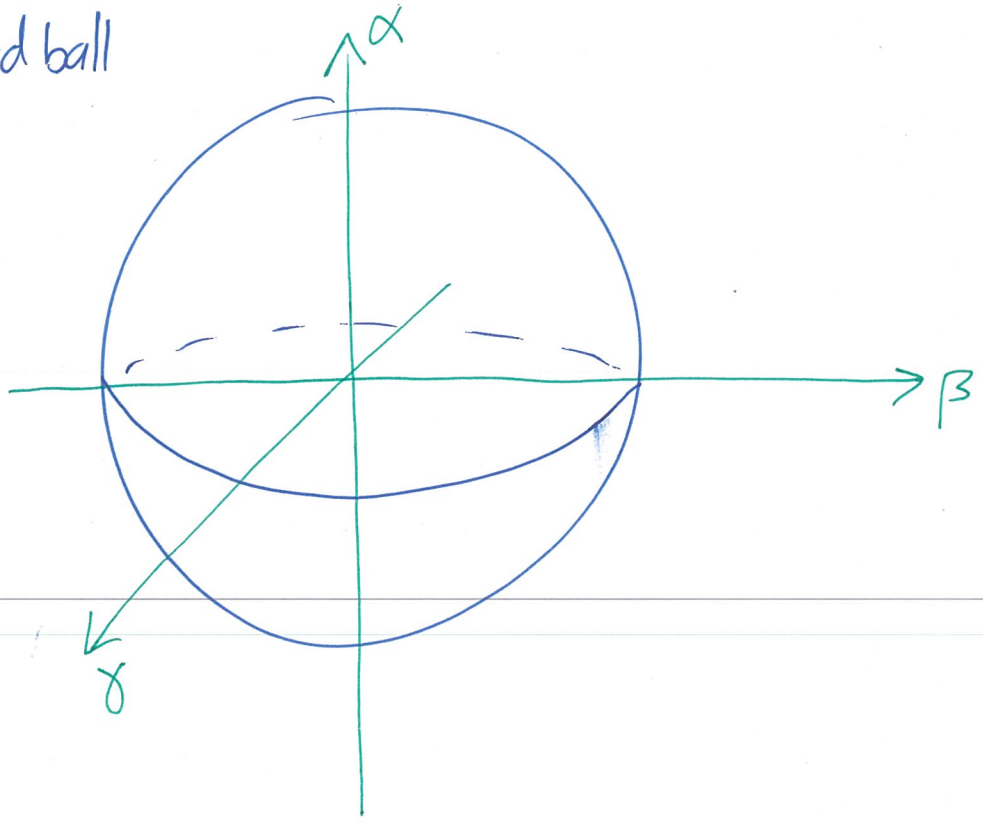
$$1 \geq \alpha^2 + 4|b|^2.$$

Writing $b = \frac{\beta + i\gamma}{2}$, $c = \frac{\beta - i\gamma}{2}$, we have that a general state is of the form

$$\rho\begin{pmatrix} w & x \\ y & z \end{pmatrix} = \frac{1+\alpha}{2} w + \frac{\beta+i\gamma}{2} x + \frac{\beta-i\gamma}{2} y + \frac{1-\alpha}{2} z$$

with $\alpha^2 + \beta^2 + \gamma^2 \leq 1$, $\alpha, \beta, \gamma \in \mathbb{R}$.

We can draw the set of states (or "Bloch region") as the solid ball



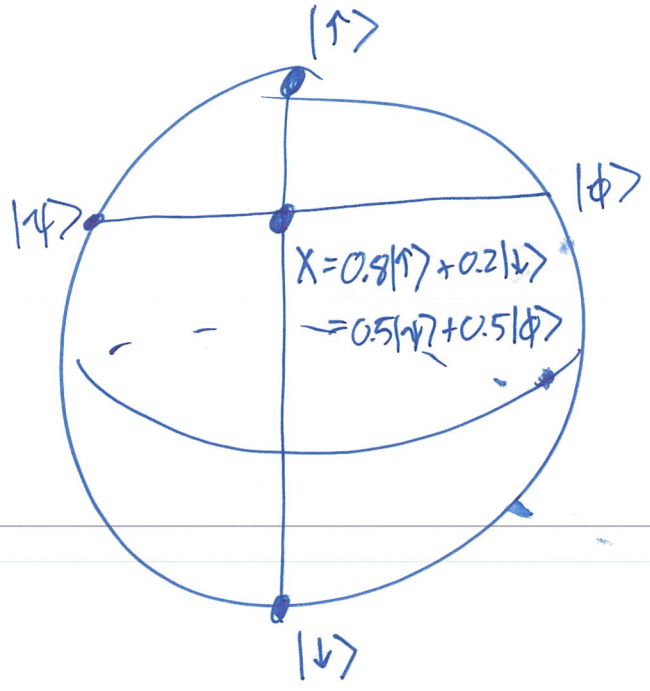
As in our classical system we have some 'extremal' states, which are not convex linear combinations of the states.

Here they are $\{(\alpha,\beta,\gamma) \mid \alpha^2+\beta^2+\gamma^2=1\}$, while for $\mathbb{C}^2$ they were $\begin{pmatrix}1\\0\end{pmatrix}$ and $\begin{pmatrix}0\\1\end{pmatrix}$.

Notice that 'classically' any non-extremal state could be written uniquely as a convex linear combination of extremal states

(that is, every state was of the form:
"with probability p, heads, and with probability $1-p$, tails"),

but this is not the case for $M_2(\mathbb{C})$.



Any rank 1 projection P (that is, an orthogonal projection onto a line) in $M_2(\mathbb{C})$ gives rise to a state as follows —

$$\rho_P(X) = \mathrm{tr}(PX)$$

(Note $\rho_P(1) = \mathrm{tr}(P) = 1$

$$\rho_P(X^*X) = \mathrm{tr}(PX^*X)$$
$$= \mathrm{tr}(PP^*X^*X)$$
$$= \mathrm{tr}(P^*X^*XP)$$
$$= \mathrm{tr}((XP)^*XP)) \geq 0.)$$

In fact, these are precisely the extremal states for $M_2(\mathbb{C})$.

$$\left\{\begin{array}{c}\text{Extremal states} \\ \text{for } M_2(\mathbb{C})\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{lines in} \\ \mathbb{C}^2\end{array}\right\} \longleftrightarrow \left\{\begin{array}{c}\text{points on} \\ S^2\end{array}\right\}.$$

In fact $M_2(\mathbb{C})$ is the right algebra of observables for a "spin $\frac{1}{2}$ particle", a quantum mechanical system which is spinning in some direction, but has no other properties.

State-operator correspondence

Any linear functional $\rho: M_n(\mathbb{C}) \to \mathbb{C}$ can be expressed in the form $\rho(X) = \mathrm{tr}(RX)$ for some $R \in M_n(\mathbb{C})$.

(Example: $\mathrm{tr}\left(\begin{pmatrix} a & c \\ b & d \end{pmatrix}\begin{pmatrix} w & x \\ y & z \end{pmatrix}\right) = \mathrm{tr}\begin{pmatrix} aw+cy & ax+cz \\ bw+dy & bx+dz \end{pmatrix}$

$= aw + bx + cy + dz.$)

The linear functional is a state

$\iff \mathrm{tr}(R) = 1$, and R is a positive element of $M_n(\mathbb{C})$.

Let's now think about measurement.

We'll only, for now, describe measurement of

boolean observables,

that is observables X that only produce the results 0 and 1.
We also ask that $X^2 = X = X^*$.
Given a boolean observable X, and a state ρ,

the expected value of observing X in the state ρ

is $\rho(X)$, so we must ~~observe the~~

obtain the result 1 with probability $\rho(X)$,

and the result 0 with probability $1-\rho(X)$.

What happens to the state?

Recall that "the state" refers to our knowledge of

the system, so even in the classical setting

there's ~~no~~ every reason to expect that the state

will change, as we update our knowledge.

Supposing we measure the value 1 for a boolean observable X on a state ρ, the updated state is

$$\hat{\rho}(Y) = \frac{\rho(XYX)}{\rho(X)}.$$

(We can't be dividing by zero here — otherwise we couldn't have measured 1!)

This is a state:

$$\hat{\rho}(1) = \frac{\rho(X1X)}{\rho(X)} = \frac{\rho(X)}{\rho(X)} = 1$$

and $$\hat{\rho}(Z^*Z) = \frac{\rho(XZ^*ZX)}{\rho(X)} = \frac{\rho((ZX)^*(ZX))}{\rho(X)} \geq 0.$$

In order to justify this rule, let's see what it says in the classical setting, $A = \mathbb{C}^{\mathcal{X}}$ for some set $\mathcal{X}$.

Recall the extremal states correspond to points of $\mathcal{X}$ ("the system is definitely in the configuration $x \in \mathcal{X}$")

while a general state is some probability density on $\mathcal{X}$.

A general boolean operator in A is of the form

$$\chi_B(x) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{otherwise} \end{cases}$$

for some subset $B \subset \mathcal{X}$.

(Exercise: $\chi_B^2 = \chi_B = \chi_B^*$)

Then $p(\chi_B)$ is the probability the system is in a configuration in the subset B.

We then have, after measuring χ_B:

$$\hat{\rho}(\chi_C) = \frac{\rho(\chi_B \chi_C \chi_B)}{\rho(\chi_B)} = \frac{\rho(\chi_{B \cap C})}{\rho(\chi_B)}$$

or in words:

"the probability that the system is in a configuration in C, given that we've just observed it to be in a configuration in B, is the probability it is in both B and C, divided by the probability it is in B."

This is precisely the classical law of conditional probability, and only one step away from Bayes theorem, tell us how to update our beliefs after making an observation.

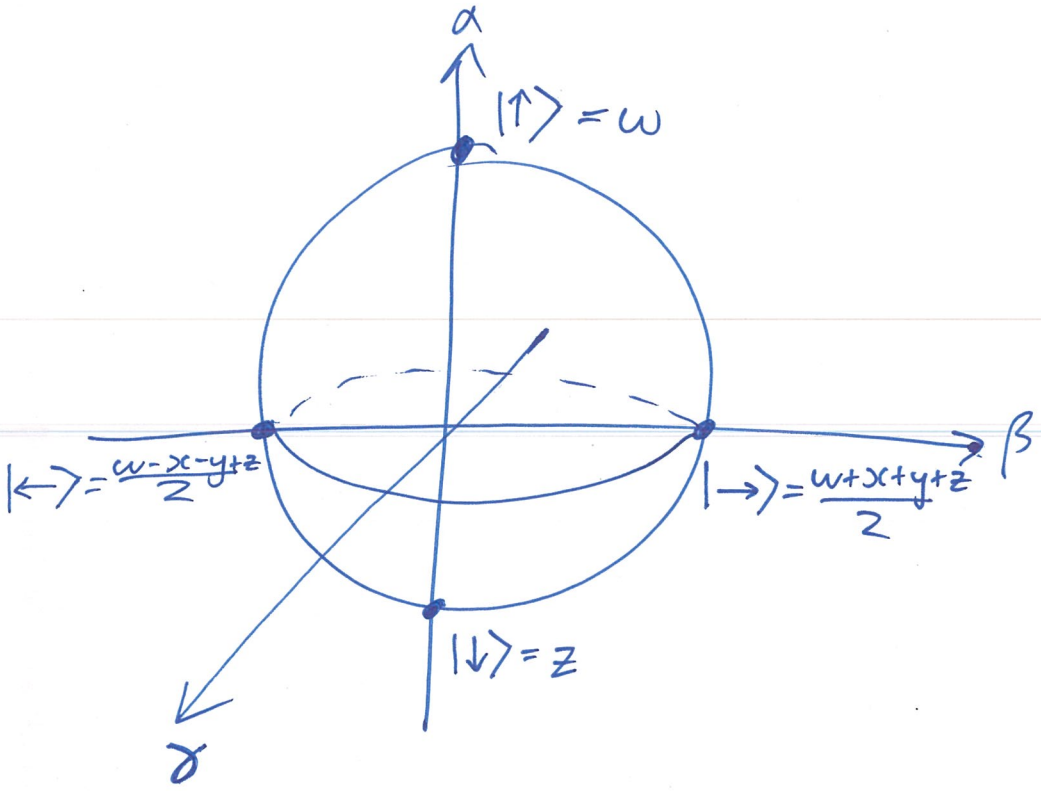
Our 'state-updating' rule is just a generalization of this rule to the non-commutative setting.

In the conventional presentation of quantum mechanics, this is described in terms of a scary-sounding "wave function collapse" process, which brings endless confusion and silliness.

Updating the state after making a measurement is no more or less mysterious than what happens when you look at a coin after you've flipped it.

What, then, makes quantum probability different from classical probability?

Recall the space of states for $M_2(\mathbb{C})$, with some labelled extremal states:



Let's measure the boolean observable $P=\begin{pmatrix}1&0\\0&0\end{pmatrix}$ on the state $\rho=|\uparrow\rangle$.

$p(P)=1$, so we definitely obtain the answer "yes".

The state doesn't change:

$$\hat{\rho}(X)=\frac{\rho(PXP)}{\rho(P)},\quad \therefore \hat{\rho}\begin{pmatrix}w&x\\y&z\end{pmatrix}=\frac{\rho\begin{pmatrix}w&0\\0&0\end{pmatrix}}{\rho\begin{pmatrix}1&0\\0&0\end{pmatrix}}=\frac{w}{1}=w,$$

$$\text{so } \hat{\rho}=\rho.$$

What if we measure $Q = \frac{1}{2}\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$?

$p(Q) = \frac{1}{2}$, so we are equally likely to measure "yes" and "no".

If we measure "yes", what is the new state?

$$\hat{\rho}(X) = \frac{\rho(QXQ)}{\rho(Q)}, \quad \hat{\rho}\begin{pmatrix} w & x \\ y & z \end{pmatrix} = \frac{\rho\left(\frac{w+x+y+z}{4}\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\right)}{\frac{1}{2}}$$

$$= \frac{w+x+y+z}{2}$$

so $\hat{\rho} = |{+}\rangle$.

Notice that if we measure P again on $\hat{\rho}$, the answer is no longer definite:

$$\hat{\rho}(P) = \frac{1}{2}.$$

Rather differently than in the classical setting, there are <u>no definite states</u>, that is, states ρ so $\rho(P) = 0$ or 1 for every boolean observable P.

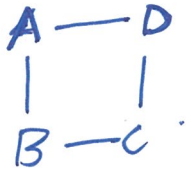
It gets even stranger!

Bell's theorem(s).

Consider four observables A, B, C, D which each take values in $\{\pm 1\}$.

Classically, in any state, definite or not, we have

$$p(AB + BC + CD - AD) \leq 2.$$

(Think of these in a square 

.

We could only get an answer bigger than 2 if A and D took different values. But then at least one pair from $\{A, B\}$, $\{B, C\}$ and $\{C, D\}$ must take different values.)

However, this inequality can be violated in non-commutative probability!

Consider the following setup in $M_4(\mathbb{C})$.

$$\rho(X) = \mathrm{tr}(RX), \quad R = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & -\frac{1}{2} & 0 \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix} \quad C = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$D = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 & & \\ -1 & -1 & & \\ & & 1 & -1 \\ & & -1 & -1 \end{pmatrix} \quad B = -\frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 & & \\ 1 & -1 & & \\ & & 1 & 1 \\ & & 1 & -1 \end{pmatrix},$$

and one easily checks

$$\rho(AB + BC + CD - AD) = \rho(AB) + \rho(BC) + \rho(CD) - \rho(AD)$$

$$= \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2}} - -\frac{1}{\sqrt{2}}$$

$$= 2\sqrt{2} \geqslant 2.$$

Measuring more general observables.

Given an observable X, we could attempt to construct the ~~doesn't~~ boolean observable $P_{X \in A}$ with result 1 if we measre X and obtain a result in the set A, and 0 otherwise.

In fact, asking that our algebra $\mathcal{A}$ is a von Neumann algebra ensures we can do this in a nice way:

$$\sum_i a_i P_{X \in [a_i, a_{i+1})} \longrightarrow X$$

if $\{a_0 \leq a_1 \leq \cdots \leq a_k$ is a partition of "the spectrum" of X that becomes finer and finer.

We find $P_{X \in A} = 0$ unless $A \cap sp(X) \neq \{\}$,

~~and so we~~ where $sp(X) = \{\lambda \mid X - \lambda I$ is not invertible $\}$

and so we say that "measuring X" gives ~~an~~ an answer in $sp(X)$.

The updated state is then $\hat{\rho}(Y) = \dfrac{\rho(P_{X=\lambda} Y P_{X=\lambda})}{\rho(P_{X=\lambda})}$.