Quantum Information

Ninnat Dangniam

Institute for Fundamental Study (IF)
Naresuan University, Phitsanulok,
Thailand 65000

ninnatd@nu.ac.th

Contents

CHAPTER 1

Information and Bits

SECTION 1.1
Probability

We will set up and formalize notations to talk about randomness using the context of statistical experiments. These can be actual ("artificial") experiments in the lab, or natural processes, the outcomes of which are random.

Denote by Ω the set of all possible outcomes (**events**) of a statistical experiment, usually called the **sample space** or the "universe".

Example | **(Coin toss)** The sample space consists of the outcome that the coin comes up head (H) or tail (T): $\Omega = \{H, T\}$.

Example | **(Die roll)** The sample space is the number on the face of a die: $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Each outcome of a statistical experiment can be associated with a **logical proposition** in the obvious way. For example, the result "1" of a die roll is associated with the proposition "the die is rolled and we obtained a 1". Every such proposition can be assigned a binary value TRUE or FALSE. These outcomes do not exhaust all possible events because we can combine events using logical (Boolean) operations such as AND or OR to create new logical propositions.

	Logical operation	Set operation
NOT	$\neg A$	$\overline{A}$
AND	$A \wedge B$	$A \cap B$
OR	$A \vee B$	$A \cup B$

Table 1. Symbols for elementary logical operations and the corresponding set operations, which we use interchangeably.

An **empty event** (set) is denoted by $\emptyset$. Self-evidently,

$$A \wedge \overline{A} = \emptyset, \qquad A \vee \overline{A} = \Omega. \tag{1.1}$$

Whenever Ω is a finite set, we can always find an elementary set of disjoint (**mutually exclusive**) propositions called **elementary events** or **atomic events** $\{E_j\}_j$:

$$E_j \wedge E_k = \begin{cases} E_j, & j = k, \\ \emptyset, & j \neq k. \end{cases} \tag{1.2}$$

The set of all possible events i.e. logical combinations of all atomic events, constitutes a *Boolean algebra* $\{0,1\}^n$ whose size is 2^n if there are n atomic events. It is the set $\{0,1\}^n$ because an event is defined by whether each atomic event is absent or present in the Boolean formula $E_{j_1} \vee E_{j_2} \vee \cdots \vee E_{j_m}$, $m \leq n$.

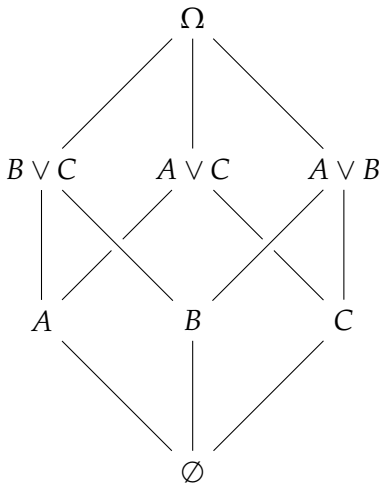

Figure 1. The Hesse diagram of a Boolean algebra with three atomic events A, B, and C. Going up in the diagram corresponds to the logical operation OR ($\vee$), while going down corresponds to AND ($\wedge$).

$\{0,1\}^n$ is also the *power set* $\mathcal{P}(\Omega)$, the set of all subsets of Ω. The power set X^Y is the set of all functions from Y to X. There are exactly $|X|^{|Y|}$ such functions, hence the notation. (Why? For each input in Y, one can choose an output from all possible choices of $x \in X$.)

So far we have not yet talked about assigning probabilities to events. There is a long-standing debate about what a probability actually means, but for us a

probability is just a number between 0 (the event never occurs) and 1 (the event occurs with certainty).

Definition 1 **(Probability axioms)**

1. $\Pr(A) \geq 0$,
2. $\Pr(A) = 1 \iff A$ is certain,
3. $\Pr(A \vee B) = \Pr(A) + \Pr(B)$ if $A \wedge B = \emptyset$

These axioms are sufficient to derive any other identities such as

$$\Pr(\neg A) = 1 - \Pr(A), \tag{1.3}$$
$$\Pr(A \vee B) = \Pr(A) + \Pr(B) - \Pr(A \wedge B). \tag{1.4}$$

Let us formalize one more thing: a **random variable** X is a variable that takes a value $x \in \Omega$ with probability $\Pr(X = x)$. We usually write $X \sim f$ to mean that values of X are distributed according to a probability distribution f.

Random variable $\leftrightarrow$ Statistical experiment

We will use interchangeably the notations

$$\Pr(X = x) \equiv p_X(x) \equiv p(x). \tag{1.5}$$

Definition 2 **(kth moment)**

$$\underbrace{\mathbb{E}[X^k]}_{\substack{\text{Mathematicians'}\\ \text{notation}}} \equiv \underbrace{\left\langle X^k \right\rangle}_{\substack{\text{Physicists'}\\ \text{notation}}} \equiv \left\langle x^k \right\rangle = \sum_x x^k p(x) \tag{1.6}$$

The **mean value**, also known as the average or the expectation value, is the first moment, whereas the **variance** is the part of the second moment that is independent of the first moment.[1]

$$\mathrm{Var}(X) \equiv \sigma_x^2 = \langle (X - \langle X \rangle)^2 \rangle = \left\langle X^2 \right\rangle - \langle X \rangle^2 \tag{1.7}$$

[1] *The variance is the second cumulant. The Gaussian distribution is the unique probability distribution whose cumulants all vanish except the first and the second.*

Example **(Bernoulli trials)**
Suppose that X takes values n, the number of heads obtained in N independent tosses of a coin with a bias p. (More about independence later.) The probability of such an event is given by the *binomial distribution* $X \sim B(N, p)$.

$$\left(\begin{array}{c}\text{Probability of}\\ \text{obtaining a}\\ \text{particular}\\ \text{sequence with}\\ \text{exactly } n \text{ heads}\end{array}\right) = p^n(1-p)^{N-n} \tag{1.8}$$

$$\Pr(X=n) = \left(\begin{array}{c}\text{Probability of}\\ \text{obtaining } \textit{any}\\ \text{particular}\\ \text{sequence with}\\ \text{exactly } n \text{ heads}\end{array}\right) = \binom{N}{n} p^n(1-p)^{N-n} \tag{1.9}$$

The normalization can be verified directly. Let $q = 1-p$.

$$\sum_{n=0}^{N} \Pr(X=n) = \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} = (p+q)^N = 1, \tag{1.10}$$

where we have used the binomial theorem in the second-to-last equality.

$$\langle n \rangle = \sum_{n=0}^{N} np(n) = \sum_{n=0}^{N} n\binom{N}{n} p^n q^{N-n} \tag{1.11}$$

$$= p\frac{\partial}{\partial p} \sum_{n=0}^{N} \binom{N}{n} p^n q^{N-n} = p\frac{\partial}{\partial p}(p+q)^N \tag{1.12}$$

$$= Np\cancelto{1}{(p+q)^{N-1}} = Np \tag{1.13}$$

The same trick can be used to calculate the variance $\sigma^2 = Np(1-p)$.

Things become interesting when there are two or more random variables.

Definition 3 **(Joint probability)**

$$\Pr(X=x, Y=y) \equiv p_{XY}(x,y) \equiv p(x,y) \tag{1.14}$$

is the probability that both the outcomes x and y happen.

Definition 4 **(Marginal probability)**

$$p(x) = \sum_y p(x,y), \qquad p(y) = \sum_x p(x,y), \tag{1.16}$$

$$p(x) = \underbrace{p(x,y)}_{\substack{\text{Probability}\\ \text{that both } x\\ \text{and } y \text{ occur}}} + \underbrace{p(x,\neg y)}_{\substack{\text{Probability}\\ \text{that } x \text{ occurs}\\ \text{but } y \text{ doesn't}}} \tag{1.15}$$

A marginal probability is the probability that an outcome described by one of the random variables may occur without looking at the outcome of the other random variable.

Definition 5 **(Conditional probability)**

For $\Pr(Y = y) \neq 0$,

$$\Pr(X = x|\underbrace{Y = y}_{\text{Condition}}) \equiv p(x|y) = \frac{p(x,y)}{p(y)} \tag{1.17}$$

is the probability that the event x occurs *if* the event y also occurs.

A conditional probability distribution for a fixed value of Y must normalize to 1.

$$\sum_x p(x|y) = \frac{1}{p(y)} \sum_x p(x,y) = \frac{\cancel{p(y)}}{\cancel{p(y)}} = 1 \tag{1.18}$$

Other useful equalities are obtained from these basic definitions.

Lemma 1 **(Law of total probability)**

$$p(x) = \sum_y p(x|y)p(y) \tag{1.19}$$

Lemma 2 **(Bayes' theorem)**

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}. \tag{1.20}$$

Bayes' theorem emphasizes that $p(x|y)$ and $p(y|x)$ are not equal in general. As an example, the probability of having Covid given a positive test result is not equivalent to the probability of testing positive when actually having Covid. Usually when computing $p(x|y)$ via Bayes' theorem, one computes the denominator $p(y)$ using the law of total probability.

Definition 6 Events x and y are **independent** if

$$p(x,y) = p(x)p(y). \tag{1.21}$$

Otherwise, they are **correlated**. We also say that two random variables X and Y are correlated if $p(x,y) \neq p(x)p(y)$ for some x and y.

In words, events x and y are independent if looking at the outcome of one of them does not give you any information about the other event i.e. does not make the other event more or less likely to occur.

$$p(x|y) = \frac{p(x,y)}{p(y)} = \frac{p(x)\cancel{p(y)}}{\cancel{p(y)}} = p(x). \tag{1.22}$$

Mutually exclusive events cannot be independent. If I flip a coin and it comes up head, it could not have come up tail. Mathematically, this is because, if $x \wedge y = \emptyset$, then $p(x,y) = 0$ but $p(x)p(y)$ cannot be zero if both $p(x)$ and $p(y)$ are nonzero.

SECTION 1.2

Information entropy

The number of bits required to describe a random variable is quantified by the Shannon entropy.

Definition 7 **(Shannon entropy)**

$$H(X) \equiv -\sum_x p(x) \log p(x) \tag{1.23}$$

When the logarithm is base 2, the entropy is measured in terms of *bits*. The special case when all n outcomes are equally likely reduces $H(X)$ to the Hartley entropy $\log n$. In physics, sometimes the natural log is used[2] and the Shannon entropy is proportional to the Gibbs entropy, which reduces to the Boltzmann entropy in the case of the uniform distribution.

Think of the Shannon entropy as the average *surprisal*, $-\log p(x)$, upon learning an outcome x. The rarer the outcome x is, the more information you gain by observing x occurring.[3] Also worth noting is that $\log 1 = 0$ agrees with our intuition that an observation can only convey new information when there are alternatives to the outcome that we have observed.

[2] *It is still common, however, to use logarithm base 2 in quantum information science.*

[3] *You may be worried that* $-\log 0 = \infty$*, but one can show that* $\lim_{x\to 0^+} 0 \log 0 = 0$ *by L'hospital's rule.*

Example **(Horse race)**

An example that illustrates how the Shannon entropy determines the amount of information in a random source is the following.

Consider a scenario where you agree to watch a horse race and send a telegram to your friend, indicating the winning horse while minimizing the number of letters used in the message. Suppose that there are four horses in the race, and you and your friend believe that the chance of horse #1 winning is $1/2$, horse #2 winning is $1/4$, and horses #3 and #4 winning are equally likely at $1/8$ each.

A naive encoding protocol would be to assign two bits to represent each horses: 00 for horse #1, 01 for horse #2, 10 for horse #3, and 11 for horse #4. However, there is a more efficient way to achieve a better average message length by utilizing a *variable-length* code. To ensure unique decodability, you may want to use, say, a prefix-free code where no codeword is a prefix of another. In this case, you can encode the horses as follows: 0 for horse #1, 10 for horse #2, 110 for horse #3, and 111 for horse #4. The average length of this encoding is

$$1 \cdot \frac{1}{2} + 2 \cdot \frac{1}{4} + 2 \cdot 3 \cdot \frac{1}{8} = 1.75 < 2. \tag{1.24}$$

Notice that the computation of the average message length is precisely the computation of the Shannon entropy. In fact, by an application of Kraft inequality, one can show that the Shannon entropy upper bounds the expected length of any prefix-free code. Also, by thinking about Huffman coding which achieves the optimal expected message length, one can also see that the

Shannon entropy gives the expected number of binary questions one needs to ask to completely remove the uncertainty from a random variable.

Example (**Typical sequences**)
Consider a multinomial distribution which generalizes the binomial distribution to d outcomes.

$$p(n_1, n_2, \ldots, n_d) = \frac{N!}{n_1! \cdots n_d!} p_1^{n_1} \cdots p_d^{n_d}. \tag{1.25}$$

In the limit of a long-running experiment $N \to \infty$, the number of x_j that appears in a sequence is $\langle n_j \rangle = N p_j$. A sequence in which x_j appears exactly $\langle n_j \rangle$ times is said to be a *typical sequence*.

$$\begin{pmatrix} \text{Probability of} \\ \text{obtaining a} \\ \text{particular typical} \\ \text{sequence} \end{pmatrix} = p_1^{n_1} \cdots p_d^{n_d} = p_1^{N p_1} \cdots p_d^{N p_d} \tag{1.26}$$

$$= \exp(N p_1 \log p_1) \cdots \exp(N p_d \log p_d) \tag{1.27}$$

$$= \exp[N(p_1 \log p_1 + \cdots + p_d \log p_d)] \tag{1.28}$$

$$= e^{-NH(X)} \tag{1.29}$$

If we count the number of such a typical sequence, we will find that

$$\ln\left(\frac{N!}{n_1! \cdots n_d!}\right) = \ln(N!) - \sum_j \ln(n_j!) \tag{1.30}$$

$$= N \ln N - \cancel{N} - \sum_j (n_j \ln n_j - \cancel{n_j}) \tag{1.31}$$ Stirling's formula

$$= N \ln N - \sum_j N p_j \ln(N p_j) \tag{1.32}$$

$$= \cancel{N \ln N} - \cancel{N \ln N} - N \sum_j p_j \log p_j \tag{1.33}$$

$$= NH(X). \tag{1.34}$$

Remarkably, each typical sequence appears with probability $e^{-NH(X)}$ and there are $e^{NH(X)}$ of them, so they monopolize all the probabilities. This *asymptotic equipartition property* lies at the heart of Shannon's information theory.

Definition 8 (**Joint entropy**)

$$H(X, Y) = -\sum_{xy} p(x, y) \log p(x, y) \tag{1.35}$$

Definition 9 (**Conditional entropy**) Since $p(x|y)$ defines a probability distribution for a

fixed y. Let

$$H(X|Y=y) \equiv -\sum_x p(x|y)\log p(x|y). \tag{1.36}$$

Then the conditional entropy is the average

$$H(X|Y) = \sum_y p(y)H(X|Y=y). \tag{1.37}$$

Lemma 3

$$H(X|Y) = H(X,Y) - H(Y) \tag{1.38}$$

In words, the conditional entropy is the uncertainty left in X after observing the value of Y.

$$H(X|Y) = \sum_y p(y)H(X|Y=y) = \sum_y p(y)\sum_x p(x|y)\log p(x|y) \tag{1.39}$$

$$= -\sum_{xy} p(x,y)[\log p(x,y) - \log p(y)] \tag{1.40}$$

$$= H(X,Y) + \sum_y \underbrace{\sum_x p(x,y)}_{p(y)} \log p(y) = \boxed{H(X,Y) - H(Y)} \tag{1.41}$$

Definition 10

(**Mutual information**) The common information shared between X and Y is

$$H(X:Y) = \sum_{x,y} p(x,y)\log\left(\frac{p(x,y)}{p(x)p(y)}\right) \tag{1.42}$$

$$= H(X) + H(Y) - H(X,Y) \tag{1.43}$$

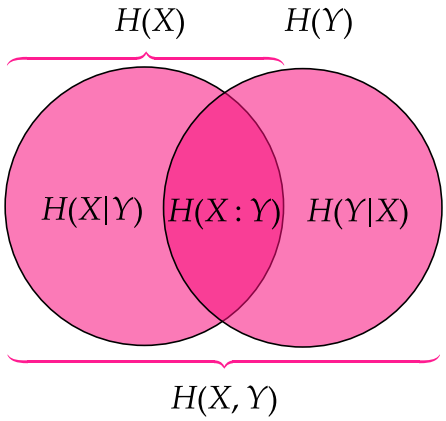

Figure 2. Venn diagram showing the relationships between various entropic quantities associated with two random variables.

CHAPTER 2

Mathematical Prerequisite

SECTION 2.1 Transition from wave mechanics

The stationary states from a complete, orthonormal set

$$\boxed{\begin{aligned}\int dx\, \varphi_n^*(x)\varphi_m(x) &= \delta_{nm} \qquad (2.1)\\ \sum_n \varphi_n^*(x)\varphi_n(x') &= \delta(x-x') \qquad (2.2)\end{aligned}}$$

(2.1) Orthonormality

(2.2) Completeness

The Dirac delta "function" is a generalized function, also known as a distribution, and only has meaning inside an integral.

$$\int dx'\, f(x')\delta(x-x') = f(x) \tag{2.3}$$

Think of the delta function as a linear map from $\mathcal{H}$ to $\mathbb{C}$.

Given a function $\psi(x)$. Suppose that we can write

$$\psi(x) = \sum_n c_n \varphi_n(x). \tag{2.4}$$

Orthogonality implies that

$$\begin{aligned}\int dx\, \varphi_n^*(x)\psi(x) &= \sum_m c_m \int dx\, \varphi_n^*(x)\varphi_m(x) \qquad (2.5)\\ &= \sum_m c_m \delta_{nm} = c_n. \qquad (2.6)\end{aligned}$$

In other words, we have an identity

$$\boxed{c_n = \int dx\, \varphi_n^*(x)\psi(x)}. \tag{2.7}$$

Equipped with the orthonormality relation, we can show that the closure relation is equivalent to the spanning property.[4]

[4] *Of course, a set that is not orthonormal (nor orthogonal) can also span the space, in which case the closure relation is generalized to the* frame condition.

Lemma 4 An orthonormal set $\{\varphi_n(x)\}_n$ spans the space of function if and only if it satisfies the closure relation (2.2).

PROOF " $\Longleftarrow$ " direction

Suppose that the closure relation is true. For an arbitrary $\psi(x)$, compute

$\psi(x) = \sum_n c_n \varphi_n(x)$. We want to show that $\sum_n c_n \varphi_n(x)$ is nothing but $\psi(x)$.

$$\sum_n c_n \varphi_n(x) = \int dx' \, \psi(x') \sum_n \varphi_n^*(x') \varphi_n(x) \tag{2.8}$$

$$= \int dx' \, \psi(x') \delta(x - x') = \psi(x) \tag{2.9}$$

as desired.
" $\Longrightarrow$ " direction
Suppose that the expansion (2.4) is valid for an arbitrary function,

$$\psi(x) = \int dx' \, \psi(x') \sum_n \varphi_n^*(x') \varphi_n(x), \tag{2.10}$$

then it must be that the sum acts like a Dirac delta function, which completes the proof. □

In quantum theory, $|c_n|^2$ has the meaning of the probability to find the system in the state $\varphi_n(x)$. It is straightforward to verify that

$$\sum_n |c_n|^2 = \int dx \, |\psi(x)|^2. \tag{2.11}$$

More generally, if $\psi(x) = \sum_n c_n \varphi_n(x)$ and $\phi(x) = \sum_n d_n \varphi_n(x)$. Then the inner product in the function space is the same as the inner product between infinitely long vectors of sequences $\{c_n\}_n$ and $\{d_n\}_n$:

$$\int dx \, \phi^*(x) \psi(x) = \sum_{nm} d_n^* c_m \int dx \, \varphi_n^*(x) \varphi_m(x) \tag{2.12}$$

$$= \sum_{nm} d_n^* c_m \delta_{nm} = \sum_n d_n^* c_n \tag{2.13}$$

$$= \begin{pmatrix} d_1^* & d_2^* & \dots \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \end{pmatrix} \tag{2.14}$$

Technically, these two kinds of inner product define different Hilbert spaces. One is the space l^2 of *square-summable* sequences,

$$\sum_n |c_n|^2 < \infty. \tag{2.15}$$

The other is the space L^2 of *square-integrable* functions,

$$\int dx \, |\psi(x)|^2 < \infty. \tag{2.16}$$

In Heisenberg's original conception of quantum theory, he rejected ascribing (continuous-valued) dynamical variables such as position and momentum to atomic orbitals on the basis that they are unobservable and that a theory that included them so far had failed to make the correct prediction [2]. Heisenberg

focused instead on discrete quantities $n = 1, 2, \ldots$ that label the atomic orbitals. In this approach, known as *matrix mechanics*, there are only observables and "quantum jumps" between different values of an observable. Notably, there is no notion of a quantum state[5], but one can think of the coefficients $\{c_n\}_n$ as a "state". Thus, one is working in the Hilbert space l^2 of square-summable sequences.

[5] *So one is spared from having to physically interpret a superposition of quantum states.*

Schrödinger came from a different vantage point entirely. He took seriously de Broglie's idea that associated to every quantum particle is a wave and asked what the wave equation for de Broglie's waves are. This led to the Schrödinger equation and the approach known as *wave mechanics*. In this approach, there are only states, but no probability, at least at first.[6] It was Max Born that supplemented the probabilistic interpretation in this approach. As a result, one is working in the Hilbert space L^2 of square-integrable functions.

[6] *Schrödinger hihmself even mistakenly believed that the wave function describes a charge density of some sort.*

However, the two approaches were the same all along as the two Hilbert spaces are equivalent (Riesz-Fischer theorem). From this line of thinking, John von Neumann developed a unified formulation of quantum mechanics as we know today [1, 2]. For us, the whole point is that $\{c_n\}_n$ and $\psi(x)$ are simply two representations of the same element $|\psi\rangle$ of an abstract Hilbert space $\mathcal{H}$.

Example (Particle in a box)

$$|\varphi_n\rangle \longleftrightarrow \varphi_n(x) = \sqrt{\frac{2}{L}} \cdot \begin{cases} \sin(k_n x), & \text{even } n, \\ \cos(k_n x), & \text{odd } n, \end{cases} \tag{2.17}$$

where $k_n = n\pi/L$. Note the important fact that the ket $|\varphi_n\rangle$ has no x-dependence.

Example (Spin-1/2)

$$|\psi\rangle = \alpha\,|\uparrow\rangle + \beta\,|\downarrow\rangle \tag{2.18}$$

SECTION 2.2

Inner product spaces

The prerequisites of this section are definitions and basic properties related to vector spaces and linear operators, which can be found in Appendix A.

Definition 11 A pairing

$$V \times V \to \mathbb{C}, \tag{2.19}$$
$$\phi, \psi \mapsto (\varphi, \psi) \tag{2.20}$$

is said to be an **inner product** if the following properties are satisfied.

- *Linearity in the second argument*:

$$(\phi, a\psi_1 + b\psi_2) = a(\phi, \psi_1) + b(\phi, \psi_2) \tag{2.21}$$

- *Conjugate symmetry:*

$$(\psi, \phi) = (\phi, \psi)^* \tag{2.22}$$

- *Positive-definiteness:*

$$(\psi, \psi) \geq 0 \text{ with equality iff } |\psi\rangle = 0 \tag{2.23}$$

Inner products in real vector spaces are bilinear, but inner products in complex vector spaces are *sesquilinear* they are , i.e., they are conjugate linear in the first argument. Sesquilinearity follows from (2.21) and (2.22)

$$\begin{aligned}(a\phi_1 + b\phi_2, \psi) &= (\psi, a\phi_1 + b\phi_2)^* && (2.24)\\ &= a^*(\psi, \phi_1)^* + b^*(\psi, \phi_2)^* && (2.25)\\ &= a^*(\phi_1, \psi) + b^*(\phi_2, \psi) && (2.26)\end{aligned}$$

Sesquilinearity is required to make (ψ, ψ) non-negative.

Example (Weird inner product)
The following formula defines an inner product in $\mathbb{R}^2$.

$$(x, y) = 5x_1y_1 + x_1y_2 + x_2y_1 + 3x_2y_2. \tag{2.27}$$

The positive definiteness can be shown by noting that

$$(x, x) = 5x_1^2 + 2x_1x_2 + 3x_2^2 = (x_1 + x_2)^2 + 4x_1^2 + 2x_2^2 \geq 0, \tag{2.28}$$

with equality iff every term is zero.

An inner product with a fixed vector ϕ always can be thought as a mapping

$$\begin{aligned}(\phi, \_) : V &\to \mathbb{C}, && (2.29)\\ \psi &\mapsto (\phi, \psi) && (2.30)\end{aligned}$$

This is an example of a *linear functional*, a linear map from V to $\mathbb{C}$.

Definition 12 The **dual space** V^* of V is defined as the vector space of linear functionals with addition defined as[7]

$$(f + g)(\psi) = f(\psi) + g(\psi). \tag{2.31}$$

[7] *Scalar multiplication in V^* follows directly from the the linearity of linear functionals: $f(a\psi) = af(\psi)$.*

Lemma 5 $\dim V^* = \dim V$.

PROOF For any given basis $\{e_j\}_j$ for V, we can define a basis $\{f_k\}_k$ for V^* by the relation $f_k(e_j) = \delta_{jk}$. If there is an element g of V^* that cannot be represented as a linear combination of $\{f_k\}_k$, then $g(e_j) = 0$ for every j, which implies that g is the zero map. Therefore, $\{f_k\}_k$ spans V^* and there are $\dim V$ of them. □

For finite dimensional vector spaces, the characterization of linear functionals are obvious in matrix form. By definition, any linear map from an n-

dimensional vector space to $\mathbb{C}$ can be represented by a 1-by-n matrix,

$$f\left[\begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}\right] = \begin{pmatrix} * & * & \dots & * \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix}, \tag{2.32}$$

which suggests that all linear functionals can be realized as an inner product.[8] This is indeed true (Riez representation theorem). An inner product induces an antilinear map † that identifies V and V^*,

$$|\psi\rangle \overset{\dagger}{\longleftrightarrow} \langle\psi| , \tag{2.33}$$

but the specification of the map itself will depend on the choice of the inner product. If we choose an orthonormal basis $\{|e_j\rangle\}_{j=1}^n$,

$$(e_j, e_k) = \delta_{jk}, \tag{2.34}$$

this map is

$$c_1 |e_1\rangle + \cdots + c_n |e_n\rangle \overset{\dagger}{\longleftrightarrow} c_1^* \langle e_1| + \cdots + c_n^* \langle e_n| . \tag{2.35}$$

In other words, $f \in V^*$ is identified with the vector $\phi_f = \sum_j [f(e_j)]^* e_j \in V$, for suppose that $\psi = \sum_j c_j e_j$, then

$$(\phi_f, \psi) = \sum_j c_j (v_f, e_j) \tag{2.36}$$

Linearity in the second argument

$$= \sum_{jk} c_j \{[f(e_k)]^*\}^* (e_k, e_j) = \sum_{jk} c_j [f(e_k)] \delta_{jk} \tag{2.37}$$

Conjugate linearity in the first argument

$$= \sum_j c_j f(e_j) = f\left[\sum_j c_j e_j\right] = f(\psi), \tag{2.38}$$

as it should be.

[8] *But note that the notion of linear functionals of the dual space does not require the notion of an inner product.*

Theorem 1 **(Cauchy-Schwarz)**

$$|\langle\phi|\psi\rangle|^2 \leq \langle\phi|\phi\rangle \langle\psi|\psi\rangle$$

PROOF Given a pair of vectors $|\phi\rangle$ and $|\psi\rangle$, one can consider the triangle formed by $\left|\psi^{\parallel}\right\rangle$, the component of $|\psi\rangle$ parallel to $|\phi\rangle$:

$$|\psi^{\parallel}\rangle = \frac{\langle\phi|\psi\rangle}{\langle\phi|\phi\rangle} |\phi\rangle , \tag{2.39}$$

and the perpendicular component $\left|\psi^{\perp}\right\rangle = |\psi\rangle - |\psi^{\parallel}\rangle$, see Fig 3.

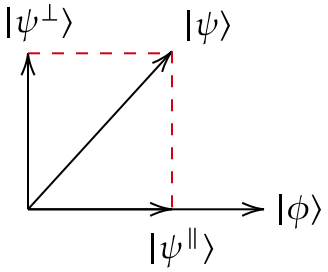

Figure 3.

The Pythagorean theorem tells us that

$$\langle\psi|\psi\rangle = \langle\psi^\perp|\psi^\perp\rangle + \langle\psi^\parallel|\psi^\parallel\rangle \tag{2.40}$$

$$\langle\psi|\psi\rangle - \frac{|\langle\phi|\psi\rangle|^2}{\langle\phi|\phi\rangle} = \langle\psi^\perp|\psi^\perp\rangle \geq 0. \tag{2.41}$$

from which the Cauchy-Schwarz inequality

$$|\langle\phi|\psi\rangle|^2 \leq \langle\phi|\phi\rangle\,\langle\psi|\psi\rangle \tag{2.42}$$

emerges. Additionally, we see from (2.41) that the inequality is tight iff $|\phi\rangle$ and $|\psi\rangle$ are scalar multiple of each other. □

Definition 13 The norm of $|\psi\rangle$ is $\||\psi\rangle\| \equiv \|\psi\| := \sqrt{\langle\psi|\psi\rangle}$.

Since an inner product defines a norm, a Hilbert space is also a normed vector space.[9] By definition, a norm needs to satisfy the triangle inequality

$$\||\psi\rangle + |\phi\rangle\| \leq \|\psi\| + \|\phi\|. \tag{2.43}$$

But for a norm induced from an inner product, the triangle inequality can be proved using the Cauchy-Schwarz inequality. Since the latter holds for any inner product, the former holds irrespective of the choice of inner product used to define the norm.

[9] *The converse is not true. Even though an inner product can be constructed from the 2-norm via the polarization identity, (the p-norm is defined as* $\|\psi\|_p = \sqrt[p]{\sum \left|c_j\right|^p}$*), a normed space is not always an inner product space. Take the max norm,* $\|\psi\|_\infty = \max_j \left|c_j\right|$*, for example.*

PROOF

$$\||\psi\rangle + |\phi\rangle\|^2 = \|\psi\|^2 + \|\phi\|^2 + 2\mathrm{Re}\,\langle\psi|\phi\rangle \tag{2.44}$$

$$\leq \|\psi\|^2 + \|\phi\|^2 + 2|\langle\psi|\phi\rangle| \tag{2.45}$$

$$\leq \|\psi\|^2 + \|\phi\|^2 + 2\|\psi\| \cdot \|\phi\| \tag{2.46}$$ Cauchy-Schwarz inequality

$$= (\|\psi\| + \|\phi\|)^2 \tag{2.47}$$

□

The vector space in quantum theory is a *separable Hilbert space*. A **Hilbert space**, denoted by $\mathcal{H}$, is a vector space equipped with an inner product and some convergence properties. Separability means that there exists a countable basis for the space.[10] Additionally, since the norm of a vector has the meaning of a probability, we demand that the norm of every vector is finite. (As a consequence of the Cauchy-Schwarz inequality, this also means that every inner product is finite.) For a finite-dimensional vector space, all these issues about separability and finiteness of the norm never arises; in this case, a Hilbert space *is simply an inner product space*.

[10] *Stationary states* $|\varphi\rangle_n$ *form a countable basis, but continuous eigenvectors such as* $|x\rangle$ *do not.*

Given the correspondence between linear functionals and the inner product, the *bra-ket* $\langle v|u\rangle$ is often regarded as merely another notation for the inner product between two vectors in the Hilbert space. This interpretation is always valid in finite dimensions, but in an infinite-dimensional Hilbert space, there are linear functionals that have no corresponding vector. Take the bra $\langle x|$, for example. Its squared norm is

$$\langle x|x\rangle = \int dx' \langle x|x'\rangle \langle x'|x\rangle = \int dx' \delta^2(x-x') = \delta(0) = \infty. \tag{2.48}$$

Now we are ready to write down properties of an **orthonormal basis (ONB)** in the Dirac notation.

$$\boxed{\begin{aligned} \langle e_j|e_k\rangle &= \delta_{jk} \\ \sum_j |e_j\rangle\langle e_j| &= \hat{\mathbb{1}} \end{aligned}} \tag{2.49}$$ Orthonormality

(2.50) Completeness

The equivalence of the spanning property and (2.50) can be shown similar to the continuous case. Suppose

$$|\psi\rangle = \sum_j c_j |e_j\rangle . \tag{2.51}$$

Then

$$\langle e_j|\psi\rangle = \sum_k c_k \langle e_j|e_k\rangle \tag{2.52}$$

$$= \sum_k c_k \delta_{jk} = c_j. \tag{2.53}$$

That is, we have that

$$\boxed{c_j = \langle e_j|\psi\rangle} . \tag{2.54}$$

Then suppose that an arbitrary vector can be expanded in such a way,

$$|\psi\rangle = \sum_j c_j |e_j\rangle = \sum_j |e_j\rangle \langle e_j|\psi\rangle . \tag{2.55}$$

Then it must be the case that $\sum_j |e_j\rangle\langle e_j| = \hat{\mathbb{1}}$.

SECTION 2.3

Linear operators

Any linear operator takes as an input a vector and outputs another vector. From this we observe that a ket-bra like $|\psi\rangle\langle\phi|$ is a linear operator since

$$|\psi\rangle \langle\phi|\zeta\rangle = \langle\phi|\zeta\rangle |\psi\rangle , \tag{2.56}$$

which is again a vector.[11] This is the magic of the bra-ket notation.

[11] *In particular, the vector space of linear operators on $\mathcal{H}$ is $\mathcal{H}^* \otimes \mathcal{H}$.*

If an orthonormal set $\{|\varphi_j\rangle\}_j$ spans a subspace $S \subset V$, then

$$\widehat{P}_S = \sum_j |e_j\rangle\langle e_j| \tag{2.57}$$

is a projection operator onto the subspace S and $\widehat{\mathbb{1}} - \widehat{P}_S$ is the projection operator onto the orthogonal complement $S^\perp$. That is, the total space is the direct sum $V = S \oplus S^\perp$, and for $|v\rangle = |s\rangle + |p\rangle \in V$, where $s \in S$ and $p \in S^\perp$, we have that $\widehat{P}_S |v\rangle = |s\rangle$ and $(\widehat{\mathbb{1}} - \widehat{P}_S)|v\rangle = |p\rangle$.

Matrix elements.

$$\widehat{T} = \widehat{\mathbb{1}}\widehat{T}\widehat{\mathbb{1}} = \sum_{jk} |e_j\rangle \overbrace{\langle e_j|\,\widehat{T}\,|e_k\rangle}^{T_{nm}} \langle e_k| \tag{2.58}$$

$$\widehat{T}|\psi\rangle = \sum_{jk} T_{jk} \overbrace{\langle e_k|\psi\rangle}^{c_k} = \sum_j \overbrace{\left(\sum_k T_{jk} c_k\right)}^{d_j} |e_j\rangle \tag{2.59}$$

$$\leftrightarrow \begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_n \end{pmatrix} = \begin{pmatrix} T_{11} & T_{12} & \dots & T_{1n} \\ T_{21} & T_{22} & & \vdots \\ \vdots & & \ddots & \\ T_{n1} & \dots & & T_{nn} \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \tag{2.60}$$

We see that the components transform as

$$\boxed{d_j = \sum_k T_{jk} c_k}. \tag{2.61}$$

The transformation of basis vectors is in a sense “opposite":

$$\boxed{\widehat{T}\,|e_j\rangle = \sum_k T_{kj}\,|e_k\rangle}. \tag{2.62}$$

The merit of the Dirac notation is that it streamlines matrix multiplications.

$$\widehat{T}\widehat{S} = \sum_{jklm} T_{jl} S_{mk}\,|e_j\rangle\,\langle e_l|e_m\rangle\,\langle e_k| = \sum_{jk} \underbrace{\left(\sum_l T_{jl} S_{lk}\right)}_{(\widehat{T}\widehat{S})_{nm}} |e_j\rangle\langle e_k| \tag{2.63}$$

For example, (2.62) can be derived without breaking a sweat.

$$\widehat{T}\,|e_j\rangle = \sum_{kl} T_{kl}\,|e_k\rangle\,\langle e_l|e_j\rangle = \sum_{jk} T_{kj}\,|e_k\rangle \tag{2.64}$$

The same is true for (2.61).

$$d_j = \langle e_j | \widehat{T} | \psi \rangle = \sum_{klm} T_{kl} c_m \langle e_j | e_k \rangle \langle e_l | e_m \rangle = \sum_l T_{jm} c_m \tag{2.65}$$

Trace. The manifestly basis-independent definition of the trace is

$$\mathrm{tr}(|\psi\rangle\langle\phi|) = \langle\phi|\psi\rangle\,, \tag{2.66}$$

which, in an ONB, equivalent to

$$\mathrm{tr}\,\widehat{T} = \sum_{jk} T_{jk}\,\mathrm{tr}\big(|e_j\rangle\langle e_k|\big) \tag{2.67}$$

Linearity of the trace

$$= \sum_{jk} T_{jk} \langle e_k | e_j \rangle = \sum_j T_{jj} = \sum_j \langle e_j | \widehat{T} | e_j \rangle\,. \tag{2.68}$$

The trace has the cyclic property

$$\mathrm{tr}\Big(\hat{A}\hat{B}\Big) = \mathrm{tr}\Big(\hat{B}\hat{A}\Big). \tag{2.69}$$

Beware that for three or more matrices in the product, the cyclic property means that

$$\mathrm{tr}\Big(\hat{A}\hat{B}\hat{C}\Big) = \mathrm{tr}\Big(\hat{B}\hat{C}\hat{A}\Big), \tag{2.70}$$

whereas

$$\mathrm{tr}\Big(\hat{A}\hat{B}\hat{C}\Big) \neq \mathrm{tr}\Big(\hat{B}\hat{A}\hat{C}\Big). \tag{2.71}$$

SECTION 2.4

Spectral theorem

Definition 14 The **adjoint** $\widehat{T}^\dagger$ of a linear operator $\widehat{T}$ is defined implicitly by its action

$$(\widehat{T}^\dagger v, u) = (v, \widehat{T} u) \tag{2.72}$$

on any $u, v \in \mathcal{H}$.

a priori this may not be the same as the † map between V and V^*, but we can show immediately that the matrix elements of $\widehat{T}^\dagger$ in an ONB must be the conjugate transpose of those of $\widehat{T}$

As is written, (2.72) is not straightforward to write in the Dirac notation, which does not distinguish between the left- and right-action of an operator.

$$(\langle v|\,\widehat{T})\,|u\rangle = \langle v|\,(\widehat{T}\,|u\rangle) = \langle v|\widehat{T}|u\rangle \tag{2.73}$$

Nevertheless, using the conjugate symmetry of the inner product, one can ex-

press an equivalent condition,

$$\langle v|\widehat{T}^\dagger|u\rangle = \langle u|\widehat{T}|v\rangle^* . \tag{2.74}$$

Since (2.72) is valid for any $u, v \in \mathcal{H}$, it must hold true for members of an orthonormal basis set $\{|e_j\rangle\}_j$. Hence, we can immediately see that

$$(\widehat{T}^\dagger)_{jk} = \langle e_j|\widehat{T}^\dagger|e_k\rangle = \langle e_k|\widehat{T}|e_j\rangle^* = T^*_{kj}. \tag{2.75}$$

Other properties of the adjoint follow:

$$(a\widehat{T} + b\widehat{S})^\dagger = a^*\widehat{T}^\dagger + b^*\widehat{S}^\dagger, \tag{2.76}$$

$$(\widehat{T}\widehat{S})^\dagger = \widehat{S}^\dagger\widehat{T}^\dagger, \tag{2.77}$$

$$(\widehat{T}^\dagger)^\dagger = \widehat{T}. \tag{2.78}$$

A particularly helpful rule is that

$$(|u\rangle\langle v|)^\dagger = (\langle v|^\dagger)(|u\rangle^\dagger) = |v\rangle\langle u| . \tag{2.79}$$

The reversing of the multiplication order can be see directly, not as a result of the transposition, but as a result of the definition of the adjoint.

$$\underbrace{\langle v|}_{V^*}\underbrace{\widehat{T}|u\rangle}_{V} = \underbrace{\langle v|}_{W^*}\underbrace{\widehat{T}\,|u\rangle}_{W} \tag{2.80}$$

The most important result from linear algebra that lies at the foundation of quantum theory is the spectral theorem.

Theorem 2

(Spectral theorem)
Eigenvectors of $\widehat{T}$ can be chosen to be an ONB $\iff$ $\widehat{T}$ is a **normal operator**: $\widehat{T}\widehat{T}^\dagger = \widehat{T}^\dagger\widehat{T}$.

A sketch of the proof utilizing projection operators is provided in Appendix A. The theorem implies that the matrix of $\widehat{T}$ is diagonal in the basis of eigenvectors $\{|e_j\rangle\}_j$:

$$\widehat{T} = \sum_j \lambda_j |e_j\rangle\langle e_j| \longleftrightarrow \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & 0 \\ & & \lambda_3 & \\ & 0 & & \ddots \\ & & & & \lambda_n \end{pmatrix} \tag{2.81}$$

As a result, we can take any analytic function of $\widehat{T}$ by taking the function of the eigenvalues directly.

Definition 15

A linear operator $\widehat{T}$ is said to be

1. **Hermitian** if $\widehat{T}^\dagger = \widehat{T}$,

2. **positive** if $\langle\psi|\widehat{T}|\psi\rangle \geq 0$ for every $|\psi\rangle \in \mathcal{H}$,
3. a **projection operator** if it is Hermitian and $\widehat{T}^2 = \widehat{T}$,[12]
4. **unitary** if $\widehat{T}\widehat{T}^\dagger = \widehat{T}^\dagger\widehat{T} = \widehat{\mathbb{1}}$.

[12] *Hermiticity is a necessary condition, as there is an operator that satisfies $\widehat{T}^2 = \widehat{T}$, but is not a projection operator:*

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

Normal operators are analogous to complex numbers. Any linear operator can be written as a sum of its "real part" $\widehat{H}$ and i times the "imaginary part" $\widehat{G}$, both of which are **Hermitian**. That is, $\widehat{H}^\dagger = \widehat{H}$ and $\widehat{G}^\dagger = \widehat{G}$.

$$\widehat{T} = \overbrace{\underbrace{\frac{\widehat{T} + \widehat{T}^\dagger}{2}}_{\text{Hermitian}}}^{\widehat{H}} + i \overbrace{\underbrace{\frac{\widehat{T} - \widehat{T}^\dagger}{2i}}_{\text{Hermitian}}}^{\widehat{G}} \tag{2.82}$$

$\widehat{H}$ and $\widehat{G}$ do not commute in general, but they commute precisely when $\widehat{T}$ is normal.

$$[\widehat{H}, \widehat{T}] = \frac{1}{4i}[\widehat{T} + \widehat{T}^\dagger, \widehat{T} - \widehat{T}^\dagger] \tag{2.83}$$

$$= \frac{1}{4i}([\widehat{T}^\dagger, \widehat{T}] - [\widehat{T}, \widehat{T}^\dagger]) = \frac{1}{2i}[\widehat{T}^\dagger, \widehat{T}] \tag{2.84}$$

Relations between subclasses of normal operators are visualized as a Venn diagram in Figure 4. Their eigenvalues belong to the corresponding subclasses of complex numbers, see Table 2.

Number	Matrix
Complex	Normal
Unit modulus	Unitary
Real	Hermitian
Positive	Positive
Idempotent (0&1)	Projection

Table 2. Taxonomy of operators and their analogous types of numbers.

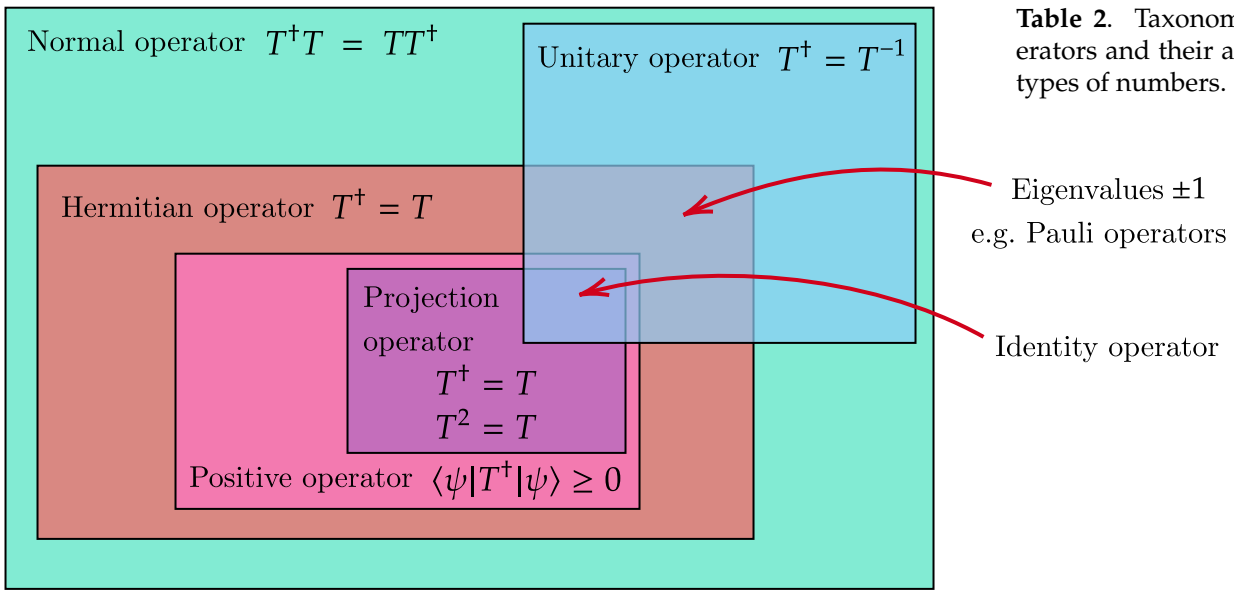

Figure 4. Taxonomy of linear operators that possess the spectral decomposition.

Lemma 6 Over the field $\mathbb{C}$, positivity implies Hermiticity.[13]

$$\langle\psi|\widehat{T}|\psi\rangle\,, \forall\,|\psi\rangle \in \mathcal{H} \implies \widehat{T}^\dagger = \widehat{T}. \tag{2.85}$$

PROOF Let $\widehat{T}$ be a positive operator. Write $\hat{T} = \widehat{H} + i\widehat{G}$, and let $|g_j\rangle$ be normalized eigenvectors of $\widehat{C}$ corresponding to eigenvalues λ_j. For

$$0 \leq \langle g_j|\widehat{T}|g_j\rangle = \underbrace{\langle g_j|\widehat{H}|g_j\rangle}_{\text{Real}} + i\lambda_j \tag{2.86}$$

to be true, λ_j must vanish for all j. Since $\widehat{G}$ is Hermitian and all its eigenvalues vanish, it must be the zero operator. Therefore, $\widehat{T}$ is Hermitian. □

Lemma 7 For any linear operator $\widehat{T}$, $\widehat{T}^\dagger\widehat{T}$ and $\widehat{T}\widehat{T}^\dagger$ are positive operators

PROOF

$$\langle\psi|\widehat{T}^\dagger\widehat{T}|\psi\rangle = (\langle\psi|\,\widehat{T}^\dagger)(\widehat{T}\,|\psi\rangle) = \left\|\widehat{T}\,|\psi\rangle\right\|^2 \geq 0. \tag{2.87}$$

□

Lemma 8 Since any unitary operator is normal, $\widehat{T}^\dagger\widehat{T} = \widehat{\mathbb{1}} \iff \widehat{T}\widehat{T}^\dagger = \widehat{\mathbb{1}}$ in finite dimensions.[14]

[13] *A counterexample over* $\mathbb{R}$*:*

$$(x\ y)\begin{pmatrix}2 & 2\\ 0 & 0\end{pmatrix}\begin{pmatrix}x\\ y\end{pmatrix} = x^2 + y^2 + (x+y)^2 \geq 0$$

[14] *A counterexample in an infinite dimension: the left shift* $\widehat{L}$ (A.14) *and the right shift* $\widehat{R}$ (A.15) *are adjoints of each other, and it is obvious that* $\widehat{L}\widehat{R} = \widehat{\mathbb{1}}$*, but* $\widehat{R}\widehat{L} \neq \widehat{\mathbb{1}}$ *because the left shift irrecoverbly kills the first component of a vector.*

CHAPTER 3

Postulates of quantum theory

SECTION 3.1

Quantum states and basis measurements

$$\boxed{\begin{aligned} \text{State} &\longleftrightarrow \text{Vector} && (3.1)\\ \text{Measurement} &\longleftrightarrow \text{Projection operator} && (3.2)\\ \text{Dynamics} &\longleftrightarrow \text{Unitary operator} && (3.3) \end{aligned}}$$

Any ONB gives a **basis measurement**, which defines a probability distribution,

$$p_k = |\langle e_k|\psi\rangle|^2 = \langle\psi|\overbrace{|e_k\rangle\langle e_k|}^{\widehat{P}_k}\psi\rangle = \langle\psi|\widehat{P}_k|\psi\rangle\,, \tag{3.4}$$

$\sum_k \widehat{P}_k = \widehat{\mathbb{1}}$. For p_k to be a probability, the state must therefore be normalized to 1:

$$\|\psi\| = \sqrt{\langle\psi|\psi\rangle} = 1. \tag{3.5}$$

After the measurement, the state is updated to the basis state $|e_k\rangle$ through the action of the projector,

$$\frac{\widehat{P}_k\,|\psi\rangle}{\sqrt{p_k}} = \frac{\widehat{P}_k\,|\psi\rangle}{\sqrt{\langle\psi|\widehat{P}_k|\psi\rangle}}. \tag{3.6}$$

Observe that multiplying a state vector by a complex phase has no effect on any empirical prediction. This can be seen from the following equation.

$$|\,\langle\phi|e^{i\delta}|\psi\rangle\,|^2 = \cancel{|e^{i\delta}|^2}^{\,1}|\langle\phi|\psi\rangle|^2 \tag{3.7}$$

This type of phase is commonly referred to as a **global phase** as opposed to a **relative phase**, which, on the other hand, gives rise to interference phenomena, a hallmark of quantum theory.

$$\left|\langle\phi|\left(|\psi_1\rangle + e^{i\delta}\,|\psi_2\rangle\right)\right|^2 = |\overbrace{\langle\phi|\psi_1\rangle}^{\psi_1} + e^{i\delta}\overbrace{\langle\phi|\psi_2\rangle}^{\psi_2}|^2 \tag{3.8}$$

$$= |\psi_1|^2 + |\psi_2|^2 + \underbrace{2\mathrm{Re}(e^{i\delta}\psi_1^*\psi_2)}_{\text{Quantum interference}} \tag{3.9}$$

SECTION 3.2

Case study: qubit

A two-level system, or a *qubit*, is any quantum sustem that can be described by a two-dimensional Hilbert space. We denote the standard ONB as $\{|0\rangle, |1\rangle\}$.

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{3.10}$$
$$= ae^{i\gamma}|0\rangle + be^{i\delta}|1\rangle \tag{3.11}$$
$$= e^{i\gamma}\left[\cos\left(\frac{\theta}{2}\right)|0\rangle + e^{i\varphi}\cos\left(\frac{\theta}{2}\right)|1\rangle\right] \tag{3.12}$$

Polar form where a, b, γ, δ are real numbers.
$a^2 + b^2 = 1$ and $\varphi = \delta - \gamma$.

By normalizing the state vector and eliminating the global phase, we have parametrized the state of a qubit by two spherical coordinate $0 \le \theta \le \pi$ and $0 \le \varphi < 2\pi$.

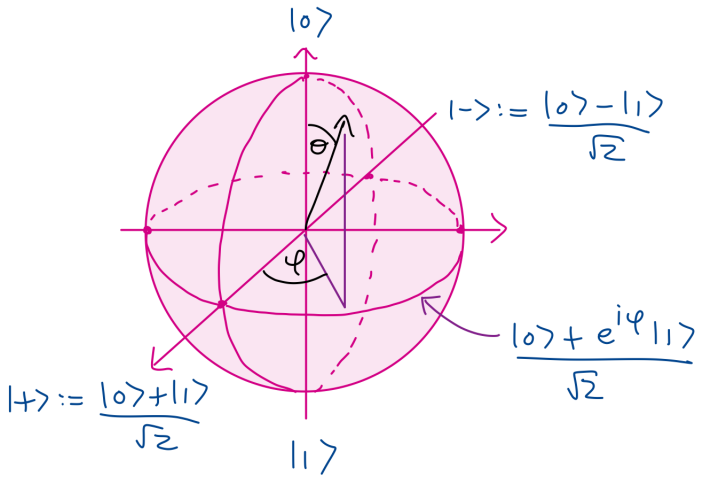

Figure 5. Space of qubit states is represented by the Bloch sphere.

Why the one half in $\theta/2$? Look at the operator space,[15]

$$|\hat{\mathbf{n}}\rangle = \cos(\theta/2)|0\rangle + \sin(\theta/2)e^{i\varphi}|1\rangle \tag{3.13}$$
$$|-\hat{\mathbf{n}}\rangle = -\sin(\theta/2)e^{-i\varphi}|0\rangle + \cos(\theta/2)|1\rangle \tag{3.14}$$

[15] *Given a general state of a two-level system, $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$, the orthogonal state (up to a phase) can be written as $|\psi^\perp\rangle = -\beta^*|0\rangle + \alpha^*|1\rangle$.*

$$|\hat{\mathbf{n}}\rangle\langle\hat{\mathbf{n}}| \longleftrightarrow \frac{1}{2}\begin{pmatrix} 1+\cos\theta & \sin\theta\, e^{-i\varphi} \\ \sin\theta\, e^{i\varphi} & 1-\cos\theta \end{pmatrix}$$
$$= \frac{1}{2}\begin{pmatrix} 1+z & x-iy \\ x+iy & 1-z \end{pmatrix} \tag{3.15}$$
$$= \boxed{\frac{\mathbb{1} + \hat{\boldsymbol{n}}\cdot\boldsymbol{\sigma}}{2}}, \tag{3.16}$$

Angel doubling formulae:

$$\sin(\theta) = 2\sin(\theta/2)\cos(\theta/2),$$
$$\cos(\theta) = \cos^2(\theta/2) - \sin^2(\theta/2)$$
$$= 2\cos^2(\theta/2) - 1.$$

where

$$\boldsymbol{\sigma} = \begin{pmatrix} \hat{\sigma}_x \\ \hat{\sigma}_y \\ \hat{\sigma}_z \end{pmatrix} \tag{3.17}$$

is the vector of Pauli matrices

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \hat{\sigma}_x = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad \hat{\sigma}_x = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{3.18}$$

Similarly,[16]

$$|-\hat{\mathbf{n}}\rangle\langle-\hat{\mathbf{n}}| \longleftrightarrow \frac{1}{2}\begin{pmatrix} 1-\cos\theta & -\sin\theta\, e^{-i\varphi} \\ -\sin\theta\, e^{i\varphi} & 1+\cos\theta \end{pmatrix} = \frac{\mathbb{1} + (-\hat{\boldsymbol{n}})\cdot\boldsymbol{\sigma}}{2} \tag{3.19}$$

[16] *The matrix form can also be obtained from that of* $|\hat{\mathbf{n}}\rangle\langle\hat{\mathbf{n}}|$ *by setting*

$$\theta \leftarrow \pi - \theta,$$
$$\varphi \leftarrow \pi + \varphi.$$

Spin observable in any direction $\hat{\mathbf{n}}$

$$\hat{\sigma}_{\hat{\mathbf{n}}} \equiv \hat{\boldsymbol{n}}\cdot\boldsymbol{\sigma} = |\hat{\mathbf{n}}\rangle\langle\hat{\mathbf{n}}| - |-\hat{\mathbf{n}}\rangle\langle-\hat{\mathbf{n}}| \longleftrightarrow \begin{pmatrix} \cos\theta & \sin\theta\, e^{-i\varphi} \\ \sin\theta\, e^{i\varphi} & \cos\theta \end{pmatrix} \tag{3.20}$$

To further investigate how relation between the geometry of the Hilbert space and that of the Euclidean space (inner product $\implies$ geometry)

Algebraic properties of Pauli operators

1. $\hat{\sigma}_j^\dagger = \hat{\sigma}_j$
2. $\hat{\sigma}_j^2 = \mathbb{1}$
3. $\hat{\sigma}_j\hat{\sigma}_k = i\hat{\sigma}_l$ where (j,k,l) are cyclic.
4. $\operatorname{tr}\hat{\sigma}_j = 0$ (true for an arbitrary $\hat{\sigma}_{\hat{\mathbf{n}}}$.)

Properties 2 and 3 can be written in a compact form using the Einstein summation convection as

$$\boxed{\hat{\sigma}_j\hat{\sigma}_k = \delta_{jk}\widehat{\mathbb{1}} + i\epsilon_{jkl}\hat{\sigma}_l}. \tag{3.21}$$

In particular, the commutator and the anticommutator are

$$[\hat{\sigma}_j, \hat{\sigma}_k] = 2i\epsilon_{jkl}\hat{\sigma}_l, \qquad \{\hat{\sigma}_j, \hat{\sigma}_k\} = 2\delta_{jk}\widehat{\mathbb{1}}. \tag{3.22}$$

Levi-Civita tensor

The Levi-Civita tensor can be used to define the d-dimensional totally anti-symmetric vector product. In three dimension,

$$\epsilon_{jkl} = \begin{cases} +1, & (j\ k\ l) \text{ is an } \textit{even} \text{ permutation } (1\ 2\ 3), (2\ 3\ 1), \text{or } (3\ 1\ 2), \\ -1, & (j\ k\ l) \text{ is an } \textit{odd} \text{ permutation } (1\ 3\ 2), (2\ 1\ 3), \text{or } (3\ 2\ 1), \\ 0, & \text{Otherwise.} \end{cases} \tag{3.23}$$

Here is a few facts about the Levi-Civita tensor. The determinant of a 3×3 matrix M can be expressed as the anti-symmetric product of its rows $\det M = \epsilon_{jkl} M_{1j} M_{2k} M_{3l}$. Thus, we have an alternative expression for the cross product:

$$\mathbf{A} \times \mathbf{B} = \hat{\mathbf{e}}_\mathbf{j} \epsilon_{jkl} A_k B_l = \det \begin{pmatrix} \hat{\mathbf{e}}_1 & \hat{\mathbf{e}}_2 & \hat{\mathbf{e}}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{pmatrix} \tag{3.24}$$

A contraction of two Levi-Civita tensors gives an expression involving the Kronecker deltas.

$$\epsilon_{jkl}\epsilon_{jmn} = \delta_{km}\delta_{lm} - \delta_{kn}\delta_{lm}, \tag{3.25}$$

$$\epsilon_{jkl}\epsilon_{jkm} = \delta_{kk}\delta_{lm} - \delta_{km}\delta_{lm} = (\dim - 1)\delta_{lm} = 2\delta_{lm} \tag{3.26}$$

$\delta_{jk}\delta_{jk} = \delta_{jj} = \operatorname{tr} \hat{\mathbb{1}}_{d\times d} =$ dim, dimension of the space

Now we can compute the inner product in a painless way without needing to use any trigonometric identity.

$$|\langle \hat{\mathbf{n}} | \hat{\mathbf{m}} \rangle|^2 = \frac{1}{4} \operatorname{tr} \left[(\mathbb{1} + \hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma}) (\mathbb{1} + \hat{\boldsymbol{m}} \cdot \boldsymbol{\sigma}) \right] \tag{3.27}$$

$$= \frac{1}{4} \operatorname{tr} \left[\mathbb{1} + \underbrace{\hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma} + \hat{\boldsymbol{m}} \cdot \boldsymbol{\sigma}}_{\text{Traceless}} + (\hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma})(\hat{\boldsymbol{m}} \cdot \boldsymbol{\sigma}) \right] \tag{3.28}$$

$$= \frac{1}{2} + \frac{1}{4} \operatorname{tr}[(\hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma})(\hat{\boldsymbol{m}} \cdot \boldsymbol{\sigma})] \tag{3.29}$$

What is the trace of $(\hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma})(\hat{\boldsymbol{m}} \cdot \boldsymbol{\sigma})$? Use the index notation.

$$(\hat{\boldsymbol{n}} \cdot \boldsymbol{\sigma})(\hat{\boldsymbol{m}} \cdot \boldsymbol{\sigma}) = n_j m_k \hat{\sigma}_k \hat{\sigma}_k \tag{3.30}$$

$$= n_j m_k (\delta_{jk} \mathbb{1} + i\epsilon_{jkl}) \tag{3.31}$$

$$= n_j m_j \mathbb{1} + i \underbrace{\epsilon_{jkl} n_j m_k}_{(\hat{\mathbf{n}} \times \hat{\mathbf{m}})_l} \hat{\sigma}_l \tag{3.32}$$

$$= (\hat{\mathbf{n}} \cdot \hat{\mathbf{m}}) \mathbb{1} + \underbrace{(\hat{\mathbf{n}} \times \hat{\mathbf{m}}) \cdot \boldsymbol{\sigma}}_{\text{Traceless}} \tag{3.33}$$

To sum up,

$$\boxed{|\langle \hat{\mathbf{n}}|\hat{\mathbf{m}}\rangle|^2 = \frac{1+\hat{\mathbf{n}}\cdot\hat{\mathbf{m}}}{2}}, \tag{3.34}$$

relating the geometry of the Hilbert space and the Euclidean space. In other words, if the angle between the Euclidean vectors $\hat{\mathbf{n}}$ and $\hat{\mathbf{n}}$ is Θ, then

$$|\langle \hat{\mathbf{n}}|\hat{\mathbf{m}}\rangle|^2 = \frac{1+\cos\Theta}{2} = \cos^2\left(\frac{\Theta}{2}\right). \tag{3.35}$$

Notice the angle doubling effect again.

SECTION 3.3

Quantum dynamics

$\frac{\mathrm{d}}{\mathrm{d}t}$ acts linearly on the state $|\psi\rangle$, so we should be able to represent it by a linear operator $\hat{G}$. What kind of operator $\hat{G}$ is for the time evolution to preserve the inner product?

$$0 = \frac{\mathrm{d}}{\mathrm{d}t}\langle\psi|\psi\rangle = \left(\frac{\mathrm{d}}{\mathrm{d}t}\langle\psi|\right)|\psi\rangle + \langle\psi|\left(\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle\right) \tag{3.36}$$

$$= \langle\psi|\hat{G}^\dagger|\psi\rangle + \langle\psi|\hat{G}|\psi\rangle \tag{3.37}$$

$$= \langle\psi|(\hat{G}^\dagger + \hat{G})|\psi\rangle \tag{3.38}$$

implying that $\widehat{G}$ is anti-Hermitian, $\hat{G}^\dagger = -\hat{G}$. The dimension of $\widehat{G}$ should be that of $[\text{time}]^{-1}$. Physicists set it to be $\widehat{G} = \widehat{H}/i\hbar$, where $\widehat{H}$ is a Hermitian *Hamiltonian operator*.[17]

[17] *This "derivation" of the Schrödinger equation contains no physics, so we cannot expect it to give the value of the constant $\hbar$.*

Definition 16 **(Schrödinger equation)**

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi(t)\rangle = \widehat{H}|\psi(t)\rangle \tag{3.39}$$

The Schrödinger equation relates an instantaneous change in the state vector to the Hamiltonian operator. A more "global", operator-version of Schrödinger equation is also illuminating when $\widehat{H}$ does not depend on time,[18]

[18] *This requires in particular that $\widehat{H}$ at different times commute with itself. Otherwise, we need to use time-ordered exponentials.*

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}\widehat{U}(t)|\psi(0)\rangle = \widehat{H}\widehat{U}(t)|\psi(0)\rangle \tag{3.40}$$

$$i\hbar\frac{\mathrm{d}U(t)}{\mathrm{d}t} = \widehat{H}\widehat{U}(t) \tag{3.41}$$

$$\frac{\mathrm{d}\widehat{U}(t)}{\widehat{U}(t)} = -\frac{i}{\hbar}\widehat{H}\mathrm{d}t \tag{3.42}$$

$$\boxed{\widehat{U}(t) = e^{-i\widehat{H}t/\hbar}}. \tag{3.43}$$

The meaning of this strange-looking matrix different equation is that we have differential equations for the eigenvalues in the basis in which $\widehat{U}$ is diagonal.

SECTION 3.4

*Looking ahead

The Schrödinger equation applies to a closed system. “Closed" here does not mean the absence of an energy or mass exchange with the surroundings; it means the absence of *communication* with the outside world. For an *open quantum system*, the quantum formalism has to be modifed as follow.

$$\boxed{\begin{aligned} \text{State} &\longleftrightarrow \text{Positive operator} && (3.44)\\ \text{Measurement} &\longleftrightarrow \text{Positive operator} && (3.45)\\ \text{Dynamics} &\longleftrightarrow \text{Completely positive map} && (3.46) \end{aligned}}$$

The meaning of complete positivity is outside the scope of this lecture.

In any case, it is more accurate to say that quantum phenomena happen to systems that are *informationally isolated*. Therefore, the notion of information is already baked into the formalism of quantum theory at a fundamental level. In my opinion, Rolf Landauer’s slogan “information is physical" is at its most profound in the quantum world.

CHAPTER 4

Quantum Information and Ebits (Entanglement Bits)

SECTION 4.1

Tensor product

The quantum state of a joint system AB lives in the tensor product space $\mathcal{H}_A \otimes \mathcal{H}_B$. The tensor product space can be defined without relying on specific bases on $\mathcal{H}_A$ and $\mathcal{H}_B$, but for the sake of simplicity, we will define the tensor product via ONBs.

Suppose that $\{|e_j\rangle\}_j$ and $\{|f_k\rangle\}_k$ are ONBs for $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. (Their dimensions may not be the same.) $\mathcal{H}_A \otimes \mathcal{H}_B$ is simply the span (formal linear combinations) of elements in the Cartesian product $\{|e_j\rangle\}_j \times \{|f_k\rangle\}_k$, denoting the elements by $|e_j\rangle \otimes |f_k\rangle$.

A vector in $\mathcal{H}_A \otimes \mathcal{H}_B$ is said to be a **product state** if it can be written as a simple product $|\psi\rangle \otimes |\phi\rangle$ of some $|\psi\rangle \in \mathcal{H}_A$ and $|\phi\rangle \in \mathcal{H}_B$. Otherwise, a vector is said to represent an **entangled state**. If needed, subscripts may be added to indicate which vector belongs to which Hilbert space, for example, $|\psi\rangle_A \otimes |\phi\rangle_B$. It is common to omit the tensor product symbol $\otimes$ and write a product state simply as $|\psi\rangle\,|\phi\rangle$, or even $|\psi\phi\rangle$ when no confusion may arise. The latter is extremely common when the qubit basis states are labeled by binary numbers, for example $|00\rangle = |0\rangle \otimes |1\rangle$, $|01\rangle = |0\rangle \otimes |1\rangle$, and so on.)

Scalar multiplication

$$\lambda(|\psi\rangle \otimes |\phi\rangle) = (\lambda\,|\psi\rangle) \otimes |\phi\rangle = \lambda(|\psi\rangle \otimes (\lambda\,|\phi\rangle)) \tag{4.1}$$

Vector addition

$$(|\psi_1\rangle + |\psi_2\rangle) \otimes |\phi\rangle = |\psi_1\rangle \otimes |\phi\rangle + |\psi_2\rangle \otimes |\phi\rangle \tag{4.2}$$

$$|\psi\rangle \otimes (|\phi_1\rangle + |\phi_2\rangle) = |\psi\rangle \otimes |\phi_1\rangle + |\psi\rangle \otimes |\phi_2\rangle \tag{4.3}$$

Compare these to the rules A.11 and A.12 for the direct product (which is equivalent to the direct sum) of vector spaces.

Linear combinations of product vectors with no common factor (entangled states) are genuinely new objects in $\mathcal{H}_A \otimes \mathcal{H}_B$.

Inner product.

$$(\langle\eta| \otimes \langle\xi|)(|\psi\rangle \otimes |\phi\rangle) = \langle\eta|\psi\rangle\,\langle\xi|\phi\rangle \tag{4.4}$$

Partial inner product.

$${}_A\langle\eta|\,(|\psi\rangle_A \otimes |\phi\rangle_B) = \langle\eta|\psi\rangle\,|\phi\rangle_B \tag{4.5}$$

Linear operators.

$$\widehat{A} \otimes \widehat{B}(|\psi\rangle \otimes |\phi\rangle) = (\widehat{A}\,|\psi\rangle) \otimes (\widehat{B}\,|\phi\rangle) \tag{4.6}$$

$$(\widehat{A} \otimes \widehat{B})(\widehat{C} \otimes \widehat{D}) = (\widehat{A}\widehat{C}) \otimes (\widehat{B}\widehat{D}) \tag{4.7}$$

Thus, the trace factorizes.

$$\mathrm{tr}\left(\widehat{A} \otimes \widehat{B}\right) = \sum_{jk} \langle e_j| \langle f_k|\, \widehat{A} \otimes \widehat{B}\, |e_j\rangle\, |f_k\rangle \tag{4.8}$$

$$= \sum_{jk} \langle e_j|\widehat{A}|e_j\rangle\ \langle f_k|\widehat{B}|f_k\rangle \tag{4.9}$$

$$= \mathrm{tr}\left(\widehat{A}\right) \otimes \mathrm{tr}\left(\widehat{B}\right) \tag{4.10}$$

Local operations always commute.

$$[\widehat{A} \otimes \widehat{\mathbb{1}}, \widehat{\mathbb{1}} \otimes \widehat{B}] = \widehat{A} \otimes \widehat{B} - \widehat{A} \otimes \widehat{B} = 0 \tag{4.11}$$

If one wants to represent a tensor product operator in a matrix form, one needs to fix the ordering of the bases of $\mathcal{H}_A \otimes \mathcal{H}_B$.[19] When the lexicographic ordering is chosen, the matrix form has the form of the Kronecker product, which only shown here for the 2-by-2 case:

$$\widehat{A}\widehat{B} \longleftrightarrow \left(\begin{array}{c:c} A_{00}B & A_{01}B \\ \hdashline A_{10}B & A_{11}B \end{array}\right). \tag{4.12}$$

[19] *By doing this, we are representing higher-rank tensors as a two-tensor (a matrix), and there is no canonical way to do it. It is more natural to represent them as they are, using quantum circuit diagrams or tensor diagrams [3].*

The concept of tensor product is typically introduced to studentss in a highly unintuitive setting of quantum theory, and as a result, the idea may appear alien at first. However, the tensor product is already present in ordinary probability theory as a self-evident means of combining the probabilities of two independent random variables.

$$\begin{pmatrix} p \\ 1-p \end{pmatrix} \otimes \begin{pmatrix} q \\ 1-q \end{pmatrix} = \begin{pmatrix} pq \\ p(1-q) \\ (1-p)q \\ (1-p)(1-q) \end{pmatrix} \tag{4.13}$$

In ordinary probability theory, a state that cannot be written as a product state $|p\rangle \otimes |q\rangle$ is a **correlated state**. Entanglement is a form of correlation, but we will see in Section A that it can be stronger than any classical correlation.

SECTION 4.2

Communication using entanglement

The four **Bell states** are defined to be

$$|\Phi_\pm\rangle \equiv \frac{|00\rangle \pm |11\rangle}{\sqrt{2}}, \tag{4.14}$$

$$|\Psi_\pm\rangle \equiv \frac{|01\rangle \pm |10\rangle}{\sqrt{2}}. \tag{4.15}$$

It is straightforward to verify that they form an ONB. Therefore, they constitute a basis measurement called the **Bell measurement**.[20] Alternatively, they can be arranged in a convenient form,

$$|\Omega_{ab}\rangle \equiv \hat{\mathbb{1}} \otimes \hat{X}^a \hat{Z}^b |\Omega\rangle\,, \tag{4.16}$$

Parity bit — Phase bit

obtained from $|\Omega\rangle \equiv |\Phi_+\rangle$ via a local Pauli operation. In particular,

$$|\Omega_{00}\rangle = \hat{\mathbb{1}} \otimes \hat{\mathbb{1}} |\Omega\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \tag{4.17}$$

$$|\Omega_{01}\rangle = \hat{\mathbb{1}} \otimes \hat{Z} |\Omega\rangle = \frac{|00\rangle - |11\rangle}{\sqrt{2}}, \tag{4.18}$$

$$|\Omega_{10}\rangle = \hat{\mathbb{1}} \otimes \hat{X} |\Omega\rangle = \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \tag{4.19}$$

$$|\Omega_{11}\rangle = \hat{\mathbb{1}} \otimes \hat{X}\hat{Z} |\Omega\rangle = \frac{|01\rangle - |10\rangle}{\sqrt{2}}. \tag{4.20}$$

[20] *The Bell measurement is a prime example of a basis measurement that is not conventionally associated to any observable (i.e. a Hermitian operator).*

Lemma 9 For any linear operator $\hat{T}$,

$$\hat{\mathbb{1}} \otimes \hat{T} |\Omega\rangle = \hat{T}^T \otimes \hat{\mathbb{1}} |\Omega\rangle\,, \tag{4.21}$$

where ^{T} denotes the matrix transposition. In particular, we have that

$$|\Omega_{11}\rangle = \hat{\mathbb{1}} \otimes \hat{X}\hat{Z} |\Omega\rangle = \hat{Z}\hat{X} \otimes \hat{\mathbb{1}} |\Omega\rangle\,. \tag{4.22}$$

In the followings, we suppress the tensor product symbol when writing two-body Pauli operators: $\hat{\sigma}_j\hat{\sigma}_k = \hat{\sigma}_j \otimes \hat{\sigma}_k$. Since

$$\hat{Z}\hat{Z} |x_1\rangle |x_2\rangle = (-1)^{x_1 \oplus x_2} |x_1\rangle |x_2\rangle\,, \quad \hat{X}\hat{X} |x_1 x_2\rangle = (-1)^{x_1 \oplus x_2} |x_1 \oplus 1\rangle |x_2 \oplus 1\rangle\,, \tag{4.23}$$

we can see that $\hat{Z}\hat{Z}$ measures the parity bit and $\hat{X}\hat{X}$ measures the phase bit,

$$\hat{Z}\hat{Z} |\Omega_{ab}\rangle = (-1)^a |\Omega_{ab}\rangle\,, \tag{4.24}$$

$$\hat{X}\hat{X} |\Omega_{ab}\rangle = (-1)^b |\Omega_{ab}\rangle\,. \tag{4.25}$$

For $\hat{Y}\hat{Y}$, note that

$$-\hat{Y}\hat{Y} = (\hat{Z}\hat{Z})(\hat{X}\hat{X}). \tag{4.26}$$

Therefore

$$-\widehat{Y}\widehat{Y}\,|\Omega_{ab}\rangle = (-1)^{a\oplus b}\,|\Omega_{ab}\rangle\,. \tag{4.27}$$

Special attention should be paid to the **singlet state** $|\Psi_-\rangle = (|01\rangle - |10\rangle)/\sqrt{2}$ with the property that simultaneous spin measurements on the two qubits in the same direction are always *anti-correlated*. In particular, the state can be written as

$$\frac{|\hat{\mathbf{n}}\rangle\,|-\hat{\mathbf{n}}\rangle - |-\hat{\mathbf{n}}\rangle\,|\hat{\mathbf{n}}\rangle}{\sqrt{2}} \tag{4.28}$$

The singlet state can be prepared, for instance, via the process of spontaneous parametric down-conversion (SPDC) that produces a photon pair with opposite polarizations.

Example (**Superdense coding**)
Equation (4.16) implies that if Alice and Bob possess an entangled pair $|\Omega\rangle$, Alice can communicate two bits of information (the parity bit and the phase bit) to Bob by choosing to perform a Pauli operation on her side and send the qubit to Bob. In the form of *Bennett's law*, we state that

$$1 \text{ ebit} + 1 \text{ qubit} \succ 2 \text{ bits}, \tag{4.29}$$

where $\succ$ means "can perform a task in place of".

Example (**Teleportation**)
The "converse" to (4.29),

$$1 \text{ ebit} + 2 \text{ bits} \succ 1 \text{ qubit}, \tag{4.30}$$

is *quantum teleportation*. Alice and Bob again begins the protocol with a pre-shared entanglement $|\Omega\rangle$, but now Alice has an unknown state $|\psi\rangle$ that she wants to send to Bob. Rather than sending the qubit directly (which would need a high-fidelity quantum channel, for example), what Alice can do is performing the Bell measurement on her unknown state and one half of the Bell pair. Remarkably, depending on the two-bit outcome ab that Alice obtained, the post-measurement state on Bob's side becomes

$$|\psi'\rangle = \widehat{X}^b\widehat{Z}^a\,|\psi\rangle\,, \tag{4.31}$$

which is not exactly the state $|\psi\rangle$ that Alice wants to send, but the extra Pauli operators can be corrected away with the information of the bits ab that Alice can tell Bob by calling him on a phone, for example.

Proofs of the teleportation protocol found in textbooks are often quite long and involve some unintuitive steps. The following short proof is inspired from the tensor diagrammatic proof [3]. Let the subscripts A, B, and C denote the Hilbert space of Alice's unknown qubit state, the Hilbert space of one half of the Bell pair that Alice holds, and the Hilbert space of the other half of the Bell

pair that Bob holds, respectively.

$$ {}_{AB}\langle\Omega_{ab}|\left(|\psi\rangle_A|\Omega_{00}\rangle_{BC}\right) = \frac{1}{2}\left[\sum_j \langle j|_A \langle j|_B (\widehat{Z}^a\widehat{X}^b)^\dagger_A \otimes \widehat{\mathbb{1}}_B\right]|\psi\rangle_A \sum_k |k\rangle_B |k\rangle_C \tag{4.32} $$

$$ = \frac{1}{2}\sum_{jk}\langle j|_A \underbrace{\widehat{X}^b\widehat{Z}^a|\psi\rangle_A}_{|\psi'\rangle}\underbrace{\langle j|_B|k\rangle_B}_{\delta_{jk}}|k\rangle_C \tag{4.33} $$

$$ = \frac{1}{2}\sum_j \langle j|\psi'\rangle|j\rangle_C \tag{4.34} $$

$$ = \frac{1}{2}|\psi'\rangle_C \tag{4.35} $$

The subnormalization factor $1/2$ tells us that the outcome probability for the Bell measurement is uniform.

SECTION 4.3
EPR argument

A philosophical troubling aspect of quantum theory is that one cannot in general think of an act of measuring as revealing a pre-existing value of the measured property. When the system is in an eigenstate of an observable $\hat{A}$ with an eigenvalue λ, subsequent measurements of $\hat{A}$ do not alter the state, hence there is a tendency to think of the value λ as pre-existed. Einstein, Padolsky, and Rosen (EPR) devised a clever argument using quantum correlation and the principle of relativity to argue that values of incompatible (non-commuting) observables can simultaneously pre-exist.

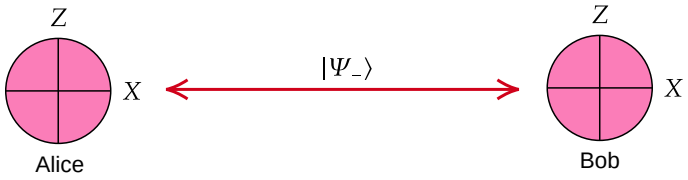

Figure 6. Measurement settings in the EPR argument.

The singlet state $|\Omega_{11}\rangle$ is an eigenstate shared by $\widehat{Z}\widehat{Z}$ and $\widehat{X}\widehat{X}$ with both eigenvalues -1. (Verify that they commute.) Now, while Alice cannot make a local Z and X measurements at the same time, if she chooses to measure one, say $\widehat{Z}_A$ and find the value $z_A = \pm 1$, then the spin of Bob's particle would need to have the opposite value to satisfy $z_A z_B = -1$. Since the spins are anti-correlated in every direction in the singlet stat, the same conclusion follows if Alice were to measure $\widehat{X}_A$. But, EPR argued, the act of measurement by Alice over here cannot effect the state of Bob's particle over there. Thus, the fact that Alice could have chosen to measure either observable and inferred z_B or x_B without

disturbing Bob's particle means that those values already existed before the measurement.

$$\begin{pmatrix} \textbf{Entanglement} \\ \text{Anti-correlation} \\ \text{in the singlet state} \end{pmatrix} + \begin{pmatrix} \textbf{Relativity} \\ \text{The choice of} \\ \text{measurement on} \\ \text{A cannot have an} \\ \text{influence on B} \end{pmatrix} \Longrightarrow \begin{pmatrix} \textbf{Quantum theory} \\ \textbf{is incomplete} \end{pmatrix}$$

The EPR correlation in the Z and X measurement outcomes can be explained by classical hidden variables, in particular, Spekkens' toy theory [5]. See also [6] for diagrams similar to Figures 6 and 8.

SECTION 4.4

CHSH inequality

The hope that quantum theory can be completed with *hidden variables* that behave entirely classically is dashed by an experimental violation of the CHSH inequality.[21] Here we present a derivation of the CHSH inequality in the setting of a communication game, called a *nonlocal game* in this context [4].

[21] *Named after John Clauser, Michael Horne, Abner Shimony, and Richard Hol, the inequality is formulated to be more suitable for experimental tests, an improvement over Bell's inequality.*

Consider a communication game involving two players, Alice and Bob. In this cooperative game, Alice and Bob initially have the opportunity to consult and share their strategies before separating and traveling to two distant referees. The players are allowed to communicate in advance, but once the game begins, they are unable to communicate due to the restriction imposed by the speed of light.

Game structure. At the referee stations, two binary questions are randomly chosen, denoted by $x \in \{0,1\}$ for Alice's question and $y \in \{0,1\}$ for Bob's question. Alice and Bob must give answers immediately, represented by variables $a \in \{0,1\}$ and $b \in \{0,1\}$, respectively. Because of the random choices, neither player knows the question in advance, and they cannot communicate the questions they received fast enough to change their answers. The referees then record the answers and later compare notes.

Winning Condition. The goal for Alice and Bob to win the game is for the logical AND of their answers $a \wedge b$ to be equal to the product of the choices of the questions xy. Here, $a \wedge b$ represents whether the answers agree (0) or disagree (1). In other words, they win if the following condition is satisfied:

$$xy = a \wedge b. \tag{4.36}$$

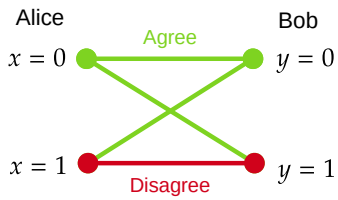

Figure 7. Diagram illustrating the rule of the nonlocal game.

The only instance where Alice and Bob's answers must disagree is when $x = y = 1$, see Figure 7. The agreement-disagreement relationship between the questions and answers forms a *frustration graph*. It becomes evident that it is impossible to assign simultaneous edge values (a and b) to satisfy the winning condition. In the case of a deterministic strategy where Alice and Bob always answer yes or no, the game can be won with a probability of 3/4. Remarkably, this represents the best achievable winning probability for Alice and Bob. (The

winning probability for mixed strategies is at most equal to that of deterministic strategies.) This bound on the winning probability

$$\Pr_{\substack{\text{win}\\ \text{classical}}} \leq \frac{3}{4}. \tag{4.37}$$

is one form of the CHSH inequality; it signifies the limit of any classical local hidden variable theory.

$$\left(\begin{array}{c} \textbf{Classical} \\ \textbf{correlation} \end{array} \right) + \left(\begin{array}{c} \textbf{Relativity} \\ \text{The choice of} \\ \text{measurement on} \\ \text{A cannot have an} \\ \text{influence on B} \end{array} \right) \Longrightarrow (\textbf{CHSH inequality})$$

Now we show that if Alice and Bob have shared the singlet state beforehand, they can win the game with a probability greater than what is permissible by the CHSH bound.

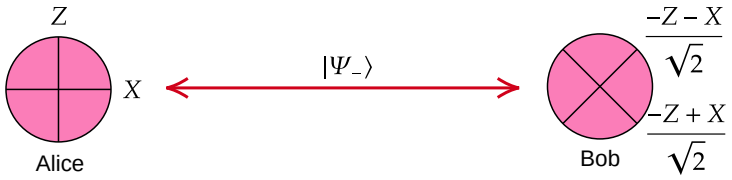

Figure 8. Measurement settings that leads to a violation of the CHSH inequality given that Alice and Bob pre-shared the singlet state.

$$\Pr_{\text{win}} = \frac{1}{4}(p_{00}^{(+)} + p_{01}^{(+)} + p_{10}^{(+)} + p_{11}^{(-)}) \tag{4.38}$$

$$= \frac{1}{4}\left\langle \widehat{P}_{00}^{(+)} + \widehat{P}_{01}^{(+)} + \widehat{P}_{10}^{(+)} + \widehat{P}_{11}^{(-)} \right\rangle \tag{4.39}$$

Solve for $\widehat{P}^{(\pm)}$ in terms of the observables.

$$\begin{cases} \widehat{P}^{(+)} + \widehat{P}^{(-)} & = \widehat{\mathbb{1}}, \\ \widehat{P}^{(+)} - \widehat{P}^{(-)} & = \widehat{\sigma}_{xy}, \end{cases} \Longrightarrow \widehat{P}^{(\pm)} = \frac{\widehat{\mathbb{1}} + \widehat{\sigma}_{xy}}{2} \tag{4.40}$$

Thus,

$$\Pr_{\text{win}} = \frac{1}{2} + \frac{1}{8}\left\langle \widehat{\sigma}_{00} + \widehat{\sigma}_{01} + \widehat{\sigma}_{10} - \widehat{\sigma}_{11} \right\rangle. \tag{4.41}$$

Specialize to the measurement angles given in...

$$\Pr_{\text{win}} = \frac{1}{2} + \frac{1}{8}\left\langle \widehat{Z} \otimes \frac{-\widehat{Z}+\widehat{X}}{\sqrt{2}} + \widehat{Z} \otimes \frac{-\widehat{Z}-\widehat{X}}{\sqrt{2}} + \widehat{X} \otimes \frac{-\widehat{Z}+\widehat{X}}{\sqrt{2}} - \widehat{X} \otimes \frac{-\widehat{Z}-\widehat{X}}{\sqrt{2}} \right\rangle \tag{4.42}$$

$$= \frac{1}{2} + \frac{1}{8\sqrt{2}}\left\langle \widehat{Z}\widehat{Z} + \widehat{X}\widehat{X} \right\rangle = \frac{1}{2} + \frac{\sqrt{2}}{4} \approx 85\% > \frac{3}{4} \geq \Pr_{\substack{\text{win} \\ \text{classical}}}, \tag{4.43}$$

where the expectation value is taken with respect to the singlet state.

The original form of the CHSH inequality, $S_{\text{classical}} \leq 2$, is phrased in terms of expectation values, or two-point correlation functions,

$$S = |\langle \widehat{\sigma}_{00} \rangle + \langle \widehat{\sigma}_{01} \rangle + \langle \widehat{\sigma}_{10} \rangle - \langle \widehat{\sigma}_{11} \rangle|. \tag{4.44}$$

The maximum violation allowed by quantum theory is $S = 2\sqrt{2}$, which is what we achieved in the computation above, called the *Tsirelson bound*. The winning probability of 1 corresponds to the value $S = 4$, which is achievable in non-quantum theories that, surprisingly enough, still do not violate special relativity (called *non-signaling theories*) [7]. However, there can be some highly unlikely consequences in a universe in which such *super-quantum* theory is obeyed such as trivial communication complexity [8].

CHAPTER A

Linear Algebra

SECTION A.1
Vector space, basis, and dimensionality

To define vector space axiomatically, it is convenient to define an abelian group first.

Definition 17

A **group** G is a set closed under an associative binary operation $\cdot$ (*group multiplication*) satisfying the following properties.

1. There exists an **identity** element e such that $e \cdot g = g \cdot e = g$ for every element $g \in G$.
2. For every $g \in G$, there exists an **inverse** denoted by g^{-1} such that $g \cdot g^{-1} = g^{-1} \cdot g = e$.

We often omit the symbol $\cdot$ and simply write a product $g \cdot h$ as gh. A group in which the order of multiplication doesn't matter ($gh = hg$) is called a **commutative** group or an **abelian** group.

Example $(\mathbb{Z}, +)$ (The group of integers with addition as group multiplication) is an abelian group with 0 as the identity and $-a$ as the inverse of a.

Example $(\mathbb{Z}, -)$ (The group of integer with subtraction as group multiplication) is **not** a group, since subtraction is not associative ($a - (b - c) \neq (a - b) - c$).

Example $(\mathbb{C}/\{0\}, \times)$ (The group of complex numbers excluding zero with multiplication) is an abelian group with 1 as the identity and $1/a$ as the inverse of a.

Example $(\{\pm 1, \pm i\}, \times)$ is an abelian group with 1 as the identity. -1 is its own inverse, and $\pm i$ is an inverse of each other.

Example $(\mathrm{GL}(n), \times)$ (The group of invertible matrices with matrix multiplication) is a non-abelian group with the identity matrix $\mathbb{1}$ as the identity and the matrix inverse A^{-1} as the inverse.

Example **(Quantum information example)** $(\{\pm\mathbb{1}, \pm iX, \pm iY, \pm iZ\}, \times)$, where

$$X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \qquad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{A.1}$$

are *Pauli matrices*, forms a non-abelian group with $\mathbb{1}$ as the identity and $\pm iX$ is an inverse of each other, for example.

Armed with the definition of an abelian group, we now proceed to define a vector space over the number field $K = \mathbb{R}$ (reals) or $\mathbb{C}$ (complex numbers).

Definition 18 A **vector space** V over K is a set closed under the following operations.

1. **Vector addition:** $u + v$, where $u, v \in V$,
2. **Scalar multiplication:** au, where $a \in K$ and $u \in V$,

such that $(V, +)$ is an abelian group, and

$$1u = u, \tag{A.2}$$
$$a(u + v) = au + av, \tag{A.3}$$
$$(a + b)u = au + bu, \tag{A.4}$$
$$(ab)u = a(bu), \tag{A.5}$$

where $u, v \in V$ and $a, b \in K$

The three-dimensional Euclidean space $\mathbb{R}^3$ is a prototypical example of a vector space over $\mathbb{R}$. As a set, $\mathbb{R}^3$ contains infinitely many vectors, but a finite representative set $\{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$, called a *basis*, is sufficient to represent any vector in $\mathbb{R}^3$. To define a basis, we first define the notion of a span.

Definition 19 Given a set $S = \{u_1, u_2, \ldots, u_n\} \subset V$ of a finite number of vectors, span S is the set of all linear combination (L.C.) of these vectors

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n, \qquad \forall a_j \in K. \tag{A.6}$$

A vector space V is said to be **finite dimensional** if it is spanned by some finite set of vectors in V. Otherwise, it is **infinite dimensional**.

Definition 20 A set $S = \{u_1, u_2, \ldots, u_n\} \subset V$ is **linearly independent (L.I.)** if the equation

$$a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0 \tag{A.7}$$

implies that $a_j = 0$ for all $j = 1, \ldots, n$.

In other words, linear independence means that no vector in S can be written as a non-trivial L.C. of other vectors in S. Otherwise, if $a_1 \neq 0$, for instance, then

$$u_1 = -(a_2u_2 + \cdots + a_nu_n)/a_1. \tag{A.8}$$

Definition 21 $S \subset V$ is a **basis** for V if it is L.I. and spans V.

The size of S is an invariant of a vector space called the **dimension**, denoted by $\dim V$. The dimension is the minimum numbers of vectors that spans V. For any nontrivial (not $\{0\}$) vector space, there are infinitely many choices of bases.

Example The field K itself is a one-dimensional vector space over K.

Example K^n is an n-dimensional vector space over K, with the standard basis $\{u_j\}_{j=1,\ldots,n}$, where u_j is a vector with 1 at the jth entry and 0 elsewhere.

In the two-dimensional case, the followings are equivalent fot a set $S = \{(a_1, a_2), (b_1, b_2)\}$.

- S is L.I.
- S spans K^2.
- The determinant $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} = a_1 b_2 - a_2 b_1 \neq 0$.

Example The set of $n \times m$ K-valued matrices is an $n \times m$-dimensional vector space with the basis $\{E_{jk}\}_{j=1,\ldots,n}^{k=1,\ldots,m}$, where E_{jk} is a matrix with 1 at the jkth entry and 0 elsewhere.

Example Polynomials over an indeterminate x form an infinite-dimensional vector space. Restricting to n-degree polynomials gives an $(n+1)$-dimensional vector space with a basis $\{1, x, x^2, \ldots, x^n\}$.

Example The set of solutions of a linear, kth-order differential equation, for example,

$$f'' + f' + f = 0$$

for $k = 2$, forms a vector space of dimension k.

By choosing a basis, every vector in a finite-dimensional vector space can be specified by its **components** v_j defined as.

$$v = \sum_j v_j u_j, \tag{A.9}$$

which is typically stacked into a column vector

$$v \leftrightarrow \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}. \tag{A.10}$$

The arrow $\leftrightarrow$ is there to remind us that the components constitute only a *representation* of the vector. The vector itself is a geometric object (visualized as an arrow, for example) that exists independent of the choice of basis, whereas the components change when we perform a *change of basis*. (More on that later.)

A subset $S \subset V$ of a vector space that is also a vector space is called a **subspace**. A vector space U is said to be a **direct sum** $V \oplus W$ if every $u \in U$ can be decomposed uniquely as a sum of $v \in V$ and $w \in W$. Equivalently, any $u \in U$ can be specified by a unique pair $(v, w) \in V \times W$ such that[22]

$$(v_1, w_1) + (v_2, w_2) = (v_1 + v_2, w_1 + w_2), \tag{A.11}$$

$$a(v, w) = (av, aw). \tag{A.12}$$

[22] *Hence for vector spaces, the direct sum and the direct product are the same. Compare properties A.11 and A.12 to the algebraic rules for the tensor product.*

SECTION A.2

Linear maps

Morphisms that preserve the structure of vector spaces are linear maps.

Definition 22 Given a map $T : U \to V$, T is said to be a **linear map** if

$$T(au + bv) = aT(u) + bT(v). \tag{A.13}$$

For now, we will focus on linear maps between the same input and output vector space, in which case linear maps are also called **linear operator**, or just operators in short. Just keep in mind that linear maps between different vector spaces are no less important. For instance, we will see in a moment that a bra $\langle u|$ is a linear map from a vector space over a field K to K itself, a one-dimensional vector space.

Example Left shift

$$(c_1, c_2, c_3, \dots) \mapsto (c_2, c_3, c_4, \dots), \tag{A.14}$$

and right shift

$$(c_1, c_2, c_3, \dots) \mapsto (0, c_1, c_2, \dots), \tag{A.15}$$

are linear operators in an infinite-dimensional vector space.

Example Differentiation

$$\frac{\mathrm{d}}{\mathrm{d}x}(af + bg) = a\frac{\mathrm{d}f}{\mathrm{d}x} + b\frac{\mathrm{d}g}{\mathrm{d}x} \tag{A.16}$$

is a linear operator in the space of differentiable functions.

Example The trace $\mathrm{tr}(\hat{A}) = \sum_j \langle e_j|\hat{A}|e_j\rangle$ is a linear map from the vector space of $L(\mathcal{H})$ of linear operators over $\mathcal{H}$ to $\mathbb{C}$.

As vectors can be specified by their components, so too can linear operators by their "matrix elements". Pick a basis $\{u_1, u_2, \dots, u_n\}$ for U. $T(u_j)$ is another vector in U, and hence can be expanded using the same basis. The **matrix element** T_{jk} is defined as

$$T(u_j) = \sum_{k=1}^{n} T_{jk} u_k, \tag{A.17}$$

where j is the row index and k is the column index.

$$T \leftrightarrow \begin{pmatrix} T_{11} & T_{12} & \dots & T_{1n} \\ T_{21} & T_{22} & & \vdots \\ \vdots & & \ddots & \\ T_{n1} & \dots & & T_{nn} \end{pmatrix} \tag{A.18}$$

SECTION A.3

Eigenvalues and eigenvectors

From this point onward, suppose that $\widehat{T} : U \to U$ maps a vector space to the same space. That is, $\widehat{T}$ is a linear operator. The equation of the form

$$\widehat{T}\,|u\rangle = \lambda\,|u\rangle \tag{A.19}$$

is called an *eigenvalue equation*. The scalar λ is called an **eigenvalue** of the linear operator $\widehat{T}$, and $|u\rangle$ is called an **eigenvector** of $\widehat{T}$. It is straightforward to show that eigenvectors corresponding to the same eigenvalue form a subspace called an **eigenspace**.

Not all linear operators possess an eigenvalue. Take rotations in $\mathbb{R}^2$, for example,

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{A.20}$$

No vector with real coefficients remain fixed under a rotation by an arbitrary angle. However, if one allows for complex linear combinations, $\hat{\mathbf{x}} \pm i\hat{\mathbf{y}}$ are eigenvectors of the rotations. In fact, this example highlights a broader phenomenon when we work in an algebraically closed field.

Theorem 3 In an algebraically closed field such as $\mathbb{C}$, every linear operator always possess at least one eigenvalue (and eigenvector).

The followings is a determinant theorem can be proved without using the determinant as follows [9].

PROOF In an n-dimensional vector space V, pick any nonzero vector $|v\rangle \in V$ and consider the set $\{|v\rangle, \widehat{T}\,|v\rangle, \widehat{T}^2\,|v\rangle, \ldots, T^n\,|v\rangle\}$. Since the cardinality of the set is $n+1$, the set must be linearly *dependent*. There exist $c_0, c_1, \cdots, c_n$, not all of them zero, such that

$$0 = (c_o\widehat{\mathbb{1}} + c_1\widehat{T} + \cdots + c_n\widehat{T}^n)\,|v\rangle \tag{A.21}$$

$$= \alpha\prod_{k=1}^{d}(\widehat{T} - \lambda_k\widehat{\mathbb{1}})\,|v\rangle\,, \tag{A.22}$$

where in the second line we have used the fact that every polynomial can be factorized into linear factors over $\mathbb{C}$:

$$c_0 + c_1 z + c_2 z^2 + \cdots + c_n z^2 = \alpha\prod_{k=1}^{n}(z - \lambda_k). \tag{A.23}$$

Thus, some $\widehat{T} - \lambda_k\widehat{\mathbb{1}}$ must be the zero operator, and the input of such linear factor is an eigenvector of $\widehat{T}$. □

SECTION A.4

Proof of the spectral theorem for normal operators

The key property for a linear operator $\widehat{T}$ to be diagonalizable is that, if $S \subset V$ is an T-invariant subspace, then the orthogonal complement $S^\perp$ is also T-invariant. You might sort of already see how this is sufficient for diagonalization, because if $|v\rangle$ is a eigenvector $\widehat{T}$ (at least one of which always exist because of Theorem 3), then one can cut down V to a smaller subspace $S^\perp$ orthogonal to $|v\rangle$. But if $S^\perp$ is also T-invariant, then one can again find at least one eigenvector in $S^\perp$. This process can then be repeated until we find all the eigenvectors of $\widehat{T}$ which form an ONB for V.

Let $\widehat{P}$ be a projection operator onto S respectively. Then

$$S \text{ is } T\text{-invariant} \iff (\widehat{\mathbb{1}} - \widehat{P})\widehat{T}\widehat{P} = 0, \tag{A.24}$$

$$S^\perp \text{ is } T\text{-invariant} \iff \widehat{P}\widehat{T}(\widehat{\mathbb{1}} - \widehat{P}) = 0. \tag{A.25}$$

The goal is to show that (A.24) implies (A.25) if and only if $\widehat{T}$ is normal. In particular, (A.25) is the statement that $\widehat{X} \equiv \widehat{P}\widehat{T}(\widehat{\mathbb{1}} - \widehat{P})$ is unequivocally the zero operator, which we is equivalent to the vanishing of the Hilbert-Schmidt norm

$$\left\|\widehat{X}\right\|^2 = \mathrm{tr}\left(\widehat{X}^\dagger \widehat{X}\right) = 0. \tag{A.26}$$

$$\begin{aligned}
\mathrm{tr}\left(\widehat{X}^\dagger \widehat{X}\right) &= \mathrm{tr}\left[(\widehat{\mathbb{1}} - P)\widehat{T}^\dagger \widehat{P}\widehat{P}\widehat{T}(\widehat{\mathbb{1}} - \widehat{P})\right] && \text{(A.27)} \\
&= \mathrm{tr}\left[(\widehat{\mathbb{1}} - \widehat{P})^2 \widehat{T}^\dagger \widehat{P}^2 \widehat{T}\right] && \text{(A.28)} \quad \text{Cyclicity of the trace} \\
&= \mathrm{tr}\left[(\widehat{\mathbb{1}} - \widehat{P})\widehat{T}^\dagger \widehat{P}\widehat{T}\right] && \text{(A.29)} \quad \text{Property of projection operators} \\
&= \mathrm{tr}\left(\widehat{T}\widehat{T}^\dagger \widehat{P} - \widehat{T}^\dagger \underbrace{\widehat{P}\widehat{T}\widehat{P}}_{\widehat{T}\widehat{P} \text{ by (A.24)}}\right) && \text{(A.30)} \\
&= \mathrm{tr}\left[(\widehat{T}\widehat{T}^\dagger - \widehat{T}^\dagger \widehat{T})\widehat{P}\right]. && \text{(A.31)}
\end{aligned}$$

The last line vanishes if $\widehat{T}$ is normal, thus concluding the proof.

The singular value decomposition makes it obvious the significance of the condition $\widehat{T}\widehat{T}^\dagger = \widehat{T}^\dagger \widehat{T}$. For an arbitrary matrix $\widehat{T}$ (may be non-square), the SVD implies that

$$\widehat{T} = \sum_j \sigma_j \left|e_j\right\rangle\!\left\langle f_j\right|, \tag{A.32}$$

where the two orthonormal sets $\{|e_j\rangle\}_j$ and $\{|f_k\rangle\}_k$ are eigenvectors of $\widehat{T}^\dagger \widehat{T}$ and $\widehat{T}\widehat{T}^\dagger$, respectively, and σ_j are non-negative *singular values*. It is clear then, that the equality $\widehat{T}\widehat{T}^\dagger = \widehat{T}^\dagger \widehat{T}$ implies that the two orthonormal sets are the same i.e. it is the basis in which the matrix of $\widehat{T}$ is diagonal.

References

[1] N. P. Landsman, *Lecture notes on Hilbert spaces and quantum mechanics*

[2] Max Jammer, *The Conceptual Development of Quantum Mechanics*, 2nd ed., Tomash Publishers, American Institute of Physics (1989).

[3] Bob Coecke, *Quantum picturalism*, Contemporary Physics 51 59–83 (2009).

[4] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner, *Bell nonlocality*, Reviews of Modern Physics, **86**, 419 (2014).

[5] Robert W. Spekkens, *Evidence for the epistemic view of quantum states: A toy theory*, Physical Review A **75**, 032110 (2007).

[6] Christopher J. Wood and Robert W. Spekkens, *The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning*, New Journal of Physics **17**, 033002 (2015).

[7] Sandu Popescu and Daniel Rohrlich, *Quantum nonlocality as an axiom*, Foundations of Physics, **24**, 379–385 (1994).

[8] Wim van Dam, *Implausible consequences of superstrong nonlocality*, Natural Computing, **12**, 9–12 (2013).

[9] Sheldon Axler, *Linear Algebra Done Right*, 2nd ed., Springer (2004).