THE LOGIC IN COMPUTER SCIENCE COLUMN

BY

YURI GUREVICH

Computer Science & Engineering
University of Michigan, Ann Arbor, Michigan, USA
gurevich@umich.edu

Wigner's quasidistribution
and Dirac's kets

Andreas Blass

Mathematics

University of Michigan

Yuri Gurevich

Computer Science & Engineering

University of Michigan

Alexander Volberg

Mathematics, Michigan State University

Hausdorff Center for Mathematics, University of Bonn

Abstract

In every state of a quantum particle, Wigner's quasidistribution is the unique quasidistribution on the phase space with the correct marginal distributions for position, momentum, and all their linear combinations.

The only difference between a probabilistic classical world and the equations of the quantum world is that somehow or other it appears as if the probabilities would have to go negative . . . Okay, that's the fundamental problem. I don't know the answer to it, . . . if I try my best to make the equations look as near as possible to what would be imitable by a classical probabilistic computer, I get into trouble.

— *Richard Feynman,*
Simulating Physics with Computers, 1982 [7, p. 480]

1   Introduction

The story of negative probabilities starts with the 1932 article [13] by Eugene Wigner. In quantum mechanics, probability distributions of the position and momentum of a particle make physical sense but their joint distribution doesn't. Yet Wigner exhibited such a joint distribution. It had some desired properties. However, some of its values were negative. "But of course," wrote Wigner, "this must not hinder the use of it in calculations."

In 1987, Jacqueline and Pierre Bertrand proposed "a new derivation of Wigner's function based on the property of positivity of its integrals along straight lines in phase space" [1]. In 2014, in this Bulletin [2], we sketched a mathematical proof of a characterization of Wigner's quasidistribution as the unique quasidistribution on the phase space $\mathbb{R}^2$ that yields the correct marginal distributions not only for position and momentum but for all their linear combinations. In 2021, that sketch was developed into a complete proof that the characterization is valid in "nice" states, namely the states given by smooth functions with compact support [3].

In this paper, we prove that the characterization is valid in all states, with no exception. Furthermore, the new proof is simpler, conceptually and technically. In particular, the uniqueness is derived from a purely measure-theoretic observation that we prove in §3. The simplicity of the new proof gave us the idea to present it in this Bulletin. We made an effort to give our readers a comprehensible and maybe even enjoyable introduction to some foundational issues of science.

Quisani[1]**:** What are negative probabilities?

Authors: The axiomatic definition of probabilities readily generalizes to *quasiprobabilities*, or *signed probabilities*, where negative values are allowed [3]. Basically, you drop the requirement that probabilities take values in the real segment $[0, 1]$ and allow arbitrary real values.

Q: But what is the intuition behind negative probabilities? An urn cannot have -3 red balls.

A: We don't know. At this point, we find it more fruitful to think about what quasiprobabilities are good for.

---

[1]A former student of the second author.

Q: You once compared the generalization of probabilities to quasiprobabilities with the generalization of real numbers to complex ones. Complex numbers became indispensable, e.g., in solving algebraic equations. Are there important theoretical problems that have been solved using negative probabilities?

A: We don't know such problems, but we expect that quasiprobabilities will be used to solve theoretical problems. They are already used in practice.

Q: Yes, you mentioned quantum tomography in our 2014 conversation [2]. Being a software engineer, I realize the paramount value of practical applications. But today I would like you to address basic questions. Some of these basic questions you seemed to dodge during our 2014 conversation. For example, you spoke about marginal distributions not only for the position and momentum but also for all their linear combinations. But what are marginal distributions for linear combinations? You never defined them properly.

A: Addressing basic questions is fine, and we will define those marginal distributions. We will try to explain things the best we can.

Q: Do explain. But please take into account that, in the meantime, I was busy with computer engineering. I didn't have time to study quantum mechanics or measure theory.

A: Understood.

2 Preliminaries

2.1 Measures

We recall some basic definitions of measure theory.

A *measurable space M* is a pair (Ω, Σ) where Ω is a nonempty set and Σ a σ-algebra of subsets of Ω. In other words, Σ is a Boolean algebra closed under countable unions. Members of Σ are *measurable sets* of M.

`Example:` The real line $\mathbb{R}$ with the collection of real Borel sets, which is the least σ-algebra containing every open real interval (a, b). ◁

`Example:` The real plane $\mathbb{R}^2$ with the collection of Borel subsets of $\mathbb{R}^2$, which is the least σ-algebra containing every open rectangle $(a, b) \times (c, d)$. ◁

A *measure μ* on a nonempty set Ω is a function such that

1. the domain of μ is a σ-algebra of subsets of Ω known as μ-*measurable* sets,

2. μ assigns a real number or ∞ to each μ-measurable set, and

3. μ is *countably additive* which means that, for all pairwise disjoint measurable sets s_n, we have
$$\mu\left(\bigcup_{n=1}^{\infty} s_n\right) = \sum_{n=1}^{\infty} \mu(s_n).$$

   Since the union $\bigcup s_n$ is independent of the order of the sets s_n, so is the sum $\sum \mu(s_n)$. That implies absolute convergence by a well-known theorem of Riemann.

If μ does not take value ∞, then μ is *finite*. If μ has no negative values then it is *nonnegative*. Every measure we consider in this paper is either finite or nonnegative.

Example: The Lebesgue measure on Euclidean spaces $\mathbb{R}^k$ and finite-dimensional Hilbert spaces $\mathbb{C}^k$, called *length* in the case of $\mathbb{R}$, called *area* in the cases of $\mathbb{R}^2$ and $\mathbb{C}$, and called *volume* in the case of $\mathbb{R}^3$ and in general. A careful treatment of the Lebesgue measure, with all the necessary proofs, is somewhat involved [12, Chapter 11], but the definition itself is simple, and we give a version of it on the example of the interval $(0, 1)$ in $\mathbb{R}$.

An open set O in $(0, 1)$ is the disjoint union of its maximal intervals; define the length of O to be the sum of the lengths of its maximal intervals. For any set $s \subseteq (0, 1)$, the *outer measure $\lambda^*(s)$* of s is the infimum of the lengths of the open sets O that cover s. The *inner measure $\lambda_*(s)$* is defined as $1 - \lambda^*((0, 1) - s)$.

If $\lambda^*(s) = \lambda_*(s)$, then s is *Lebesgue measurable* and the outer (and also the inner) measure $\lambda^*(s)$ is called the *Lebesgue measure $\lambda(s)$*.

The Lebesgue measure on $\mathbb{R}$ is defined by applying this construction to intervals $(i, i + 1)$ for all integers i and adding the resulting measures if all the pieces are measurable.

Lemma 1. *If μ, ν are finite measures on a measurable space M, then their difference $(\mu - \nu)(s) = \mu(s) - \nu(s)$ is a finite measure on M.*

Proof. If measurable sets $s_1, s_2, \ldots$ are pairwise disjoint, then

$$(\mu - \nu)\bigcup_n s_n = \mu\bigcup_n s_n - \nu\bigcup_n s_n = \sum_n \mu(s_n) - \sum_n \nu(s_n)$$

$$= \sum_n (\mu(s_n) - \nu(s_n)) = \sum_n (\mu - \nu)(s_n).$$

The third equality holds because of absolute convergence. □

A function $f : \Omega_1 \to \Omega_2$ from a measurable space $M_1 = (\Omega_1, \Sigma_1)$ to a measurable space $M_2 = (\Omega_2, \Sigma_2)$ is *measurable for M_1, M_2* if the f-preimage of every measurable set in M_2 is measurable in M_1. If M_2 is the real line $\mathbb{R}$ or complex line $\mathbb{C}$ endowed with the σ-algebra of Borel sets, then f is a *measurable function on M_1*. ◁

2.2 $L^2(\mathbb{R})$, and test functions

If f is a measurable function on a Euclidean space $\mathbb{R}^k$ and μ is a nonnegative measure on $\mathbb{R}^k$, then

$$\int f d\mu = \int_{\mathbb{R}^k} f d\mu = \int_{\mathbb{R}^k} f(x) d\mu(x)$$

means the Lebesgue integral of f with respect to measure μ. For real-valued f, the integral is defined by approximating f with so-called *simple functions*, i.e. functions $g(x)$ taking only finitely many values v_i, each on a measurable set s_i. The integral $\int g(x) d\mu(x)$ is simply $\sum_i v_i \mu(s_i)$. If the supremum of the integrals of simple functions $g(x) \leqslant f(x)$ for all x coincides with the infimum of the integrals of simple functions $g(x) \geqslant f(x)$ for all x, then their common value is the integral $\int f d\mu$, in which case f is *integrable* with respect to measure μ. See details in Chapter 11 of [12].

For $\mathbb{C}$-valued functions f, just integrate the real and imaginary parts separately. It is easy to check that every bounded continuous function is integrable with respect to any finite measure μ.

If s is a measurable subset of $\mathbb{R}^k$, we let $\chi_s(x)$ be 1 for $x \in s$ and 0 otherwise. Then $\int_s f(x) d\mu(x)$ means $\int_{\mathbb{R}^k} \chi_s(x) f(x) d\mu(x)$.

Proviso. By default, Euclidean spaces $\mathbb{R}^k$ come with the Lebesgue measure. ◁

$L^2(\mathbb{R})$ is the Hilbert space of square integrable functions $\psi : \mathbb{R} \to \mathbb{C}$ with the inner product $\langle \psi \mid \varphi \rangle$ given by the (Lebesgue) integral $\int_{\mathbb{R}} \psi^*(x)\varphi(x)\, dx$.

Two $L^2(\mathbb{R})$ functions are considered equivalent if they differ only on a set of measure zero. Strictly speaking, $L^2(\mathbb{R})$ vectors are the equivalence classes. It is more convenient though to work with individual functions modulo the equivalence relation.

The forward Fourier transform $\mathcal{F}$ sends an $L^2(\mathbb{R})$ function $\psi(x)$ to

$$\widehat{\psi}(\xi) = \frac{1}{\sqrt{2\pi}} \int \psi(x)\, e^{-i\xi x}\, dx$$

provided that the integral exists. Similarly, the inverse Fourier transform $\mathcal{F}^{-1}$ sends a function $\varphi(\xi)$ to

$$\widecheck{\varphi}(x) = \frac{1}{\sqrt{2\pi}} \int \varphi(\xi) e^{i\xi x}\, d\xi,$$

Mathematically x and ξ are real variables. In applications, the dimension of ξ is the inverse of that of x so that ξx is a pure number. Here and in the rest of the paper, integrals are by default integrals over $\mathbb{R}$.

The forward and inverse Fourier transforms are defined also for functions of several variables. In particular, provided the integrals exist, we have

$$\widehat{f}(\xi, \eta) = \frac{1}{2\pi} \iint f(x, y)\, e^{-i(\xi x + \eta y)}\, dx\, dy,$$

$$\widecheck{g}(x, y) = \frac{1}{2\pi} \iint g(\xi, \eta) e^{i(\xi x + \eta y)}\, d\xi\, d\eta.$$

If μ is a finite measure on $\mathbb{R}$, its Fourier transform $\widehat{\mu}$ is an $\mathbb{R} \to \mathbb{C}$ function:

$$\widehat{\mu}(\zeta) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-ix\zeta} d\mu(\zeta).$$

Similarly, if ν is a finite measure on $\mathbb{R}^2$, its Fourier transform $\widehat{\nu}$ is an $\mathbb{R}^2 \to \mathbb{C}$ function:

$$\widehat{\nu}(\xi, \eta) = \frac{1}{2\pi} \int_{\mathbb{R}^2} e^{-i(\xi x + \eta y)}\, d\nu(x, y).$$

An $L^2(\mathbb{R})$ function $\psi(x)$ is a *Schwartz function* if it is infinitely differentiable

and if it and its derivatives rapidly approach zero when $x \to \pm\infty$ in the sense that, for all nonnegative integers j, k, we have

$$\lim_{x \to \pm\infty} \left| x^j \frac{d^k \psi(x)}{dx^k} \right| = 0.$$

Schwartz functions are also known as *test functions*.

The Fourier transform of a test function is a test function. By the Plancherel theorem [9, Theorem A.19], the Fourier transform $\mathcal{F}$ is a unitary operator on the test functions. But these functions are dense in $L^2(\mathbb{R})$. By continuity, $\mathcal{F}$ is (or rather extends to) a unitary operator on the whole $L^2(\mathbb{R})$.

2.3 Distributions, and exponential operators

Dirac introduced a "function" $\delta(x)$ which is identically zero for all $x \neq 0$ while $\delta(0)$ is infinite, so infinite that $\int_{\mathbb{R}} \delta(x)dx = 1$. This makes no sense. δ is not a function in the usual sense. But it does make sense in the context of integrals of the form $\int_{\mathbb{R}} f(x)\delta(x)dx$ which should be $f(0)$, at least for well behaved functions f. Laurent Schwartz suggested viewing Dirac's δ and similar "generalized functions" as linear functionals on the space of test functions. Thus, Dirac's δ-function would be thought of as the linear functional on test functions f:

$$f \mapsto \int f(x)\delta(x)dx = f(0). \tag{1}$$

If $0 \neq c \in \mathbb{R}$, then

$$\delta(x) = \frac{1}{|c|} \delta\left(\frac{x}{c}\right). \tag{2}$$

Indeed, if we use the substitution $y = x/c$ and keep integrating from $-\infty$ to ∞, then we have

$$\int f(x) \frac{1}{|c|} \delta\left(\frac{x}{c}\right) dx = \int f(cy)\delta(y)dy = f(0).$$

Schwartz developed these ideas into a theory of *distributions*, i.e. continuous linear real-valued functionals on the space of test functions (with the suitable topology). Since then distributions play a major role in the theory of differential equations.

Some divergent integrals, e.g. $\int e^{itx} dt$, can be seen as distributions in that sense. In fact, as distributions,

$$\int e^{itx} dt = 2\pi\delta(x). \tag{3}$$

Indeed,

$$\begin{aligned}
\int dx\, f(x) \int e^{itx} dt &= \sqrt{2\pi} \int dt\, \frac{1}{\sqrt{2\pi}} \int f(x) e^{itx} dx \\
&= \sqrt{2\pi} \int \check{f}(t)\, dt \\
&= 2\pi \cdot \frac{1}{\sqrt{2\pi}} \int \check{f}(t) e^{-it0} dt = 2\pi f(0).
\end{aligned}$$

The exponential e^A of an operator A on $L^2(\mathbb{R})$ is the operator

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \ldots \tag{4}$$

provided that series converges.

If $(X\psi)(x) = x \cdot \psi(x)$, then

$$(e^X \psi)(x) = \sum_{k=0}^{\infty} \frac{1}{k!} (X^k \psi)(x) = \psi \cdot \sum_{k=0}^{\infty} \frac{1}{k!} x^k = \psi \cdot e^x.$$

If D is the derivative operator $\frac{d}{dx}$ and c a real number, then $e^{cD}\psi(x) = \psi(x + c)$. Indeed,

$$\begin{aligned}
e^{cD}\psi(x) &= \sum_{k=0}^{\infty} \frac{(cD)^k \psi(x)}{k!} = \sum_{k=0}^{\infty} \frac{D^k f(x)}{k!} c^k \\
&= \psi(x) + \frac{\psi'(x)}{1!} c + \frac{\psi''(x)}{2!} c^2 + \frac{\psi'''(x)}{3!} c^3 + \ldots
\end{aligned}$$

which is the Taylor series of $\psi(x + c)$ around point x; think of c as Δx.

Q: I worry about convergence of the Taylor series.

A: The Taylor series certainly converges on analytic functions, in particular on Gaussian functions

$$\exp\left(-\frac{(x-a)^2}{2b^2}\right).$$

The linear combinations of Gaussian functions are dense in $L^2(\mathbb{R})$, and there is a unique continuous extension of e^{cD} to L^2, namely the shift $f(x) \mapsto f(x+c)$.

3 Pushforward measures, and uniqueness theorem

We recall the definition of pushforward measures and then prove a measure-theoretic uniqueness theorem used in the proof of our main theorem in §9.

Consider measurable spaces $M_1 = (\Omega_1, \Sigma_1)$ and $M_2 = (\Omega_2, \Sigma_2)$. Let μ be a measure on M_1 and let a function $f : \Sigma_1 \to \Sigma_2$ be measurable for M_1, M_2.

Definition 2 (§3.6 in [5])**.** The *pushforward of μ along f*, a.k.a. the *f-pushforward of μ* or the *f-image* of μ, is the measure

$$f_*\mu(e) = \mu\big(f^{-1}(e)\big)$$

on M_2. $\lhd$

It is easy to check that $\nu = f_*\mu$ is indeed a measure on M_2.

Proposition 3. *With notation as above, for every measurable function $g : \Omega_2 \to \mathbb{C}$ on M_2, if $g(f(x))$ is μ-integrable then g is ν-integrable and*

$$\int_{M_2} g(y)d\nu(y) = \int_{M_1} g(f(x))d\mu(x).$$

For real-valued f, Proposition 3 is a modification of Theorem 3.6.1 in book [5] as described in the comments following the proof of theorem in the book. The generalization to $\mathbb{C}$-valued functions is straightforward.

Recall that, by default, real Euclidean spaces $\mathbb{R}^k$ are equipped with the Lebesgue measure. Accordingly, measurable subsets of $\mathbb{R}^k$ are Lebesgue measurable, and integrals are Lebesgue integrals.

Theorem 4 (Uniqueness)**.** *Let μ_1, μ_2 be finite Borel measures on $\mathbb{R}^2$. If $(ax + by)_*\mu_1 = (ax + by)_*\mu_2$ for all a, b not both zero, then $\mu_1 = \mu_2$.*

Proof. Let a, b range over pairs of reals not both zero. By Lemma 1, $\mu = \mu_1 - \mu_2$ is a measure on $\mathbb{R}^2$. It suffices to prove that μ is the zero measure $\mathbb{R}^2$. Let $\nu_{ab} = (ax + by)_*\mu$.

Every ν_{ab} is the zero measure on $\mathbb{R}$. Indeed, if s is a Borel subset s of $\mathbb{R}$, let $S = \{(x, y) : ax + by \in s\}$. We have

$$(ax + by)_*\mu(s) = \mu(S) = \mu_1(S) - \mu_2(S)$$
$$= (ax + by)_*\mu_1(s) - (ax + by)_*\mu_2(s) = 0.$$

We have.

$$\widehat{\nu}_{ab}(\zeta) = \int_{\mathbb{R}} e^{i\zeta t}\, d\nu_{ab}(t) = \int_{\mathbb{R}^2} e^{i\zeta(ax+by)}\, d\mu(x, y)$$
$$= \int_{\mathbb{R}^2} e^{i(a\zeta x + b\zeta y)}\, d\mu(x, y),$$

where the second equality uses Proposition 3 with $g(t) = e^{i\zeta t}$; the integrand $e^{i\zeta(ax+by)}$ is a bounded continuous function and therefore is integrable with respect to the (finite) measure μ.

Comparing this with the Fourier transform of μ,

$$\widehat{\mu}(\xi, \eta) = \int_{\mathbb{R}^2} e^{i(\xi x + \eta y)}\, d\mu(x, y),$$

we get

$$\widehat{\mu}(a\zeta, b\zeta) = \widehat{\nu}_{ab}(\zeta).$$

Since every ν_{ab} is the zero measure, every $\widehat{\nu}_{ab}(\zeta) = 0$ for all ζ. It follows that $\widehat{\mu}$ is the function zero.

By Proposition 3.8.6 in book [5], if two Borel measures on $\mathbb{R}^2$ have equal Fourier transforms, then they coincide. Applying this to μ and the zero measure, we conclude that μ is the zero measure. $\qquad\square$

The theorem generalizes to higher dimensions, but we restrict our attention to $\mathbb{R}^2$.

4   Marginal distributions

Traditionally, for a (signed) probability distribution of several variables, the marginal (signed) distributions are defined only for single variables or subsets

of the variables. We extend this definition to linear functions of the variables, restricting attention to just two variables.

Let $\mathcal{P}$ be a quasiprobability distribution on real plane $\mathbb{R}^2$ with coordinate axes x and y, and let (a, b) range over pairs of real numbers not both zero.

Definition 5. The $(ax + by)$ *marginal* of $\mathcal{P}$ is the pushforward measure (and in fact quasidistribution) $(ax + by)_* \mathcal{P}$ of $\mathcal{P}$ along the function $z = ax + bp : \mathbb{R}^2 \to \mathbb{R}$. ◁

In particular, $(x)_* \mathcal{P}(s) = \mathcal{P}(s \times \mathbb{R})$, and $(y)_* \mathcal{P}(s) = \mathcal{P}(\mathbb{R} \times s)$, so that $(x)_* \mathcal{P}$ and $(y)_* \mathcal{P}$ are traditional marginals.

Now suppose that $\mathcal{P}$ is given by a density function $f(x, y)$, so that $\mathcal{P}(s) = \iint_s f(x, y) \, dx \, dy$ for all measurable subsets s of $\mathbb{R}^2$. We show that in this case every marginal $(ax + by)_* \mathcal{P}$ is given by a density function which will be denoted $(ax + by)_* f$.

Lemma 6. *For every pair (a, b), the function*

$$
g(z) = \begin{cases} \dfrac{1}{|b|} \displaystyle\int f\left(x, \dfrac{1}{b}(z - ax)\right) dx & \text{if } b \neq 0 \\ \dfrac{1}{|a|} \displaystyle\int f\left(\dfrac{z}{a}, y\right) dy & \text{otherwise} \end{cases}
$$

is the density function $(ax + by)_* f$.

Proof. We consider the case $b \neq 0$; the other case is similar (and a bit simpler). It suffices to prove that $(ax + by)_* \mathcal{P}[u, v] = \int_u^v g(z) \, dz$ on intervals $[u, v]$ with $u \leqslant v$. We have

$$
(ax + by)_* \mathcal{P}[u, v] = \iint_{u \leqslant ax + by \leqslant v} f(x, y) \, dx \, dy.
$$

Let $z = ax + by$, so that $y = \frac{1}{b}(z - ax)$. Change variables in the integral, from x, y to x, z. The absolute value of the Jacobian determinant of this transformation is $\frac{1}{|b|}$, so we obtain

$$
(ax + by)_* \mathcal{P}[u, v] = \iint_{u \leqslant z \leqslant v} \frac{1}{|b|} f\left(x, \frac{1}{b}(z - ax)\right) dx \, dz
$$

$$
= \int_u^v dz \int_{\mathbb{R}} \frac{1}{|b|} f\left(x, \frac{1}{b}(z - ax)\right) dx = \int_u^v g(z) \, dz. \qquad \square
$$

Lemma 7. $(ax + by)_* \mathcal{P}(u, v) = (acx + bcy)_* \mathcal{P}(cu, cv)$ *for every real* $c \neq 0$ *and every open interval* (u, v) *of* $\mathbb{R}$.

Proof.

$$
\begin{aligned}
(ax + by)_* \mathcal{P}(u, v) &= \mathcal{P}\{(x, y) : u < ax + by < v\} \\
&= \mathcal{P}\{(x, y) : cu < acx + bcy < cv\} \\
&= (acx + cby)_* \mathcal{P}(cu, cv).
\end{aligned}
$$

□

Lemma 8 (Lemma 6.4 in [3])**.** *For all real* a, b *not both zero and every function* $g : \mathbb{R} \to \mathbb{R}$, *the following claims are equivalent.*

1. *g is the density function* $(ax + bp)_* f$.

2. *$\hat{g}(\zeta) = \sqrt{2\pi} \cdot \hat{f}(a\zeta, b\zeta)$ where $\hat{f}$ and $\hat{g}$ are (forward) Fourier transforms of f and g respectively.*

The computation that proves the lemma was essentially done in our proof of the uniqueness theorem above.

5 Position, momentum, and their linear combinations

Consider one particle moving in one dimension. A generalization to more particles in more dimensions is relatively straightforward.

In classical mechanics, the position x and momentum p of the particle determine its current state. The set of all possible classical states is the phase space of the particle. In quantum mechanics, the state space of the particle is the Hilbert space $L^2(\mathbb{R})$.

Using Dirac's bra-ket notation, we write $|\psi\rangle$ for the vector given by function ψ. Unit vectors $|\psi\rangle$ represent states of the particle, and two unit vectors represent the same state if and only if they differ by a scalar factor $e^{i\theta}$.

Q: How come a whole $\mathbb{R} \to \mathbb{C}$ function is needed to represent just one quantum state?

A: Because, in quantum mechanics, a particle is also a wave. If it is in state $|\psi\rangle$, then $|\psi(x)|^2$ is the probability density at x for finding the particle. $\langle \psi | \psi \rangle$ is the total probability. Accordingly, $\langle \psi | \psi \rangle$ must be 1, and this is why unit vectors are used to represent states.

In quantum mechanics, observable quantities are represented by Hermitian operators on the state space. In particular, the position observable X and momentum observable P are (represented by) operators

$$(X\psi)(x) = x \cdot \psi(x),$$
$$(P\psi)(x) = -i\hbar\frac{d\psi}{dx}$$

where $\hbar = h/2\pi$ is the reduced Planck constant; h is the (unreduced) Planck constant. We will simplify notation by assuming (by proper choice of units) that $\hbar = 1$.

Q: Functions $X\psi$ and $P\psi$ may fail to be square integrable.

A: Indeed, the operators X, P are undefined in some states.

Q: The formula for P looks mysterious to me. Is momentum also related to the wave character of our particle?

A: Yes, it is. The momentum p corresponds to the wavelength $\lambda = h/p$ (de Broglie relation). So, if a particle had an exact value p of the momentum, its wave function would be (up to a scalar factor)

$$\psi(x) = e^{2\pi i x/\lambda} = e^{ipx/\hbar} = e^{ipx},$$

which is an eigenfunction of P with eigenvalue p. (Yes, this ψ isn't in our Hilbert space. We'll return to this point in §6.) P is designed to be the operator whose eigenvalues are momenta, just as X is the operator whose eigenvalues (corresponding to "eigenfunctions" $\delta(x-q)$) are positions q.

There is a simple mathematical connection between the two operators: $\mathcal{F} P \mathcal{F}^{-1} = X$ where $\mathcal{F}$ is the Fourier transform. It suffices to verify this equality

on test functions.

$$\mathcal{F}P(\psi)(\xi) = \int P\psi(x)e^{-i\xi x}dx = \int -i\frac{d\psi(x)}{dx}e^{-i\xi x}dx$$
$$= i\int \psi(x)\frac{de^{-i\xi x}}{dx}dx = \xi \int \psi(x)e^{-i\xi x}dx =$$
$$= \xi \cdot \mathcal{F}(\psi)(\xi) = X\mathcal{F}(\psi)(\xi).$$

The third equality uses integration by parts; the extra terms disappear because $e^{-i\xi x}$ is bounded and the test function ψ approaches zero when the argument goes to $\pm\infty$.

Q: The equality $\mathcal{F}P\mathcal{F}^{-1} = X$ makes me worry about the dimensions. You mentioned earlier that, when x is a length, as here, then the variable ξ of the Fourier transform is a reciprocal length. But here the variable of the Fourier transform seems to be a momentum. How do you reconcile these dimensions?

A: By convention, we're using units where $\hbar = 1$, and the dimension of $\hbar$ is momentum times length, so our convention makes reciprocal length the same as momentum. Thus, our P is dimensionally correct.

6 Dirac's kets

The spectral theory of self-adjoint operators in finite dimensional Hilbert spaces is relatively simple. Suppose that $\mathcal{H}$ is n-dimensional, and consider a self-adjoint operator A on $\mathcal{H}$. Since $\mathcal{H}$ is self-adjoint, all its eigenvalues are real. There exists an orthonormal basis $|1\rangle, \ldots, |n\rangle$ for $\mathcal{H}$ composed of eigenvectors of A. Let $\lambda_1, \ldots, \lambda_n$ be the corresponding eigenvalues.

For simplicity of exposition, we assume that all eigenvalues λ_k are distinct (and thus non-degenerate since the number of them equals the dimension of $\mathcal{H}$). This suffices for our purposes in this paper, and it allows us to label the eigenvectors with the corresponding eigenvalues. We may write $|\lambda_k\rangle$ instead of $|k\rangle$.

It will be convenient, in the infinite dimensional case, to use an alternative characterization of the "distinct eigenvalues" assumption. The operator A is called *cyclic* if there is a vector $|\psi\rangle$ in $\mathcal{H}$ such that the vectors

$$|\psi\rangle, A|\psi\rangle, A^2|\psi\rangle, \ldots, A^{n-1}|\psi\rangle$$

span the whole space $\mathcal{H}$; such vector $|\psi\rangle$ is *A-cyclic*.

Claim 9. *A is cyclic if and only if the eigenvalues $\lambda_1, \ldots, \lambda_n$ are all distinct.*

Proof. If the values are distinct, then $\sum_{k=1}^{n} |k\rangle$ is *A*-cyclic. If, on the other hand, $\lambda_i = \lambda_j$, then for any vector $|\psi\rangle = \sum c_k |k\rangle$, all linear combinations of vectors $A^l |\psi\rangle$ have coefficients of $|i\rangle$ and $|j\rangle$ in the same ratio $c_i : c_j$. Thus these linear combinations fail to span $\mathcal{H}$. $\qquad\square$

Think of $\mathcal{H}$ as the state space of a quantum system where unit vectors represent states of the system (and two unit vectors represent the same state if and only if they are collinear). Then every state $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_j |\lambda_j\rangle \langle \lambda_j \,|\, \psi\rangle \tag{5}$$

where scalars $\langle \lambda_j \,|\, \psi\rangle$ are probability amplitudes of the wave function $|\psi\rangle$, so that the corresponding probabilities are $|\langle \lambda_j \,|\, \psi\rangle|^2$. Accordingly,

$$A|\psi\rangle = \sum_j A|\lambda_j\rangle \langle \lambda_j \,|\, \psi\rangle = \sum_j \lambda_j |\lambda_j\rangle \langle \lambda_j \,|\, \psi\rangle$$

and the expectation of *A* in a state $|\psi\rangle$ is

$$\langle \psi | A | \psi\rangle = \sum_j \lambda_j \langle \psi \,|\, \lambda_j\rangle \langle \lambda_j \,|\, \psi\rangle = \sum_j \lambda_j \,|\langle \lambda_j \,|\, \psi\rangle|^2 \,.$$

The infinite dimensional case is much more involved. Let *A* be a self-adjoint operator on a infinite-dimensional space $\mathcal{H}$. If you measure *A*, you still receive some real number but it is not necessarily an eigenvalue. It is just an element of the spectrum

$$\sigma(A) = \{\lambda : \text{operator } A - \lambda I \text{ is not invertible}\}$$

of *A*. Notice that the set of eigenvalues is

$$\{\lambda : \text{operator } A - \lambda I \text{ is not one-to-one}\} \,.$$

In finite dimensions, any one-to-one linear operator is invertible, but this principle fails in infinite dimensions.

As in the finite dimensional case, we make the simplifying assumption that the spectrum is simple in the sense that A is *cyclic*. That is, there is a vector $|\psi\rangle$, called an *A-cyclic* vector, such that the vectors

$$|\psi\rangle, \; A|\psi\rangle, \; A^2|\psi\rangle, \; \ldots$$

are dense in whole Hilbert space $\mathcal{H}$.

How to generalize the spectral theory of self-adjoint operators in the finite dimensional case to the infinite dimensional case? Paul Dirac came up with an elegant heuristic generalization [6] which works for operators like $aX + bP$ on $L^2(\mathbb{R})$. We sketch Dirac's generalization, restricting attention to the Hilbert space $L^2(\mathbb{R})$ and to cyclic self-adjoint operators A on $L^2(\mathbb{R})$ with $\sigma(A) = \mathbb{R}$.

For each spectrum value $r \in \mathbb{R}$, there is a generalized eigenvector, in short *eigenket*, $|r\rangle$ for A, so that $A|r\rangle = r|r\rangle$. As in the finite dimensional case, we take advantage of the cyclicity assumption to label eigenkets $|r\rangle$ by the corresponding spectrum value r.

Q: What do you mean by eigenvector being generalized?

A: That it does not necessarily belong to $L^2(\mathbb{R})$.

Q: That is confusing. Give me an example.

A: If A is the momentum operator P, then $|r\rangle$ is the function $x \mapsto e^{irx}/\sqrt{2\pi}$. Indeed,

$$P\left(\frac{e^{irx}}{\sqrt{2\pi}}\right) = -i\frac{d}{dx}e^{irx} = r\left(\frac{e^{irx}}{\sqrt{2\pi}}\right).$$

The finite-dimensional orthonormality requirement is replaced by δ-*normality*:

$$\langle s \,|\, r \rangle = \delta(r - s).$$

In the case of P, using (3) we have

$$\int \frac{e^{-isx}}{\sqrt{2\pi}} \cdot \frac{e^{irx}}{\sqrt{2\pi}} \, dx = \frac{1}{2\pi}\int e^{i(r-s)x}dx = \delta(r - s).$$

The finite-dimensional expansion (5) with respect to the eigenvectors becomes

the *Dirac basis expansion*:

$$|\psi\rangle = \int |r\rangle\langle r\,|\,\psi\rangle\,dr$$

where $\langle r\,|\,\psi\rangle$ is the density of probability amplitude, so that the corresponding probability density is

$$D_A(r) = |\langle r\,|\,\psi\rangle|^2. \tag{6}$$

The probability distribution on $\mathbb{R}$, given by the probability density function $D_A(r)$, will be denoted $\mathcal{P}_A$.

We have $A|\psi\rangle = \int A|r\rangle\langle r\,|\,\psi\rangle\,dr = \int r|r\rangle\langle r\,|\,\psi\rangle\,dr$, and the expectation of A in state $|\psi\rangle$ is

$$\langle\psi|A|\psi\rangle = \int r\,|\langle r\,|\,\psi\rangle|^2\,dr. \tag{7}$$

In the case $A = P$, the density of probability amplitude is

$$\langle r\,|\,\psi\rangle = \frac{1}{\sqrt{2\pi}}\langle e^{irx}\,|\,\psi(x)\rangle = \frac{1}{\sqrt{2\pi}}\int e^{-irx}\psi(x)\,dx = \widehat{\psi}(r),$$

the probability density is

$$D_P(r) = \left|\widehat{\psi}(r)\right|^2, \tag{8}$$

and $\mathcal{P}_P$ is the corresponding probability distribution.

Q: Is there mathematical justification of Dirac's heuristic generalization?

A: The theory of rigged Hilbert spaces, see [4] for example, mathematically justifies the use of generalized eigenvectors[2]. The operators A are subject to some constraints which are satisfied by the operators X, P and their linear combinations. See §3 in [10] in this connection.

Q: It this "rigged" as in rigged elections? What witty guy came up with this term?

A: This is a translation of a Russian term[3] meaning *equipped* or *rigged* as in "rigging a ship for sailing."

Q: Will you tell me more about rigged Hilbert spaces?

A: Rigged Hilbert spaces deserve a separate column article, but the basic idea

---

[2]An alternative justification is provided by the spectral theory of operators [9].
[3]оснащенный; see [8].

is that a Hilbert space $\mathcal{H}$ is augmented with two additional spaces to obtain a so-called *Gelfand triple*.

In our case, the triple is

$$\Phi \subset L^2(\mathbb{R}) \subset \Phi^\times$$

where Φ comprises test functions and $\Phi^\times$ comprises distributions. Recall that we touched upon distributions in §2.3.

Q: What kind of distributions are the eigenkets $|r\rangle = e^{irx}$ of P?

A: An antilinear functional

$$\langle f \, | \, r \rangle = \frac{1}{\sqrt{2\pi}} \int f^*(x) e^{irx} dx$$

on test functions f [11].

Q: I am confused. Earlier, in §2.3, you said that distributions are linear, not antilinear, functionals.

A: Yes, the elements of $\Phi^\times$ are antilinear functionals serving as generalized kets while distributions are linear functionals serving as generalized bras. In our case, it is safe to ignore the distinction and call both of them distributions.

Q: Explain.

A: To see what goes on, consider the finite dimensional case. View elements of $\mathbb{C}^n$ as column vectors, so the inner product $\langle \varphi \, | \, \psi \rangle$ is given by the matrix product $\varphi^\dagger \cdot \psi$. Thus ψ acts antilinearly on φ whereas φ acts linearly on ψ, but both are column vectors in the same space $\mathbb{C}^n$. The same entities serve as linear and antilinear functionals.

From this point of view, an eigenket of an operator $A : \mathbb{C}^n \to \mathbb{C}^n$ for the eigenvalue λ is a column vector ψ such that $A \cdot \psi = \lambda\psi$. An eigenbra is a column vector φ such that $\varphi^\dagger \cdot A = \lambda\varphi^\dagger$; equivalently, $A^\dagger \cdot \varphi = \lambda^*\varphi$. When A is self-adjoint and its eigenvalues λ are therefore real, the eigenbras and eigenkets coincide.

The situation is similar in infinite dimensions, and the operators whose eigenkets and eigenbras we use are (essentially) self-adjoint. Thus we can safely use the same entities as eigenkets and eigenbras.

7 Linear combinations of position and momentum operators

In this section, we apply Dirac's machinery to linear combinations $aX + bP$ of the operators X and P where a, b are real numbers not both zero.

One can argue that, from physical considerations, the spectrum of every $aX + bP$ is $\mathbb{R}$. This is supported by theory. Every operator $aX + bP$ is self-adjoint[4]. The spectrum of every self-adjoint operator consists of reals [9, Theorem 9.17]. Hence, for every $aX + bP$, the spectrum $\sigma(aX + bP) \subseteq \mathbb{R}$. Furthermore, for every $aX + bP$, we will provide an eigenket $|r\rangle$ of $aX + bP$ for every real r, so that $\sigma(aX + bP) = \mathbb{R}$.

Lemma 10. *Let* $Z = aX + bP$ *where* a, b *are real numbers not both zero, let* $0 \neq c \in \mathbb{R}$, *and let* (u, v) *be an interval in* $\mathbb{R}$. *Then, in every state* $|\psi\rangle$,

$$\mathsf{P}[Z \in (u, v)] = \mathsf{P}[cZ \in \{cr : u < r < v\}] = \begin{cases} \mathsf{P}[cZ \in (cu, cv)] & \text{if } c > 0, \\ \mathsf{P}[cZ \in (cv, cu)] & \text{if } c < 0. \end{cases}$$

Q: This seems obvious. The two events are the same and so have the same probability.

A: Z is an observable, and the probabilities are determined by the rules of quantum mechanics.

Q: But we can view Z also as a random variable by repeatedly measuring it in a given state. It is gratifying that both views give the same result, isn't it?

A: Agreed.

Proof. We consider the case $c > 0$; the case $c < 0$ is similar.

As we saw in the previous section, there is a δ-normal system $\langle f(r, x) : r \in \mathbb{R} \rangle$ where $f(r, x)$ is an eigenket of Z for spectrum value r. Then $\left\langle \frac{1}{\sqrt{c}} f(\frac{r}{c}, x) : r \in \mathbb{R} \right\rangle$ is a δ-normal system for cZ. Indeed,

$$(cZ) \left(\frac{1}{\sqrt{c}} f\left(\frac{r}{c}, x\right) \right) = \sqrt{c}\, Z f\left(\frac{r}{c}, x\right) = \sqrt{c}\, \frac{r}{c} f\left(\frac{r}{c}, x\right) = r \cdot \frac{1}{\sqrt{c}} f\left(\frac{r}{c}, x\right),$$

---

[4]More exactly, $aX + bP$ is essentially self-adjoint [9, Proposition 9.40], and every essentially self-adjoint operator has a unique self-adjoint extension [9, Proposition 9.11].

and, using the δ-normality of $\langle f(r,x) : r \in \mathbb{R}\rangle$ and using (2), we have

$$\left\langle \frac{1}{\sqrt{c}} f\left(\frac{s}{c}, x\right) \,\middle|\, \frac{1}{\sqrt{c}} f\left(\frac{r}{c}, x\right) \right\rangle = \frac{1}{c}\delta\left(\frac{s}{c} - \frac{r}{c}\right) = \delta(r - s).$$

By (6), in a state $|\psi\rangle$, the probability density functions for Z and cZ are

$$|\langle f(r,x) \,|\, \psi(x)\rangle|^2 \quad \text{and} \quad |\langle f(r/c, x) \,|\, \psi(x)\rangle|^2$$

Using the substitution $r = cs$, we have

$$\mathsf{P}[cZ \in (cu, cv)] = \int_{cu}^{cv} dr \left| \int_{\mathbb{R}} dx \, \frac{1}{\sqrt{c}} f^*\left(\frac{r}{c}, x\right) \psi(x) \right|^2$$

$$= \int_u^v ds \left| \int_{\mathbb{R}} dx \, f^*(s,x)\psi(x) \right|^2 = \mathsf{P}[Z \in (u,v)]. \qquad \square$$

For every real r, the eigenket $|r\rangle$ of X for spectrum value r is the delta function $\delta(x - r)$, which acts on test functions according to

$$f(x) \mapsto \langle f \,|\, r\rangle = \int f^*(x)\delta(x - r)\,dx = f^*(r).$$

$X|r\rangle$ is the distribution that sends a test function $f(x)$ to

$$\int f^*(x)x\delta(x - r)\,dx = \int f^*(x + r)(x + r)\delta(x)\,dx$$

$$= rf^*(r) = r\langle f \,|\, r\rangle.$$

Thus, $X|r\rangle = r|r\rangle$. In particular, $\sigma(X) = \mathbb{R}$. Furthermore, these eigenkets form a δ-normal system:

$$\langle s \,|\, r\rangle = \int \delta(x - s)\delta(x - r)\,dx = \delta(r - s).$$

The density of probability amplitude is $\langle r \,|\, \psi\rangle = \int \delta(x - r)\psi\,dx = \psi(r)$, the probability density function is

$$D_X(r) = |\psi(r)|^2. \tag{9}$$

and $\mathcal{P}_X$ is the corresponding probability distribution.

Next we consider a case $Z = -cX + P$. It is still a special case, but the general case $Z = aX + bP$ easily reduces to that special case; we return to this issue in the next section.

The mapping defined by $U|\psi\rangle = e^{icx^2/2}|\psi\rangle$ is a unitary operator on the test functions:

$$\langle U\psi \mid U\varphi \rangle = \langle \psi | U^\dagger U | \varphi \rangle = \langle \psi \mid \varphi \rangle.$$

We have $UPU^{-1} = Z$ on test functions [3, §6]. Indeed,

$$\left(UPU^{-1}\right)\psi = e^{icx^2/2} \cdot \left(-i\frac{d}{dx}\left(e^{-icx^2/2}\psi\right)\right) = -cx\psi - i\frac{d\psi}{dx}$$
$$= -cX\psi + P\psi = Z\psi.$$

Accordingly, one may expect that U transforms generalized eigenvectors of P into those of Z. This intuition happens to be correct.

For each real r, the distribution $e^{irx+icx^2/2}$ is a generalized eigenvector of Z for r:

$$Ze^{irx+icx^2/2} = (-cX + P)e^{irx+icx^2/2} = -cxe^{irx+icx^2/2} - i\frac{d}{dx}e^{irx+icx^2/2}$$
$$= e^{irx+icx^2/2}\left[-cx - i(ir + icx)\right] = re^{irx+icx^2/2}.$$

It follows that the spectrum of Z is $\mathbb{R}$.

For each real r, let $|r\rangle$ be the eigenket $\frac{1}{\sqrt{2\pi}}e^{irx+icx^2/2}$ of Z for generalized eigenvalue r. These eigenkets form a δ-normal system:

$$\int \frac{e^{-isx-icx^2/2}}{\sqrt{2\pi}} \cdot \frac{e^{irx+icx^2/2}}{\sqrt{2\pi}}\, dx = \frac{1}{2\pi}\int e^{i(r-s)x}dx = \delta(r-s).$$

Accordingly, the density of probability amplitude is

$$\langle r \mid \psi \rangle = \frac{1}{\sqrt{2\pi}}\langle e^{irx+icx^2/2} \mid \psi(x)\rangle = \frac{1}{\sqrt{2\pi}}\int e^{-irx+icx^2/2}\psi(x)\, dx,$$

the probability density is

$$D_Z(r) = \frac{1}{2\pi}\left|\langle e^{irx+icx^2/2} \mid \psi(x)\rangle\right|^2, \tag{10}$$

and the corresponding probability distribution is $\mathcal{P}_Z$.

8 Wigner's quasiprobability distribution

In every $L^2(\mathbb{R})$ state $|\psi\rangle$, Wigner's quasiprobability $\mathcal{P}_W$ is given by probability density function

$$w(x, p) = \frac{1}{2\pi} \int_{\mathbb{R}} \psi^*(x + \frac{\gamma\hbar}{2})\psi(x - \frac{\gamma\hbar}{2})e^{i\gamma p}\, d\gamma. \tag{11}$$

The integral converges in every $L^2(\mathbb{R})$ state $|\psi\rangle$. According to Wigner, the x-marginal and p-marginal of his distribution are the probability distributions for X and P respectively [13].

Q: You don't take advantage of using units where $\hbar = 1$. I guess it doesn't hurt to keep $\hbar$ in this case.

A: Exactly.

Q: Is this obvious that the x-marginal and p-marginal of Wigner's distribution are the probability distributions for X and P?

A: It is certainly easier to verify than the claim that an arbitrary $(ax + by)$-marginal of Wigner's distribution is the probability distribution for $aX + bY$.

Lemma 11. *In every state $|\psi\rangle$, $(x)_* \mathcal{P}_W = \mathcal{P}_X$.*

Proof. It suffices to prove that, in every state $|\psi\rangle$, the x-marginal $x_* w$ of Wigner's density is the density function $|\psi|^2$. Since two density function coincide if they are proportional, we may neglect constant factors.

By Lemma 6, up to constant factors, the $x_* w$ density function is

$$x_* w = \int w(x, p) dp = \qquad\qquad \text{by Lemma 6}$$
$$\int \left[\int \psi^*(x + \frac{\gamma\hbar}{2})\psi(x - \frac{\gamma\hbar}{2})\, e^{i\gamma p}\, d\gamma \right] dp =$$
$$\int \left[\int e^{i\gamma p} dp \right] \psi^*(x + \frac{\gamma\hbar}{2})\psi(x - \frac{\gamma\hbar}{2})\, d\gamma = \qquad \text{by (3)}$$
$$\int \delta(\gamma)\psi^*(x + \frac{\gamma\hbar}{2})\psi(x - \frac{\gamma\hbar}{2})\, d\gamma = \qquad \text{by (1)}$$
$$\psi^*(x)\psi(x) = |\psi|^2 \qquad\qquad\qquad\qquad \square$$

Lemma 12. $w(x, p) = \int \widehat{\psi}(p + \frac{\gamma}{2})^* \, \widehat{\psi}(p - \frac{\gamma}{2}) \, e^{-i\gamma x} \, d\gamma$
up to a constant factor.

Proof. Ignoring constant factors and using the formula

$$\psi(x) = \int \widehat{\psi}(\xi) \, e^{i\xi x} \, d\xi$$

for the inverse Fourier transform, we get

$$
\begin{aligned}
w(x, p) &= \int \psi(x + \frac{\gamma}{2})^* \, \psi(x - \frac{\gamma}{2}) \, e^{i\gamma p} \, d\gamma \\
&= \iiint \widehat{\psi}(\xi)^* \, e^{-i\xi(x + \frac{\gamma}{2})} \, \widehat{\psi}(\eta) \, e^{i\eta(x - \frac{\gamma}{2})} \, e^{i\gamma p} \, d\xi \, d\eta \, d\gamma \\
&= \iint \left[e^{-i\gamma(\frac{\xi}{2} + \frac{\eta}{2} - p)} d\gamma \right] \widehat{\psi}(\xi)^* \, \widehat{\psi}(\eta) \, e^{-ix(\xi - \eta)} \, d\xi \, d\eta \\
&= \iint \widehat{\psi}(\xi)^* \, \widehat{\psi}(\eta) \, e^{-ix(\xi - \eta)} \, \delta(\frac{\xi}{2} + \frac{\eta}{2} - p) \, d\xi \, d\eta.
\end{aligned}
$$

The δ function makes it easy to perform the integration with respect to η. Just substitute $2p - \xi$ for η in the remaining factors.

$$w(x, p) = \int \widehat{\psi}(\xi)^* \, \widehat{\psi}(2p - \xi) \, e^{-ix(2\xi - 2p)} \, d\xi.$$

Change variables in the integral to $\gamma = 2\xi - 2p$, so ξ becomes $p + \frac{\gamma}{2}$.

$$w(x, p) = \int \widehat{\psi}(p + \frac{\gamma}{2})^* \, \widehat{\psi}(p - \frac{\gamma}{2}) \, e^{-i\gamma x} \, d\gamma. \qquad \square$$

Corollary 13. *The (p)-marginal of Wigner's distribution is the probability distribution $\mathcal{P}_P$ for P.*

The proof is similar to that of Lemma 11, except the formula of Lemma 12 is used.

The generalization to arbitrary $(ax + bp)$-marginals will be proved in §9. A key role in that proof is played by the following lemma.

Lemma 14 (Lemma 6.6 in [3]). *For all test functions ψ and real numbers $\alpha, \beta,$*

not both zero,

$$\langle\psi|e^{-i(\alpha X+\beta P)}|\psi\rangle = e^{i\alpha\beta\hbar/2}\int\psi^*(y)e^{-i\alpha y}\psi(y-\beta\hbar)\,dy.$$

Actually, Lemma 6.6 in [3] speaks about states $|\psi\rangle$ that are "nice" in the sense that the function ψ is smooth and compactly supported. But the lemma and its proof obviously remain valid for test functions ψ.

9   Characterization theorem

Theorem 15. *In every state $|\psi\rangle$ in $L^2(\mathbb{R})$, Wigner's quasidistribution $\mathcal{P}_W$ is the unique quasidistribution on $\mathbb{R}^2$ such that, for all real numbers a, b not both zero, the marginal $(ax + bp)_*\mathcal{P}_W$ is correct in the sense that it coincides with the probability distribution $\mathcal{P}_Z$ of the observable $Z = aX + bP$.*

Proof. The uniqueness follows from Theorem 4.

We need to prove the equality $(ax + bp)_*\mathcal{P}_W = \mathcal{P}_Z$ for every pair of reals a, b not both zero and in every state $|\psi\rangle$. By virtue of Lemmas 7 and 10, we restrict attention to two cases:

1. $a = 1$ and $b = 0$, so that $Z = X$,

2. $b = 1$, so that $Z = aX + P$.

Case (1) is taken care of by Lemma 11. In the rest of the proof we consider case (2). Even though $b = 1$, we sometimes write $aX + bP$ and $ax + bp$ anyway.

Fix an arbitrary real number a. We prove that, in every state,

$$(ax + bp)_*\mathcal{P}_W = (ax + p)_*\mathcal{P}_W = \mathcal{P}_Z. \tag{12}$$

In every state $|\psi\rangle$, the quasidistribution $\mathcal{P}_W$ is given by the probability density function $w(x, p)$ specified by formula (11). By Lemma 6, the marginal quasidis-

tribution $(ax + p)_*\mathcal{P}_W$ is given by quasiprobability density function

$$g(r) = \int w(x, r - ax)\, dx$$
$$= \frac{1}{2\pi} \iint \psi^*(x + \frac{\gamma\hbar}{2})\, \psi(x - \frac{\gamma\hbar}{2})\, e^{i\gamma(r-ax)}\, d\gamma\, dx.$$

The probability distribution $\mathcal{P}_Z$ is given by the probability density function D_Z of Z specified in (10). To prove (12), it suffices to prove

$$g(r) = D_Z(r). \tag{13}$$

Both sides of (13) are continuous as functions of the state in the $L^2(\mathbb{R})$ metric. By continuity, it suffices to prove that the equality (13) holds in every "nice" state $|\psi\rangle$ provided that the nice states are dense in $L^2(\mathbb{R})$.

Theorem 6.2 in [3] does just that. In that theorem, a state $|\psi\rangle$ is nice if ψ is smooth and compactly supported. While Theorem 6.2 addresses both, the uniqueness and the correctness aspects, the emphasis in its proof is on uniqueness, and the correctness proof may be a bit confusing. We explain it here. Notice that two density functions coincide if they are proportional. Accordingly, we ignore constant factors in equations below.

Since (ignoring the factor $1/2\pi$)

$$w(x, p) = \int \psi^*(x + \frac{\gamma\hbar}{2})\, \psi(x - \frac{\gamma\hbar}{2})\, e^{i\gamma p}\, d\gamma,$$

its Fourier transform is

$$\hat{w}(\alpha, \beta) = \iiint dx\, dp\, d\gamma\, \psi^*(x + \frac{\gamma\hbar}{2})\, \psi(x - \frac{\gamma\hbar}{2})\, e^{i\gamma p}\, e^{-i\alpha x}\, e^{-i\beta p}.$$

Here p occurs only in two of the exponential factors, so the integration over p produces

$$\int dp\, e^{i(\gamma - \beta)p} = \delta(\gamma - \beta).$$

The delta function now makes the integral over γ trivial; just substitute β for γ in the integrand. Thus,

$$\hat{w}(\alpha, \beta) = \int dx\, \psi^*(x + \frac{\beta\hbar}{2})\, \psi(x - \frac{\beta\hbar}{2})\, e^{-i\alpha x}.$$

Introducing a new integration variable $y = x + \frac{\beta\hbar}{2}$, we get

$$\widehat{w}(\alpha,\beta) = \int dy\, \psi^*(y)\, \psi(y - \beta\hbar)\, e^{-i\alpha y}\, e^{i\alpha\beta\hbar/2}.$$

By Lemma 14,

$$\widehat{w}(\alpha,\beta) = \langle\psi|e^{-i(\alpha X + \beta P)}|\psi\rangle.$$

In particular, if Z is defined as $aX + bP$ and if we substitute $a\zeta$ and $b\zeta$ for α and β, we get

$$\widehat{w}(a\zeta, b\zeta) = \langle\psi|e^{-i\zeta Z}|\psi\rangle.$$

By Lemma 8, $\langle\psi|e^{-i\zeta Z}|\psi\rangle$ is the Fourier transform of the marginal $g(z) = (ax + bp)_* w(x, p)$. But this same $\langle\psi|e^{-i\zeta Z}|\psi\rangle$ is also, up to a constant factor, the Fourier transform of the density function D_Z of Z in the state ψ. Indeed, by (4),

$$e^{-i\zeta Z} = \sum_k \frac{1}{k!}(-i\zeta Z)^k,$$

and we restrict attention to test functions ψ, so that convergence is no problem. We have

$$\langle\psi|e^{-i\zeta Z}|\psi\rangle = \left\langle\psi \,\Big|\, \sum_k \frac{1}{k!}(-i\zeta Z)^k \,\Big|\, \psi\right\rangle = \sum_k \frac{(-i\zeta)^k}{k!}\langle\psi|Z^k|\psi\rangle$$

Now we use Dirac's machinery. Let $|r\rangle$ be the eigenket $\frac{1}{\sqrt{2\pi}}e^{irx + icx^2/2}$ of Z for spectrum value r. We have $Z^2|r\rangle = Z(r|r\rangle) = r^2|r\rangle$ and similarly for other powers of Z. By the preceding computation, (7), and (10), we have

$$\langle\psi|e^{-i\zeta Z}|\psi\rangle = \sum_k \frac{(-i\zeta)^k}{k!}\int r^k|\langle r\,|\,\psi\rangle|^2\, dr = \int \sum_k \frac{(-i\zeta)^k}{k!} r^k|\langle r\,|\,\psi\rangle|^2\, dr$$

$$= \int e^{-i\zeta r}|\langle r\,|\,\psi\rangle|^2\, dr = \int e^{-i\zeta r} D_Z(r)\, dr,$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Acknowledgment

References

[1] Jacqueline Bertrand and Pierre Bertrand, "A tomographic approach to Wigner's function," Foundations of Physics 17:4 (1987) 397–405

[2] Andreas Blass and Yuri Gurevich, "Negative probabilities," Bulletin of EATCS, February 2014

[3] Andreas Blass and Yuri Gurevich, "Negative probabilities: What they are for," Journal of Physics A 54 (2021), article 315303, `https://doi.org/10.1088/1751-8121/abef4d`

[4] Arno Böhm, "The rigged Hilbert space and quantum mechanics," Springer 1978

[5] Vladimir I. Bogachev, "Measure theory," Vol. 1, Springer 2007

[6] Paul Dirac, "The principles of quantum mechanics," 4th edition, Oxford University Press 1958

[7] Richard P. Feynman, "Simulating physics with computers," International Journal of Theoretical Physics, Vol. 21, Nos. 6/7 (1982) 467–488

[8] Israel M. Gelfand and Naum Y. Vilenkin, "Generalized functions," Vol. 4, Academic Press 1964; the Russian original published in 1961 by Физматлит

[9] Brian C. Hall, "Quantum theory for mathematicians," Springer 2013

[10] Rafael de la Madrid, "Quantum mechanics in rigged Hilbert space language," PhD thesis, Universidad de Valladolid, 2001

[11] Rafael de la Madrid, "The role of the rigged Hilbert space in quantum mechanics," European Journal of Physics 26 (2005) 287–312, doi:10.1088/0143-0807/26/2/008

[12] Walter Rudin, "Principles of mathematical analysis," McGraw-Hill, 3rd edition, 1976

[13] Eugene P. Wigner, "On the quantum correction for thermodynamic equilibrium," Physical Review 40 (1932) 749–759