What is special about the Kirkwood-Dirac distributions?

M. Spriet[1], C. Langrenez[1], R. Brummelhuis[2], and S. De Bièvre[1]

[1]Univ. Lille, CNRS, Inria, UMR 8524, Laboratoire Paul Painlevé, F-59000 Lille, France
[2]Univ. Reims Champagne-Ardenne, CNRS, UMR 9008, laboratoire Mathématique de Reims, F-51687 Reims, France

November 5, 2025

Abstract

Among all possible quasiprobability representations of quantum mechanics, the family of Kirkwood-Dirac representations has come to the foreground in recent years because of the flexibility they offer in numerous applications. This raises the question of their characterisation: what makes Kirkwood-Dirac representations special among all possible choices? We show the following. For two observables $\hat{A}$ and $\hat{B}$, consider all quasiprobability representations of quantum mechanics defined on the joint spectrum of $\hat{A}$ and $\hat{B}$, and that have the correct marginal Born probabilities for $\hat{A}$ and $\hat{B}$. For any such Born-compatible quasiprobability representation, we show that there exists, for each observable $\hat{X}$, a naturally associated conditional expectation, given $\hat{B}$. In addition, among the aforementioned representations, only the Kirkwood-Dirac representation has the following property: its associated conditional expectation of $\hat{X}$ given $\hat{B}$ coincides with the best predictor of $\hat{X}$ by a function of $\hat{B}$, for all $\hat{X}$.

1 Introduction

Within quantum mechanics, it is not possible to properly define the joint probability distribution of two incompatible (*i.e.* noncommuting) observables in a coherent manner for all states [1–4]. This impossibility is one of the characteristic features of quantum mechanics, distinguishing it from classical mechanics. For the prototypical incompatible observables of position and momentum, Wigner famously found a way around this impossibility by associating to each quantum state a quasiprobability distribution for position and momentum, having the correct marginals, but that can take negative values [5]. A large variety of other quasiprobability distributions for position and momentum have subsequently been developed and have been proven useful in various fields of physics [6–11], even though not all these constructions yield joint quasiprobability distributions: their marginals do not always coincide with the Born-rule probabilities for position and momentum. Extensions of these constructions to abelian groups other than the space translations, and their associated Heisenberg groups, have also been developed [12–17]. Other approaches include coherent state constructions [18–21]. All these approaches fit in a general unifying setup using frames and dual frames on the quantum Hilbert space. This setup allows to define a quasiprobability distribution $Q(\hat{\rho})$ for all quantum states $\hat{\rho}$ and classical symbols $\tilde{Q}(\hat{X})$ for all quantum observables $\hat{X}$ [3], both defined on a space Λ, thus forming a quasiprobability representation of quantum mechanics.

Among the countless quasiprobability representations of quantum mechanics so obtained feature the ones introduced by Dirac [8]. Given any fixed pair of noncommuting observables $\hat{A}$ and $\hat{B}$, Dirac proposed for each quantum state $\hat{\rho}$ two joint quasiprobability distributions $Q^{\mathrm{KD},\ell}(\hat{\rho})$

and $Q^{\mathrm{KD,r}}(\hat\rho)$ (both with the correct marginals, therefore) on the space $\sigma(\hat A)\times\sigma(\hat B)$, where $\sigma(\hat A)$ and $\sigma(\hat B)$ denote the spectra of $\hat A$ and $\hat B$:

$$Q^{\mathrm{KD},\ell}_{a,b}(\hat\rho) = \mathrm{Tr}\Big(\Pi^{\hat A}_a\hat\rho\Pi^{\hat B}_b\Big),\quad Q^{\mathrm{KD,r}}_{a,b}(\hat\rho) = \mathrm{Tr}\Big(\Pi^{\hat B}_b\hat\rho\Pi^{\hat A}_a\Big), \tag{1.1}$$

where $\Pi^{\hat A}_a$ and $\Pi^{\hat B}_b$ are the spectral projectors of $\hat A$ onto the eigenspace of $a\in\sigma(\hat A)$ and of $\hat B$ onto the eigenspace of $b\in\sigma(\hat B)$, respectively. The reason for the choice of superscript ℓ, for "left" and r for "right" will become clear below. $Q^{\mathrm{KD},\ell}$ and $Q^{\mathrm{KD,r}}$ are today referred to as the Kirkwood-Dirac distributions [8]: they do indeed generalize the construction of Kirkwood [7] for the special case of position and momentum. The naturally associated Kirkwood-Dirac symbols $\tilde Q^{\mathrm{KD},\ell}(\hat X)$ and $\tilde Q^{\mathrm{KD,r}}(\hat X)$ for any operator $\hat X$, correspond to a choice of ordering between the operators $\hat A$ and $\hat B$: $\hat A$ before $\hat B$ for ℓ and $\hat B$ before $\hat A$ for r. Indeed, the Kirkwood-Dirac symbol $\tilde Q^{\mathrm{KD},\ell}(\hat X)$ of the operator $\hat X = f(\hat A)g(\hat B)$ is $f(a)g(b)$, whereas the Kirkwood-Dirac symbol $\tilde Q^{\mathrm{KD,r}}(\hat X)$ of $\hat X = g(\hat B)f(\hat A)$ is $f(a)g(b)$. We refer to Section 4 for details. The flexibility of this construction, applying as it does to general noncommuting observables $\hat A, \hat B$, has proven to be an asset in an increasing number of applications in the last few years; we refer to the recent reviews [4, 22] for details.

The question then arises: what singles out the Kirkwood-Dirac quasiprobability representations of quantum mechanics among all possible such representations. We will show in this paper that they are, in a sense to be explained below, uniquely well-behaved with respect to the notion of conditional expectations in quantum mechanics.

Let us point out that there are at least two quite natural but – as we will see – nevertheless different approaches to the definition of quantum conditional expectations. First, in probability theory, the notion of conditional expectation $\mathbb{E}_{\mathbb{P}}(X|Y)$ of a random variable X, given a random variable Y, both defined on an underlying probability space Ω with probability $\mathbb{P}$, can be defined in terms of their conditional probability distribution, which in turn is defined in terms of their joint probability distribution $\mathbb{P}(X = x, Y = y)$. Since, as pointed out above, such joint probabilities do not exist in quantum mechanics for noncommuting observables, this definition cannot straightforwardly be adapted to the quantum context. One approach consists in replacing probabilities by quasiprobabilities and then proceeding in analogy with the classical treatment [23, 24]. As we explain in Section 4, this approach leads to a notion of conditional expectation $\mathbb{E}^Q_{\hat\rho}(\hat X|\hat Y)$ as an operator that depends not only on $\hat\rho$, but also on the quasiprobability distribution Q used in its definition. In addition, as we will show, the conditional expectation $\mathbb{E}^Q_{\hat\rho}(\hat X|\hat Y)$ so defined does not generally have all the natural properties one would expect from a conditional expectation. A second, slightly more sophisticated definition of the conditional expectation $\mathbb{E}_{\mathbb{P}}(X|Y)$ in probability theory is as the function $f(Y)$ of Y that minimizes

$$\mathbb{E}_{\mathbb{P}}(|X-f(Y)|^2) = \int_\Omega |X(\omega)-f(Y(\omega))|^2\mathrm{d}\mathbb{P}(\omega), \tag{1.2}$$

$\mathbb{P}$ being the underlying probability measure. We note in passing and for later reference that the random variable $\mathbb{E}_{\mathbb{P}}(X|Y)$ is real valued when X is real valued. In statistics, $\mathbb{E}_{\mathbb{P}}(X|Y)$ is referred to as the best predictor of X by a function of Y. (We refer to Section 2 for details.) A similar construction has been used to define a conditional expectation in the quantum mechanical context [25–28]. If $\hat X$ and $\hat Y$ are operators on a Hilbert space $\mathcal{H}$ with $\hat Y$ self-adjoint and if $\hat\rho$ is a given mixed state (positive trace-class operator of trace 1), one considers the following expression Eq. (1.2):

$$\mathrm{Tr}(\hat\rho(\hat X - f(\hat Y))^\dagger(\hat X - f(\hat Y))) = \mathrm{Tr}((\hat X - f(\hat Y))\hat\rho(\hat X - f(\hat Y))^\dagger), \tag{1.3}$$

which is a natural analog of Eq. (1.2) One then defines the conditional expectation $\mathbb{E}^\ell_{\hat\rho}(\hat X|\hat Y)$ as that operator function $f(\hat Y)$ of $\hat Y$ which minimizes the expression in Eq. (1.3). As we will show in Theorem 3.3, the map $\hat X \to \mathbb{E}^\ell_{\hat\rho}(\hat X|\hat Y)$ is uniquely characterized by the following two properties, naturally associated with a conditional expectation:

(i) $\mathbb{E}^{\ell}_{\hat\rho}(f(\hat Y)\hat X|\hat Y) = f(\hat Y)\mathbb{E}^{\ell}_{\hat\rho}(\hat X|\hat Y)$, for all functions $f(\hat Y)$ of $\hat Y$ and for all $\hat X \in \mathcal{L}(H)$;

(ii) Iterated expectations: $\mathbb{E}_{\hat\rho}(\mathbb{E}^{\ell}_{\hat\rho}(\hat X|\hat Y)) = \mathbb{E}_{\hat\rho}(\hat X)$.

Here we wrote $\mathbb{E}_{\hat\rho}(\hat X) := \mathrm{Tr}(\hat X\hat\rho)$, to stress the similarity with the analogous properties that hold for the classical notion of conditional expectation, with $\hat\rho$ playing the role of the classical probability measure $\mathbb{P}$. The superscript ℓ stands for "left" because in property (i) above $f(\hat Y)$ appears to the left of $\hat X$. Although this definition of a conditional expectation seems considerably more intrinsic than the one that relies on the choice of a quasiprobability representation, the appearance of this ordering still points to the fact that this definition cannot be unique either, and that choices must have been made in its construction. It turns out that the choice of Eq. (1.3) as the quantum equivalent to Eq. (1.2) of the quantity to be minimized hides a choice of ordering. This should be expected, since the passage from the classical observable $\mathbb{E}_{\mathbb{P}}(X|Y)$ to a quantum equivalent involves a choice of quantization of observables, and as such can be expected to imply a choice of ordering. Indeed, an equally natural quantum analog of Eq. (1.2) is

$$\mathrm{Tr}(\hat\rho(\hat X - f(\hat Y))(\hat X - f(\hat Y))^\dagger) = \mathrm{Tr}((\hat X - f(\hat Y))^\dagger\hat\rho(\hat X - f(\hat Y))). \tag{1.4}$$

Note that this quantity differs from Eq. (1.3) by the order in which $(\hat X - f(\hat Y))$ and $(\hat X - f(\hat Y))^\dagger$ appear. The same minimization procedure as above then leads to a different conditional expectation, that we denote by $\mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat X|\hat Y)$. Again, the map $\hat X \to \mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat X|\hat Y)$ is uniquely characterized by property (ii) above as well as by the following version of (i):

$$\mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat X f(\hat Y)|\hat Y) = f(\hat Y)\mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat X|\hat Y), \quad \forall f(\hat Y), \forall \hat X \in \mathcal{L}(\mathcal{H}), \tag{1.5}$$

in which $f(\hat Y)$ appears to the right of $\hat X$. One has

$$\mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat X|Y) = \mathbb{E}^{\ell}_{\hat\rho}(\hat X^\dagger|Y)^\dagger. \tag{1.6}$$

Our main result in this paper (Theorem 4.5) links the "left" and "right" Kirkwood-Dirac representations to the "left" and "right" conditional expectations $\mathbb{E}^{\ell}_{\hat\rho}(\hat X|\hat Y)$ and $\mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat X|\hat Y)$. It can be paraphrased as follows. Fix an observable $\hat B$. Consider, for any observable $\hat A$ that is incompatible with $\hat B$, any quasiprobability representation Q on $\sigma(\hat A)\times\sigma(\hat B)$ that has the correct Born marginals for $\hat A$ and for $\hat B$. Then one has that

$$\forall\hat\rho \in D_{\hat B}, \forall \hat X \in \mathcal{L}(\mathcal{H}), E^{Q}_{\hat\rho}(\hat X|\hat B) = E^{\ell}_{\hat\rho}(\hat X|\hat B) \quad \Leftrightarrow \quad Q = \mathrm{Q}^{\mathrm{KD},\ell}_{a,b}. \tag{1.7}$$

Here $D_{\hat B} = \{\hat\rho|\langle\varphi^{\hat B}_b, \hat\rho\varphi^{\hat B}_b\rangle \neq 0, \forall b \in \sigma(\hat B)\}$. In other words, among all Born-compatible quasiprobability representations on $\sigma(\hat A)\times\sigma(\hat B)$, only the "left" Kirkwood-Dirac representations have an associated conditional expectation for all operators $\hat X$ that coincides with $\mathbb{E}^{\ell}_{\hat\rho}(\hat X|\hat B)$. In view of what precedes, an equivalent statement of this result is that among all Born-compatible quasiprobability representations on $\sigma(\hat A)\times\sigma(\hat B)$, only the "left" Kirkwood-Dirac representations satisfy the properties (i) and (ii) above. The same statements hold with ℓ replaced by r. These results single out the Kirkwood-Dirac representations among all possible Born-compatible quasiprobability representations on $\sigma(\hat A)\times\sigma(\hat B)$.

The paper is organized as follows. In Section 2, we recall two of the standard definitions of the notion of conditional expectation in probability theory. We then provide a characterization of this notion (Theorem 2.3), the extension of which to the quantum context will be essential for our further analysis. In Section 3 we introduce quantum conditional expectations as minimizers. We provide a characterization of these conditional expectations akin to the classical one (Theorem 3.3). In Section 4 we prove our main result, Theorem 4.5. To avoid technical complications, we restrict ourselves in this paper to finite probability spaces Ω and to finite dimensional Hilbert spaces $\mathcal{H}$.

2 Classical conditional expectation

Let Ω be a finite set. Let $X : \Omega \to \mathbb{C}$ be a complex random variable and $Y : \Omega \to \mathbb{R}$ be a real random variable. We write $\mathrm{Ran}(X)$ for the range of X, meaning

$$\mathrm{Ran}(X) = \{x \in \mathbb{C} \mid \exists \omega \in \Omega, X(\omega) = x\} = X(\Omega), \tag{2.1}$$

and similarly for Y. We denote by D_Y the set of probability measures $\mathbb{P}$ on Ω such that for each $y \in \mathrm{Ran}(Y)$, $\mathbb{P}(Y = y) > 0$. The space of complex random variables on Ω is denoted by $\mathcal{F}(\Omega; \mathbb{C})$. For $\mathbb{P} \in D_Y$, we equip this space with the following sesquilinear form:

$$\langle X, X' \rangle_{\mathbb{P}} = \mathbb{E}_{\mathbb{P}}(\overline{X} X') = \int_\Omega \overline{X}(\omega) X'(\omega) \mathrm{d}\mathbb{P} = \sum_{\omega \in \Omega} \overline{X}(\omega) X'(\omega) \mathbb{P}(\{\omega\}). \tag{2.2}$$

We first recall the elementary definition of "conditional expectation" of a random variable.

Definition 2.1. *Let $\mathbb{P} \in D_Y$. For $y \in \mathrm{Ran}(Y)$ and $x \in \mathbb{C}$, the conditional probability that X takes the value x knowing $Y = y$ is given by*

$$\mathbb{P}(X = x | Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}. \tag{2.3}$$

The conditional expectation of X knowing that $Y = y$ is

$$\mathbb{E}_{\mathbb{P}}(X | Y = y) = \sum_{x \in \mathrm{Ran(X)}} x \mathbb{P}(X = x | Y = y) = \sum_{x \in \mathrm{Ran(X)}} x \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}. \tag{2.4}$$

We will, as usual, define the conditional expectation of X, knowing Y, denoted by $\mathbb{E}_{\mathbb{P}}(X|Y)$, as the function

$$\mathbb{E}_{\mathbb{P}}(X|Y)(y) = \mathbb{E}_{\mathbb{P}}(X|Y = y). \tag{2.5}$$

Note that $\mathbb{E}_{\mathbb{P}}(X|Y)$ is a random variable that is a function of Y. Let us note that

$$\mathbb{E}_{\mathbb{P}}(f(Y)|Y) = f(Y), \tag{2.6}$$

as is readily checked. We note that this definition, and all the properties that follow in this sections, are strongly dependent on the initially chosen probability $\mathbb{P}$ on Ω. We indicate this dependence in the notation, to anticipate on the quantum conditional probability defined in the next section, which also depends on the state $\hat{\rho}$ considered.

We now provide two distinct characterizations of $\mathbb{E}_{\mathbb{P}}(X|Y)$ that will be essential in the following sections in order to define a quantum conditional expectation.

Theorem 2.2. *Let $\mathbb{P} \in D_Y$. The conditional expectation of X knowing Y is the unique minimizer of*

$$f \mapsto \mathbb{E}_{\mathbb{P}}(|X - f(Y)|^2) = \langle X - f(Y), X - f(Y) \rangle_{\mathbb{P}}, \tag{2.7}$$

over the set of complex valued functions f defined on $\mathrm{Ran}(Y)$.

We denote by $\mathcal{Y}$ the set of all complex random variables on Ω that are functions of Y. In other words, $Z : \Omega \to \mathbb{C}$ belongs to $\mathcal{Y}$ provided Z is constant on all level sets of Y. This means there exists a function $f : \mathrm{Ran}(Y) \to \mathbb{C}$ so that $Z = f(Y)$. Note that $\mathcal{Y}$ is a complex vector space ($\mathcal{Y} \subset \mathcal{F}(\Omega; \mathbb{C})$); it is in addition closed under multiplication of functions. In the usual language of Hilbert space theory, the minimizer of Eq. (2.7) – which is the conditional expectation of X given Y – is the orthogonal projection of X onto $\mathcal{Y}$. In more advanced treatments of probability, where the space Ω is not finite and where the probability measures are general, this property provides a simple way to *define* the conditional expectation, since Definition 2.1 then does not necessarily make sense. We will see in the next section that this definition adapts nicely to quantum theory as well.

Proof. For $f : \mathrm{Ran}(Y) \to \mathbb{C}$, we have

$$f(y) = \sum_{y' \in \mathrm{Ran}(Y)} f(y') \mathbb{1}_{y'}(y). \tag{2.8}$$

One then computes

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}(|X - f(Y)|^2) &= \mathbb{E}_{\mathbb{P}}(|X|^2) - 2 \sum_{y \in \mathrm{Ran}(Y)} \mathrm{Re}\left(\overline{f(y)} \mathbb{E}_{\mathbb{P}}(X \mathbb{1}_y(Y))\right) + \sum_{y \in \mathrm{Ran}(Y)} |f(y)|^2 \mathbb{P}(Y = y) \\
&= \mathbb{E}_{\mathbb{P}}(|X|^2) + \sum_{y \in \mathrm{Ran}(Y)} \mathbb{P}(Y = y) \left(|f(y)|^2 - 2 \frac{\mathrm{Re}\left(\overline{f(y)} \mathbb{E}_{\mathbb{P}}(X \mathbb{1}_y(Y))\right)}{\mathbb{P}(Y = y)} \right) \\
&= \mathbb{E}_{\mathbb{P}}(|X|^2) - \sum_{y \in \mathrm{Ran}(Y)} \frac{|\mathbb{E}_{\mathbb{P}}(X \mathbb{1}_y(Y))|^2}{\mathbb{P}(Y = y)} + \sum_{y \in \mathrm{Ran}(Y)} \mathbb{P}(Y = y) \left| f(y) - \frac{\mathbb{E}_{\mathbb{P}}(X \mathbb{1}_y)}{\mathbb{P}(Y = y)} \right|^2 .
\end{aligned}$$

This quantity is minimal if and only if for all $y \in \mathrm{Ran}(Y)$,

$$f(y) = \frac{\mathbb{E}_{\mathbb{P}}(X \mathbb{1}_y)}{\mathbb{P}(Y = y)}. \tag{2.9}$$

Moreover, we have that

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}(X \mathbb{1}_y) &= \sum_{(x, y') \in \mathrm{Ran}(X) \times \mathrm{Ran}(Y)} \mathbb{P}(X = x) x \mathbb{P}(Y = y') \mathbb{1}_y(y') = \sum_{x \in \mathrm{Ran}(X)} x \mathbb{P}(X = x) \mathbb{P}(Y = y) \\
&= \sum_{x \in \mathrm{Ran}(X)} x \mathbb{P}(X = x, Y = y) \\
&= \mathbb{P}(Y = y) \mathbb{E}_{\mathbb{P}}(X | Y = y).
\end{aligned}$$

This concludes the proof. ☐

As we shall now prove, the conditional expectation is also characterized by two simple properties:

Theorem 2.3. *Let Y be a real random variable on Ω and $\mathbb{P} \in D_Y$. Then there exists a unique linear map $\mathcal{E}_{\mathrm{cond,Y}} : \mathcal{F}(\Omega; \mathbb{C}) \to \mathcal{Y} \subset \mathcal{F}(\Omega; \mathbb{C})$ that satisfies the following properties :*

1. *For any complex random variable X on Ω and any $f : \mathrm{Ran}(Y) \to \mathbb{C}$,*

$$\mathcal{E}_{\mathrm{cond},Y}(f(Y)X) = f(Y) \mathcal{E}_{\mathrm{cond},Y}(X) \tag{2.10}$$

2. *For any complex random variable X on Ω,*

$$\mathbb{E}_{\mathbb{P}}(\mathcal{E}_{\mathrm{cond},Y}(X)) = \mathbb{E}_{\mathbb{P}}(X). \tag{2.11}$$

One has $\mathcal{E}_{\mathrm{cond},Y}(X) = \mathbb{E}_{\mathbb{P}}(X|Y)$.

Proof. Existence is clear, since it is easily checked that $X \to \mathbb{E}_{\mathbb{P}}(X|Y)$ satisfies the two desired properties. It remains to show uniqueness. Let $\mathcal{E}_{\mathrm{cond},Y} : X \to \mathcal{E}_{\mathrm{cond},Y}(X) \in \mathcal{Y}$ that satisfies properties *1* and *2*. Let X be a random variable. For any $f(Y) \in \mathcal{Y}$, we have

$$\begin{aligned}
\langle f(Y), X - \mathcal{E}_{\mathrm{cond},Y}(X) \rangle_{\mathbb{P}} &= \mathbb{E}_{\mathbb{P}}(\overline{f(Y)}(X - \mathcal{E}_{\mathrm{cond},Y}(X))) \\
&= \mathbb{E}_{\mathbb{P}}(\overline{f(Y)}X) - \mathbb{E}_{\mathbb{P}}(\mathcal{E}_{\mathrm{cond},Y}(\overline{f(Y)}X)) \text{ by } \mathit{1} \\
&= \mathbb{E}_{\mathbb{P}}(\overline{f(Y)}X) - \mathbb{E}_{\mathbb{P}}(\overline{f(Y)}X) \text{ by } \mathit{2} \\
&= 0.
\end{aligned}$$

and of course $\langle f(Y), X - \mathbb{E}_{\mathbb{P}}(X|Y)\rangle_{\mathbb{P}} = 0$ as well. It follows

$$\begin{aligned}
&\langle \mathcal{E}_{\text{cond},Y}(X) - \mathbb{E}_{\mathbb{P}}(X|Y), \mathcal{E}_{\text{cond},Y}(X) - \mathbb{E}_{\mathbb{P}}(X|Y)\rangle_{\mathbb{P}} \\
=&\langle \mathcal{E}_{\text{cond},Y}(X) - X + X - \mathbb{E}_{\mathbb{P}}(X|Y), \mathcal{E}_{\text{cond},Y}(X) - \mathbb{E}_{\mathbb{P}}(X|Y)\rangle_{\mathbb{P}} \\
=&-\langle X - \mathcal{E}_{\text{cond},Y}(X), \mathcal{E}_{\text{cond},Y}(X)\rangle_{\mathbb{P}} + \langle X - \mathbb{E}_{\mathbb{P}}(X|Y), \mathcal{E}_{\text{cond},Y}(X)\rangle_{\mathbb{P}} \\
&+ \langle X - \mathcal{E}_{\text{cond},Y}(X), \mathbb{E}_{\mathbb{P}}(X|Y)\rangle_{\mathbb{P}} - \langle X - \mathbb{E}_{\mathbb{P}}(X|Y), \mathbb{E}_{\mathbb{P}}(X|Y)\rangle_{\mathbb{P}} \\
=&0
\end{aligned}$$

since $\mathbb{E}_{\mathbb{P}}(X|Y), \mathcal{E}_{\text{cond},Y}(X) \in \mathcal{Y}$. Therefore $\mathbb{E}_{\mathbb{P}}(X|Y) = \mathcal{E}_{\text{cond},Y}(X)$ almost everywhere with respect to $\mathbb{P}$. Let $\omega \in \Omega$, then $Y(\omega) \in \text{Ran}(Y)$ and thus, since $\mathbb{P} \in D_Y$, there exists $\omega' \in \Omega$ such that $\mathbb{P}(\{\omega'\}) > 0$ and $Y(\omega) = Y(\omega')$. The fact that $\mathbb{E}_{\mathbb{P}}(X|Y) = \mathcal{E}_{\text{cond},Y}(X)$ almost everywhere with respect to $\mathbb{P}$ then implies $\mathbb{E}_{\mathbb{P}}(X|Y)(\omega') = \mathcal{E}_{\text{cond},Y}(X)(\omega')$. Moreover since they are both functions of Y we have $\mathbb{E}_{\mathbb{P}}(X|Y)(\omega') = \mathbb{E}_{\mathbb{P}}(X|Y)(\omega)$ and $\mathcal{E}_{\text{cond},Y}(X)(\omega') = \mathcal{E}_{\text{cond},Y}(X)(\omega)$. In the end we have indeed $\mathbb{E}_{\mathbb{P}}(X|Y) = \mathcal{E}_{\text{cond},Y}(X)$ on Ω. This concludes the proof. □

Let us point out that one can see quite directly that Eq. (2.10) and Eq. (2.11) imply that $I - \mathcal{E}_{\text{cond},Y}(I)$ is orthogonal to the algebra $\mathcal{Y}$, as well as belonging to $\mathcal{Y}$. This implies

$$\mathcal{E}_{\text{cond},Y}(I) = I, \tag{2.12}$$

where I is the constant function equal to 1 on Ω. Using Eq. (2.10) again, we obtain that

$$\mathcal{E}_{\text{cond},Y}(f(Y)) = f(Y). \tag{2.13}$$

These identities can also be retrieved from $\mathcal{E}_{\text{cond},Y}(X) = \mathbb{E}_{\mathbb{P}}(X|Y)$.

We finish this section by showing that one can recover the joint probability distribution of the random variables X and Y using the expectation and conditional expectation. We will discuss a similar relation between joint quasiprobabilities and the conditional expectation in the quantum realm in section 4.

Proposition 2.4. *For any* $(x, y) \in \text{Ran}(X) \times \text{Ran(Y)}$*, we have*

$$\mathbb{P}(X = x, Y = y) = \mathbb{E}_{\mathbb{P}}[\mathbb{1}_y(Y)\mathbb{E}_{\mathbb{P}}(\mathbb{1}_x(X)|Y)]. \tag{2.14}$$

Proof. We compute

$$\begin{aligned}
\mathbb{E}_{\mathbb{P}}[\mathbb{1}_y(Y)\mathbb{E}_{\mathbb{P}}(\mathbb{1}_x(X)|Y)] &= \sum_{y' \in \text{Ran}(Y)} \mathbb{P}(Y = y')\mathbb{1}_y(y')\mathbb{E}_{\mathbb{P}}(\mathbb{1}_x(X)|Y)(y') \\
&= \mathbb{P}(Y = y)\mathbb{E}_{\mathbb{P}}(\mathbb{1}_x(X)|Y)(y) \\
&= \mathbb{P}(Y = y) \sum_{x' \in \text{Ran}(X)} \mathbb{1}_x(x') \frac{\mathbb{P}(X = x', y = y)}{\mathbb{P}(Y = y)} \\
&= \mathbb{P}(X = x, Y = y).
\end{aligned}$$

□

We finally point out that Y can also be taken complex valued in what precedes. But, to stress the analogy with the quantum mechanical treatment, we have taken Y real valued here.

3 Quantum conditional expectation via minimization

In this section, we show how to adapt the classical characterization of a conditional expectation based on Theorem 2.3 to the quantum context: see Theorem 3.3. We point out that this result reflects the fact that any such adaptation needs to address the "ordering problem". Indeed, whereas for random variables X and Y, one has $f(Y)X = Xf(Y)$, this is no longer generally true when $\hat{X}$ and $\hat{Y}$ are operators on a Hilbert space. We need some preliminaries before stating and proving Theorem 3.3.

Let $\mathcal{H}$ be a Hilbert space of dimension $d < +\infty$. Let $\mathcal{L}(\mathcal{H})$ be the space of operators on $\mathcal{H}$, and $D(\mathcal{H})$ the set of density matrices (positive semi-definite operators of trace 1). We will first adapt the classical construction of a conditional expectation of Theorem 2.2 to the quantum context. For that purpose, we need to identify a suitable quantum analogue of the inner product between classical observables in Eq. (2.2). One may expect such analogues not to be uniquely determined since the inner product involves the square of a classical observable, and in general, quantizations of observables don't respect products: the quantization of the product is usually not the product of the quantizations due to ordering problems. We will see this phenomenon manifests itself here as well.

One choice that presents itself naturally is to define, for any two $\hat{X}, \hat{X}' \in \mathcal{L}(\mathcal{H})$, the sesquilinear form

$$\langle \hat{X}, \hat{X}' \rangle_\ell = \mathrm{Tr}\Big(\hat{\rho}\hat{X}^\dagger \hat{X}'\Big) = \mathrm{Tr}\Big(\hat{X}'\hat{\rho}\hat{X}^\dagger\Big). \tag{3.1}$$

Now, let $\hat{Y}$ be a self-adjoint operator on $\mathcal{H}$ with non-degenerate spectrum and $\hat{X} \in \mathcal{L}(\mathcal{H})$. We denote by $\mathcal{Y}$ the algebra of all operators and that can be expressed as a function of $\hat{Y}$. We denote by $D_{\hat{Y}}$ the convex set of density matrices $\hat{\rho} \in D(\mathcal{H})$ satisfying $\langle \varphi_y, \hat{\rho}\varphi_y \rangle \neq 0$ for all $y \in \sigma(\hat{Y})$, where φ_y is the corresponding eigenvector. For a density matrix $\hat{\rho}$, in analogy with Theorem 2.2, one can then define (following [25–28]) a quantum conditional expectation of $\hat{X}$, knowing $\hat{Y}$, that we shall denote by $\mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})$ as the minimizer of the following quantity:

$$f \mapsto \mathrm{Tr}\Big(\hat{\rho}|\hat{X} - f(\hat{Y})|^2\Big) = \langle \hat{X} - f(\hat{Y}), \hat{X} - f(\hat{Y}) \rangle_\ell, \tag{3.2}$$

over all $f : \sigma(\hat{Y}) \to \mathbb{C}$. We shall discuss the existence and uniqueness of this minimizer below. We shall refer to $\mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})$ as the left conditional expectation of $\hat{\rho}$, for reasons that will become clear shortly. There is a second, equally natural choice for the sesquilinear product, analogous to the inner product in Eq. (2.2), namely

$$\langle \hat{X}, \hat{X}' \rangle_{\mathrm{r}} = \mathrm{Tr}\Big(\hat{\rho}\hat{X}'\hat{X}^\dagger\Big) = \mathrm{Tr}\Big(\hat{X}^\dagger\hat{\rho}\hat{X}'\Big). \tag{3.3}$$

Note that, as anticipated, the difference between Eq. (3.1) and Eq. (3.3) lies indeed in the change of order of appearance of $\hat{X}^\dagger$ and $\hat{X}'$. We then define the right conditional expectation of $\hat{X}$, knowing $\hat{Y}$, that we shall denote by $\mathbb{E}^{\mathrm{r}}_{\hat{\rho}}(\hat{X}|\hat{Y})$ as the minimizer of

$$f \mapsto \langle \hat{X} - f(\hat{Y}), \hat{X} - f(\hat{Y}) \rangle_{\mathrm{r}}, \tag{3.4}$$

The existence and uniqueness of these conditional expectations is guaranteed by the following theorem, under a suitable hypothesis on $\hat{\rho}$.

Theorem 3.1. *Assume that $\hat{\rho} \in D_{\hat{Y}}$. Then the functions given in Eq.* (3.2) *and Eq.* (3.4) *admit unique minimizers given by*

$$f^\ell_0(y) = \frac{\langle \varphi_y, \hat{X}\hat{\rho}\varphi_y \rangle}{\langle \varphi_y, \hat{\rho}\varphi_y \rangle},\ y \in \sigma(\hat{Y}), \quad f^{\mathrm{r}}_0(y) = \frac{\langle \varphi_y, \hat{\rho}\hat{X}\varphi_y \rangle}{\langle \varphi_y, \hat{\rho}\varphi_y \rangle},\ y \in \sigma(\hat{Y}), \tag{3.5}$$

respectively.

Proof. We only prove the "left" case, the "right" case being similar. We compute

$$\begin{aligned}\langle \hat{X}-f(\hat{Y}),\hat{X}-f(\hat{Y})\rangle_\ell &= \mathrm{Tr}\Big(\hat\rho|\hat{X}|^2\Big)-\mathrm{Tr}\Big(\hat\rho\hat{X}^\dagger f(\hat{Y})\Big)-\mathrm{Tr}\Big(\hat\rho f(\hat{Y})^\dagger\hat{X}\Big)+\mathrm{Tr}\Big(\hat\rho|f(\hat{Y})|^2\Big)\\ &= \mathrm{Tr}\Big(\hat\rho|\hat{X}|^2\Big)+\sum_{y\in\sigma(\hat{Y})}\langle\varphi_y,\hat\rho\varphi_y\rangle\left(|f(y)|^2-2\,\mathrm{Re}(\overline{f(y)}\frac{\langle\varphi_y,\hat{X}\hat\rho\varphi_y\rangle}{\langle\varphi_y,\hat\rho\varphi_y\rangle})\right)\\ &= \mathrm{Tr}\Big(\hat\rho|\hat{X}|^2\Big)-\sum_{y\in\sigma(\hat{Y})}\frac{|\langle\varphi_y,\hat{X}\hat\rho\varphi_y\rangle|^2}{\langle\varphi_y,\hat\rho\varphi_y\rangle}+\sum_{y\in\sigma(\hat{Y})}\langle\varphi_y,\hat\rho\varphi_y\rangle\left|f(y)-\frac{\langle\varphi_y,\hat{X}\hat\rho\varphi_y\rangle}{\langle\varphi_y,\hat\rho\varphi_y\rangle}\right|^2.\end{aligned}$$

Hence this quantity is minimal if and only if

$$f(y)=\frac{\langle\varphi_y,\hat{X}\hat\rho\varphi_y\rangle}{\langle\varphi_y,\hat\rho\varphi_y\rangle} \tag{3.6}$$

for all $y\in\sigma(\hat{Y})$. ☐

Definition 3.2. *We define the left quantum conditional expectation of the operator $\hat{X}$ knowing $\hat{Y}$ in the state $\hat\rho$ as*

$$\mathbb{E}^\ell_{\hat\rho}(\hat{X}|\hat{Y})=f^\ell_0(\hat{Y})=\sum_{y\in\sigma(\hat{Y})}\frac{\langle\varphi_y,\hat{X}\hat\rho\varphi_y\rangle}{\langle\varphi_y,\hat\rho\varphi_y\rangle}\Pi_y, \tag{3.7}$$

where Π_y is the orthogonal projector to the eigenvector φ_y for $y\in\sigma(\hat{Y})$. Similarly we define the right quantum conditional expectation of $\hat{X}$ knowing $\hat{Y}$ in the state $\hat\rho$ as

$$\mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat{X}|\hat{Y})=f^{\mathrm{r}}_0(\hat{Y})=\sum_{y\in\sigma(\hat{Y})}\frac{\langle\varphi_y,\hat\rho\hat{X}\varphi_y\rangle}{\langle\varphi_y,\hat\rho\varphi_y\rangle}\Pi_y. \tag{3.8}$$

The quantum conditional expectations $\mathbb{E}^{\ell,\mathrm{r}}$ are therefore the orthogonal projections with respect to the two "inner products" induced by the state $\hat\rho$ of the bounded operator $\hat{X}$ onto the algebra of functions of the operator $\hat{Y}$. Note that

$$\mathbb{E}^{\mathrm{r}}_{\hat\rho}(\hat{X}|\hat{Y})=\mathbb{E}^\ell_{\hat\rho}(\hat{X}^\dagger|\hat{Y})^\dagger. \tag{3.9}$$

Let us point out that $\mathbb{E}^{\ell,\mathrm{r}}_{\hat\rho}(\hat{X}|\hat{Y})$ are not self-adjoint in general, even if we take $\hat{X}$ to be self-adjoint. This is different from the classical situation, where the conditional expectation of a real random variable X, given a real random variable Y, is always real.

We also note that the coefficients $\left(\frac{\langle\varphi_y,\hat{X}\hat\rho\varphi_y\rangle}{\langle\varphi_y,\hat\rho\varphi_y\rangle}\right)_{y\in\sigma(Y)}$ appearing in $\mathbb{E}^\ell_{\hat\rho}(\hat{X}|\hat{Y})$ are the weak values with initial state ρ, a weak measurement of X and a post-selection Π_y for each $y\in\sigma(\hat{Y})$, as has been remarked before [25].

Let us point out that there are many other quantum analogs of the inner product in Eq. (1.2), such as for example

$$\langle\hat{X},\hat{X}'\rangle_\alpha=\alpha\langle\hat{X},\hat{X}'\rangle_\ell+(1-\alpha)\langle X,X'\rangle_{\mathrm{r}}. \tag{3.10}$$

for $0\leqslant\alpha\leqslant1$, interpolating between $\langle X,X'\rangle_\ell$ and $\langle X,X'\rangle_{\mathrm{r}}$. Minimizing $\langle\hat{X}-f(\hat{Y}),\hat{X}-f(\hat{Y})\rangle_\alpha$ leads to a candidate notion of conditional expectation that we denote by $\mathbb{E}^\alpha_{\hat\rho}(\hat{X}|\hat{Y})$. The following theorem shows that to obtain a conditional expectation that satisfies the properties of the classical conditional expectation given in Theorem 2.3, α has to be either 0 or 1.

Theorem 3.3. *Let $\hat\rho\in D_{\hat{Y}}$. For any $\#\in\{\ell,\mathrm{r}\}$, there exist a unique map $\mathcal{E}^{\#}_{\mathrm{cond},Y}:\mathcal{L}(\mathcal{H})\to\mathcal{Y}$ satisfying the following properties:*

1. *For any $\hat{X} \in \mathcal{L}(\mathcal{H})$ and any $f : \sigma(\hat{Y}) \to \mathbb{C}$,*

$$\text{for } \# = \ell,\ \mathcal{E}^{\ell}_{\text{cond},Y}(f(\hat{Y})\hat{X}) = f(\hat{Y})\mathcal{E}^{\ell}_{\text{cond},Y}(\hat{X}) = \mathcal{E}^{\ell}_{\text{cond},Y}(\hat{X})f(\hat{Y}), \tag{3.11}$$

$$\text{for } \# = \mathrm{r},\ \mathcal{E}^{\mathrm{r}}_{\text{cond},Y}(\hat{X}f(\hat{Y})) = f(\hat{Y})\mathcal{E}^{\mathrm{r}}_{\text{cond},Y}(\hat{X}) = \mathcal{E}^{\mathrm{r}}_{\text{cond},Y}(\hat{X})f(\hat{Y}). \tag{3.12}$$

2. *For any $\hat{X} \in \mathcal{L}(\mathcal{H})$,*

$$\mathbb{E}_{\hat{\rho}}(\mathcal{E}^{\#}_{\text{cond},Y}(\hat{X})) = \mathbb{E}_{\hat{\rho}}(\hat{X}). \tag{3.13}$$

This map is given by $\mathcal{E}^{\#}_{\text{cond},Y}(\hat{X}) = \mathbb{E}^{\#}_{\hat{\rho}}(\hat{X}|\hat{Y})$.

It is readily checked that $\mathbb{E}^{\ell,\mathrm{r}}_{\hat{\rho}}(I|\hat{Y}) = I$, and thus for any $f : \sigma(\hat{Y}) \to \mathbb{C}$,

$$\mathbb{E}^{\ell,\mathrm{r}}_{\hat{\rho}}(f(\hat{Y})|\hat{Y}) = f(\hat{Y}). \tag{3.14}$$

The Theorem singles out two natural definitions of quantum conditional expectation through the conditions Eq. (3.11) and (3.13) or Eq. (3.12) and (3.13).

As in the classical case, the uniqueness in Theorem 3.3 proves that Eq. (3.14) can be deduced from Eq. (3.11)-(3.12) and Eq. (3.13). Again, one could have shown this fact directly.

Proof. Again, we write the proof for $\# = \ell$, the other case being similar. We begin by showing that the left quantum conditional expectation satisfies those two properties. For point 1, we compute

$$\begin{aligned}
\mathbb{E}^{\ell}_{\hat{\rho}}(f(\hat{Y})\hat{X}|\hat{Y}) &= \sum_{y\in\sigma(\hat{Y})} \frac{\langle \varphi_y, f(\hat{Y})\hat{X}\hat{\rho}\varphi_y\rangle}{\langle \varphi_y, \hat{\rho}\varphi_y\rangle}\Pi_y \\
&= \sum_{y\in\sigma(\hat{Y})} f(y)\frac{\langle \varphi_y, \hat{X}\hat{\rho}\varphi_y\rangle}{\langle \varphi_y, \hat{\rho}\varphi_y\rangle}\Pi_y \\
&= f(\hat{Y})\mathbb{E}^{\ell}_{\hat{\rho}}(\hat{X}|\hat{Y}) = \mathbb{E}^{\ell}_{\hat{\rho}}(\hat{X}|\hat{Y})f(\hat{Y}).
\end{aligned}$$

For point 2, we compute

$$\begin{aligned}
\mathbb{E}_{\hat{\rho}}(\mathbb{E}^{\ell}_{\hat{\rho}}(\hat{X}|\hat{Y})) &= \mathrm{Tr}\left(\hat{\rho}\sum_{y\in\sigma(\hat{Y})} \frac{\langle \varphi_y, \hat{X}\hat{\rho}\varphi_y\rangle}{\langle \varphi_y, \hat{\rho}\varphi_y\rangle}\Pi_y\right) \\
&= \sum_{y\in\sigma(\hat{Y})} \langle \varphi_y, \hat{X}\hat{\rho}\varphi_y\rangle \\
&= \mathrm{Tr}\left(\hat{X}\hat{\rho}\right) \\
&= \mathbb{E}_{\hat{\rho}}(\hat{X}).
\end{aligned}$$

Now, let $\mathcal{E}^{\ell}_{\text{cond},Y} : \mathcal{L}(\mathcal{H}) \to \mathcal{Y}$ be a map that satisfies (3.11) and (3.13). We will show that $\mathcal{E}^{\ell}_{\text{cond},Y}(\hat{X}) = \mathbb{E}^{\ell}_{\hat{\rho}}(\hat{X}|\hat{Y})$ for any $\hat{X} \in \mathcal{L}(\mathcal{H})$. Let $\hat{X} \in \mathcal{L}(\mathcal{H})$. For any $f(\hat{Y}) \in \mathcal{Y}$, we have

$$\begin{aligned}
\langle f(\hat{Y}), \hat{X} - \mathcal{E}^{\ell}_{\text{cond},Y}(\hat{X})\rangle_{\ell} &= \mathrm{Tr}\left(\hat{\rho}f(\hat{Y})^{\dagger}(\hat{X} - \mathcal{E}^{\ell}_{\text{cond},Y}(f(\hat{X})))\right) \\
&= \mathrm{Tr}\left(\hat{\rho}f(\hat{Y})^{\dagger}\hat{X}\right) - \mathrm{Tr}\left(\hat{\rho}\mathcal{E}^{\ell}_{\text{cond},Y}(f(\hat{Y})^{\dagger}\hat{X})\right) \text{ by (3.11)} \\
&= \mathrm{Tr}\left(\hat{\rho}f(\hat{Y})^{\dagger}\hat{X}\right) - \mathrm{Tr}\left(\hat{\rho}f(\hat{Y})^{\dagger}\hat{X}\right) \text{ by (3.13)} \\
&= 0,
\end{aligned}$$

and of course $\langle f(\hat{Y}), \hat{X} - \mathbb{E}_{\hat{\rho}}(\hat{X}|\hat{Y})\rangle_\ell = 0$ as well. It follows

$$\begin{aligned}
&\langle \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y}), \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})\rangle_\ell \\
=&\langle \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) - \hat{X} + \hat{X} - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y}), \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})\rangle_\ell \\
=&-\langle \hat{X} - \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}), \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X})\rangle_\ell + \langle \hat{X} - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y}), \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X})\rangle_\ell \\
&+\langle \hat{X} - \mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}), \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})\rangle_\ell - \langle \hat{X} - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y}), \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})\rangle_\ell \\
=&0
\end{aligned}$$

since $\mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}), \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y}) \in \mathcal{Y}$. Therefore, $\mathrm{Tr}\Big(\hat{\rho}|\mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})|^2\Big) = 0$. Since $\mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y}) \in \mathcal{Y}$, there exists $C_{\hat{\rho}} : \sigma(\hat{Y}) \to \mathbb{C}$ such that $\mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) - \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y}) = C_{\hat{\rho}}(\hat{Y})$. One gets

$$0 = \mathrm{Tr}\Big(\hat{\rho}|C_{\hat{\rho}}(\hat{Y})|^2\Big) = \sum_{y\in\sigma(\hat{Y})} |C_{\hat{\rho}}(y)|^2 \langle \varphi_y, \hat{\rho}\varphi_y\rangle,$$

which implies, since $\hat{\rho} \in D_{\hat{Y}}$, that $C_{\hat{\rho}} = 0$, *i.e.* $\mathcal{E}^\ell_{\mathrm{cond},Y}(\hat{X}) = \mathbb{E}^\ell_{\hat{\rho}}(\hat{X}|\hat{Y})$. □

The conditional expectation of $\hat{X}$ knowing $\hat{Y}$ is defined only for states $\hat{\rho} \in D_{\hat{Y}}$, to avoid dividing by 0. As we now prove in the following Lemma, this condition on the state $\hat{\rho}$ is not too restrictive, as the set $D_{\hat{Y}}$ is, in a topological sense, a large set (it is dense in the set of all density matrices).

Lemma 3.4. *Let $\hat{Y}$ be a self-adjoint operator on $\mathcal{H}$. Then the set*

$$D_{\hat{Y}} = \{\hat{\rho} \in \mathcal{D}(\mathcal{H}) : \mathrm{Tr}(\hat{\rho}\Pi_y) \neq 0 \ \forall y \in \sigma(\hat{Y})\}$$

is dense in $\mathcal{D}(\mathcal{H})$ (for any norm topology).

Proof. Since we are working in finite dimension, all norms are equivalent. We chose to work with the trace-class norm. Let $\hat{\rho}_0 \in D(\mathcal{H})$. If $\hat{\rho}_0 \in D_{\hat{Y}}$ there is nothing to do. Otherwise, let

$$I_0 = \{y \in \sigma(\hat{Y}) : \mathrm{Tr}(\hat{\rho}_0\Pi_y) = 0\}.$$

We denote by $|I_0|$ the cardinal of that set. For $\varepsilon > 0$, let

$$\hat{\rho}_\varepsilon = \frac{1}{1+\varepsilon|I_0|}\left(\hat{\rho}_0 + \varepsilon\sum_{y\in I_0}\Pi_y\right)$$

Then it is easily checked that $\hat{\rho}_\varepsilon \in D_{\hat{Y}}$ for any $\varepsilon > 0$ and moreover

$$\begin{aligned}
\mathrm{Tr}(|\hat{\rho}_\varepsilon - \hat{\rho}_0|) &= \mathrm{Tr}\left(\left|\hat{\rho}_0\left(\frac{1}{1+\varepsilon|I_0|} - 1\right) + \frac{\varepsilon}{1+\varepsilon|I_0|}\sum_{y\in I_0}\Pi_y\right|\right) \\
&\leqslant \left|1 - \frac{1}{1+\varepsilon|I_0|}\right| + \frac{\varepsilon|I_0|}{1+\varepsilon|I_0|} \\
&\xrightarrow[\varepsilon\to 0]{} 0.
\end{aligned}$$

This proves the lemma. □

4 Quantum conditional expectation via quasiprobability representation of quantum mechanics

As shown in the previous section, defining the quantum conditional expectation via a minimization problem as in Eq. (3.2) or Eq. (3.4) is a way of mimicking what happens classically, see Theorem 2.2. One can alternatively try to imitate the classical conditional expectation by using Definition 2.1. This definition uses the joint probability distribution of the two random variables under consideration, which is naturally defined in classical probability theory. In quantum mechanics, as recalled in the introduction, a notion of joint probability distribution does not exist for noncommuting observables, but one can try to use a quasiprobability distribution instead.

In this section we fix $\hat{A}$ and $\hat{B}$ two non-degenerate self-adjoint operators on $\mathcal{H}$. We will assume that for any couple $(a,b) \in \sigma(\hat{A}) \times \sigma(\hat{B})$, $\langle \varphi_a^{\hat{A}}, \varphi_b^{\hat{B}} \rangle \neq 0$, where $(\varphi_a^{\hat{A}})_{a\in\sigma(A)}$ are the eigenvectors of $\hat{A}$ and $(\varphi_b^{\hat{B}})_{b\in\sigma(A)}$ are the ones of $\hat{B}$. This condition can be seen as a form of incompatibility slightly stronger than merely requiring that the operators do not commute [29, 30]. In this case the projectors $(\Pi_b^{\hat{B}}\Pi_a^{\hat{A}})_{(a,b)\in\sigma(A)\times\sigma(B)}$ and $(\Pi_a^{\hat{A}}\Pi_b^{\hat{B}})_{(a,b)\in\sigma(A)\times\sigma(B)}$ form bases of $\mathcal{L}(\mathcal{H})$.

Definition 4.1. *A frame on $\sigma(\hat{A}) \times \sigma(\hat{B})$ is a family of operators $S = (S_{a,b})_{(a,b)\in\sigma(\hat{A})\times\sigma(\hat{B})} \subset \mathcal{L}(\mathcal{H})$ such that there exists $(C_1, C_2) \in \mathbb{R}_*^+$ satisfying that for all $\hat{X} \in \mathcal{L}(\mathcal{H})$*

$$C_1 \mathrm{Tr}(\hat{X}\hat{X}^\dagger) \leqslant \sum_{(a,b)\in\sigma(A)\times\sigma(B)} \left|\mathrm{Tr}(S_{a,b}^\dagger \hat{X})\right|^2 \leqslant C_2 \mathrm{Tr}(\hat{X}\hat{X}^\dagger).$$

The frame is said to be Born compatible for $\hat{A}$ and $\hat{B}$ if for any $(a_0, b_0) \in \sigma(\hat{A}) \times \sigma(\hat{B})$,

$$\sum_{a\in\sigma(\hat{A})} S_{a,b_0}^\dagger = \Pi_{b_0}^{\hat{B}} \text{ and } \sum_{b\in\sigma(\hat{B})} S_{a_0,b}^\dagger = \Pi_{a_0}^{\hat{A}}. \tag{4.1}$$

We note that a frame as defined in Definition 4.1 is in fact a basis for $\mathcal{L}(\mathcal{H})$. Moreover, Born compatibility implies that:

$$\sum_{(a,b)\in\sigma(\hat{A})\times\sigma(\hat{B})} S_{a,b} = \mathrm{I}.$$

It has a unique dual frame $(T_{a,b})_{(a,b)\in\sigma(\hat{A})\times\sigma(\hat{B})}$ that satisfies

$$\forall (a,b), (a',b') \in \sigma(\hat{A}) \times \sigma(\hat{B}), \mathrm{Tr}\left(T_{a',b'} S_{a,b}^\dagger\right) = \delta_{a,a'}\delta_{b,b'}, \tag{4.2}$$

see [31, 32].

Definition 4.2. *Let $(S_{a,b})$ be a Born-compatible frame for $\hat{A}$ and $\hat{B}$ with $(T_{a,b})$ its unique dual frame. One can associate to any state $\hat{\rho}$ on $\mathcal{H}$ a quasiprobability distribution on $\sigma(\hat{A}) \times \sigma(\hat{B})$ defined by:*

$$\forall (a,b) \in \sigma(\hat{A}) \times \sigma(\hat{B}), Q_{a,b}(\hat{\rho}) = \mathrm{Tr}\left(\hat{\rho} S_{a,b}^\dagger\right). \tag{4.3}$$

One can also associate to any operator $\hat{X} \in \mathcal{L}(\mathcal{H})$ a symbol on $\sigma(\hat{A}) \times \sigma(\hat{B})$ defined by:

$$\forall (a,b) \in \sigma(\hat{A}) \times \sigma(\hat{B}), \tilde{Q}_{a,b}(\hat{X}) = \mathrm{Tr}\left(\hat{X} T_{a,b}^\dagger\right) \tag{4.4}$$

The couple $(Q, \tilde{Q})$ is said to be a quasiprobability representation of quantum mechanics as it satisfies the "overlap formula":

$$\forall \hat{\rho} \in D(\mathcal{H}), \forall \hat{X} \in \mathcal{L}(\mathcal{H}),\ \mathrm{Tr}\left(\hat{\rho}\hat{X}^\dagger\right) = \sum_{(a,b)\in\sigma(\hat{A})\times\sigma(\hat{B})} \overline{\tilde{Q}_{a,b}(\hat{X})} Q_{a,b}(\hat{\rho}). \tag{4.5}$$

Let $(S_{a,b})$ be a Born-compatible frame for $\hat{A}$ and $\hat{B}$. Let $\hat{\rho} \in \mathcal{D}(\mathcal{H})$. It is easily checked that for any $(a_0, b_0) \in \sigma(\hat{A}) \times \sigma(\hat{B})$,

$$\sum_{a\in\sigma(\hat{A})} Q_{a,b_0}(\hat{\rho}) = \langle \varphi^{\hat{B}}_{b_0}, \hat{\rho}\varphi^{\hat{B}}_{b_0}\rangle \text{ and } \sum_{b\in\sigma(\hat{B})} Q_{a_0,b}(\hat{\rho}) = \langle \varphi^{\hat{A}}_{a_0}, \hat{\rho}\varphi^{\hat{A}}_{a_0}\rangle. \tag{4.6}$$

The function $(a, b) \mapsto Q_{a,b}(\hat{\rho})$ therefore plays the role in the quantum realm of a "joint quasiprobability distribution" associated with $\hat{A}$ and $\hat{B}$, with the particularity that it is not a positive valued function in general. One is then tempted to define, as was done in [22, 25] a conditional quasiprobability of the observable $\hat{A}$ knowing that the observable $\hat{B}$ takes the value $b \in \sigma(\hat{B})$, in the state $\hat{\rho} \in D_{\hat{B}}$ by:

$$\forall a \in \sigma(\hat{A}),\ Q_{a|b}(\hat{\rho}) = \frac{Q_{a,b}(\hat{\rho})}{\langle \varphi^{\hat{B}}_{b}, \hat{\rho}\varphi^{\hat{B}}_{b}\rangle}. \tag{4.7}$$

It is then natural to define the following conditional expectation associated to the quasiprobability distribution Q, following the classical definition given by Definition 2.1 :

Definition 4.3. *Let S be a Born-compatible frame for $\hat{A}$ and $\hat{B}$ and $(Q, \tilde{Q})$ be the associated quasiprobability representation of quantum mechanics. The Q-conditional expectation of $\hat{X} \in \mathcal{L}(\mathcal{H})$ knowing $\hat{B}$ in the state $\hat{\rho} \in D_{\hat{B}}$ is given by*

$$\mathbb{E}^{Q}_{\hat{\rho}}(\hat{X}|\hat{B}) = \sum_{(a,b)\in\sigma(\hat{A})\times\sigma(\hat{B})} \overline{\tilde{Q}_{a,b}(\hat{X}^{\dagger})} Q_{a|b}(\hat{\rho})\Pi^{\hat{B}}_{b}. \tag{4.8}$$

To see why this definition is quite natural, we first show it satisfies two properties analogous to Eq. (2.10) and to Proposition 2.4 respectively.

Proposition 4.4. *The Q-conditional expectation has the following property: for any $\hat{\rho} \in D_{\hat{B}}$ and any $\hat{X} \in \mathcal{L}(\mathcal{H})$,*

$$\mathbb{E}_{\hat{\rho}}(\mathbb{E}^{Q}_{\hat{\rho}}(\hat{X}|\hat{B})) = \mathbb{E}_{\hat{\rho}}(\hat{X}). \tag{4.9}$$

In particular, for any $\rho \in D_{\hat{B}}$ and any $(a, b) \in \sigma(\hat{A}) \times \sigma(\hat{B})$

$$Q_{a,b}(\hat{\rho}) = \mathrm{Tr}\left(\rho S^{\dagger}_{a,b}\right) = \mathbb{E}_{\hat{\rho}}(\mathbb{E}^{Q}_{\hat{\rho}}(S^{\dagger}_{a,b}|\hat{B})). \tag{4.10}$$

However, as we will establish below, generally a Q-conditional expectation does *not* satisfy the pull-through formulas Eq. (3.11) or Eq. (3.12).

Proof. It is a short computation that follows from (4.5):

$$\begin{aligned}
\mathbb{E}_{\hat{\rho}}(\mathbb{E}^{Q}_{\hat{\rho}}(\hat{X}|\hat{B})) &= \sum_{a,b\in\sigma(\hat{A})\times\sigma(\hat{B})} \overline{\tilde{Q}_{a,b}(\hat{X}^{\dagger})} \frac{Q_{a,b}(\hat{\rho})}{\langle \varphi^{\hat{B}}_{b}, \hat{\rho}\varphi^{\hat{B}}_{b}\rangle} \mathrm{Tr}\left(\hat{\rho}\Pi^{\hat{B}}_{b}\right) \\
&= \sum_{a,b} \overline{\tilde{Q}_{a,b}(\hat{X}^{\dagger})} Q_{a,b}(\hat{\rho}) \\
&= \mathrm{Tr}\left(\hat{\rho}\hat{X}\right).
\end{aligned}$$

□

We now introduce the left and right Kirkwood-Dirac distributions associated with $\hat{A}$ and $\hat{B}$. Two natural choices of Born-compatible frames for $\hat{A}$ and $\hat{B}$ are

$$S^{\ell}_{a,b} = \Pi^{\hat{A}}_{a}\Pi^{\hat{B}}_{b} \text{ and } S^{\mathrm{r}}_{a,b} = \Pi^{\hat{B}}_{b}\Pi^{\hat{A}}_{a}.$$

It is then easily shown that the dual frames are

$$T^{\ell}_{a,b} = \frac{1}{|\langle \varphi^{\hat{A}}_a, \varphi^{\hat{B}}_b\rangle|^2}\Pi^{\hat{A}}_a\Pi^{\hat{B}}_b \text{ and } T^{\mathrm{r}}_{a,b} = \frac{1}{|\langle \varphi^{\hat{A}}_a, \varphi^{\hat{B}}_b\rangle|^2}\Pi^{\hat{B}}_b\Pi^{\hat{A}}_a.$$

The associated quasiprobability representations are the left and right Kirkwood-Dirac representations:

$$\begin{aligned}
\mathrm{Q}^{\mathrm{KD},\ell}_{a,b}(\hat{\rho}) &= \langle \varphi^{\hat{B}}_b, \varphi^{\hat{A}}_a\rangle\langle \varphi^{\hat{A}}_a, \hat{\rho}\varphi^{\hat{B}}_b\rangle, \ \tilde{\mathrm{Q}}^{\mathrm{KD},\ell}_{a,b}(\hat{X}) = \frac{1}{\langle \varphi^{\hat{A}}_a, \varphi^{\hat{B}}_b\rangle}\langle \varphi^{\hat{A}}_a, \hat{X}\varphi^{\hat{B}}_b\rangle \\
\mathrm{Q}^{\mathrm{KD},\mathrm{r}}_{a,b}(\hat{\rho}) &= \langle \varphi^{\hat{A}}_a, \varphi^{\hat{B}}_b\rangle\langle \varphi^{\hat{B}}_b, \hat{\rho}\varphi^{\hat{A}}_a\rangle, \ \tilde{\mathrm{Q}}^{\mathrm{KD},\mathrm{r}}_{a,b}(\hat{X}) = \frac{1}{\langle \varphi^{\hat{B}}_b, \varphi^{\hat{A}}_a\rangle}\langle \varphi^{\hat{B}}_b, \hat{X}\varphi^{\hat{A}}_a\rangle.
\end{aligned}$$

They are called the left and right Kirkwood-Dirac representations because of the following fact. For $f : \sigma(\hat{A}) \to \mathbb{C}$ and $g : \sigma(\hat{B}) \to \mathbb{C}$, one gets

$$\tilde{\mathrm{Q}}^{\mathrm{KD},\ell}_{a,b}(f(\hat{A})g(\hat{B})) = f(a)g(b) \tag{4.11}$$

so the symbol is well behaved when $f(\hat{A})$ is "to the left", and

$$\tilde{\mathrm{Q}}^{\mathrm{KD},\mathrm{r}}_{a,b}(g(\hat{B})f(\hat{A})) = g(b)f(a), \tag{4.12}$$

the symbol is well behaved when $f(\hat{A})$ is "to the right". Following Definition 4.3, one can define the associated Q-conditional expectations: for any $\hat{\rho} \in D_{\hat{B}}$ and any $\hat{X} \in \mathcal{L}(\mathcal{H})$,

$$\begin{aligned}
\mathbb{E}^{\mathrm{Q}^{\mathrm{KD},\ell}}_{\hat{\rho}}(\hat{X}|\hat{B}) &= \sum_{(a,b)\in\sigma(\hat{A})\times\sigma(\hat{B})} \overline{\frac{1}{\langle \varphi^{\hat{A}}_a, \varphi^{\hat{B}}_b\rangle}\langle \varphi^{\hat{A}}_a, \hat{X}^\dagger\varphi^{\hat{B}}_b\rangle}\frac{\langle \varphi^{\hat{B}}_b, \varphi^{\hat{A}}_a\rangle\langle \varphi^{\hat{A}}_a, \hat{\rho}\varphi^{\hat{B}}_b\rangle}{\langle \varphi^{\hat{B}}_b, \hat{\rho}\varphi^{\hat{B}}_b\rangle}\Pi^{\hat{B}}_b \\
&= \sum_{(a,b)\in\sigma(\hat{A})\times\sigma(\hat{B})} \frac{\langle \varphi^{\hat{B}}_b, \hat{X}\varphi^{\hat{A}}_a\rangle\langle \varphi^{\hat{A}}_a, \hat{\rho}\varphi^{\hat{B}}_b\rangle}{\langle \varphi^{\hat{B}}_b, \hat{\rho}\varphi^{\hat{B}}_b\rangle}\Pi^{\hat{B}}_b \\
&= \sum_{b\in\sigma(\hat{B})} \frac{\langle \varphi^{\hat{B}}_b, \hat{X}\hat{\rho}\varphi^{\hat{B}}_b\rangle}{\langle \varphi^{\hat{B}}_b, \hat{\rho}\varphi^{\hat{B}}_b\rangle}\Pi^{\hat{B}}_b.
\end{aligned}$$

One recognizes the left conditional expectation given in Definition 3.2, in other words

$$\mathbb{E}^{\mathrm{Q}^{\mathrm{KD},\ell}}_{\hat{\rho}}(\hat{X}|\hat{B}) = \mathbb{E}^{\ell}_{\hat{\rho}}(\hat{X}|\hat{B}). \tag{4.13}$$

Similarly,

$$\mathbb{E}^{\mathrm{Q}^{\mathrm{KD},\mathrm{r}}}_{\hat{\rho}}(\hat{X}|\hat{B}) = \mathbb{E}^{\mathrm{r}}_{\hat{\rho}}(\hat{X}|\hat{B}). \tag{4.14}$$

In particular, those two Q-conditional expectations satisfy respectively Eq. (3.11) and Eq. (3.12). We shall now prove that they are the only Q-conditional expectations with this property, singling out the Kirkwood-Dirac distributions among the quasiprobability representations that are Born compatible with $\hat{A}$ and $\hat{B}$.

Theorem 4.5. *Let S be a Born-compatible frame for $\hat{A}$ and $\hat{B}$ and $(Q, \tilde{Q})$ be the associated quasiprobability representation of quantum mechanics. The Q-conditional expectation satisfies*

$$\mathbb{E}^{Q}_{\hat{\rho}}(f(\hat{B})\hat{X}|\hat{B}) = f(\hat{B})\mathbb{E}^{Q}_{\hat{\rho}}(\hat{X}|\hat{B}) \tag{4.15}$$

for any $\hat{\rho} \in D_{\hat{B}}$, $\hat{X} \in \mathcal{L}(\mathcal{H})$ and $f : \sigma(\hat{B}) \to \mathbb{C}$ if and only if $(Q, \tilde{Q})$ is the left Kirkwood-Dirac quasiprobability $(\mathrm{Q}^{\mathrm{KD},\ell}, \tilde{\mathrm{Q}}^{\mathrm{KD},\ell})$ *representation of quantum mechanics associated with $\hat{A}$ and $\hat{B}$. In* *this case, the Q-conditional expectation agrees with the conditional expectation given by equation* (3.7)*.*

Similarly the Q-conditional expectation satisfies

$$\mathbb{E}^Q_{\hat{\rho}}(\hat{X}f(\hat{B})|\hat{B}) = f(\hat{B})\mathbb{E}^Q_{\hat{\rho}}(\hat{X}|\hat{B}) \tag{4.16}$$

for any $\hat{\rho} \in D_{\hat{B}}$, $\hat{X} \in \mathcal{L}(\mathcal{H})$ and $f : \sigma(\hat{B}) \to \mathbb{C}$ if and only if $(Q, \tilde{Q})$ is the right Kirkwood-Dirac quasiprobability ($\mathrm{Q}^{\mathrm{KD,r}}$,$\tilde{\mathrm{Q}}^{\mathrm{KD,r}}$) *representation of quantum mechanics associated with $\hat{A}$ and $\hat{B}$. In this case the Q-conditional expectation agrees with the conditional expectation given by equation* (3.8)*.*

Proof. Once again, we do the "left" case, the "right" case being similar. We have shown in (4.13), that if $(Q, \tilde{Q}) = (\mathrm{Q}^{\mathrm{KD},\ell},\tilde{\mathrm{Q}}^{\mathrm{KD},\ell})$, then (4.15) is satisfied. Conversely, assume that S satisfies (4.15). Then, from Eq. 4.9 combined with Theorem 2.3, the Q-conditional expectation must agree with the conditional expectation given by equation (3.7), i.e. for any $\hat{\rho} \in D_{\hat{B}}$ and any $\hat{X} \in \mathcal{L}(\mathcal{H})$,

$$\sum_{a,b \in \sigma(\hat{A})\times\sigma(\hat{B})} \overline{\tilde{Q}_{a,b}(\hat{X}^\dagger)} Q_{a|b}(\hat{\rho})\Pi^{\hat{B}}_b = \sum_{b\in\sigma(\hat{B})} \frac{\langle \varphi^{\hat{B}}_b, \hat{X}\hat{\rho}\varphi^{\hat{B}}_b\rangle}{\langle \varphi^{\hat{B}}_b, \hat{\rho}\varphi^{\hat{B}}_b\rangle}\Pi^{\hat{B}}_b. \tag{4.17}$$

By taking the complex conjugate of equation (4.2), one sees that for any $(a, b), (a', b') \in \sigma(\hat{A}) \times \sigma(\hat{B})$,

$$\tilde{Q}_{a',b'}(S_{a,b}) = \delta_{a,a'}\delta_{b,b'}.$$

Let $(a, b) \in \sigma(\hat{A}) \times \sigma(\hat{B})$. Plugging $\hat{X} = S^\dagger_{a,b}$ into (4.17) yields

$$\frac{\mathrm{Tr}\Big(\hat{\rho}S^\dagger_{a,b}\Big)}{\langle \varphi^{\hat{B}}_b, \hat{\rho}\varphi^{\hat{B}}_b\rangle}\Pi^{\hat{B}}_b = \sum_{b'\in\sigma(\hat{B})} \frac{\mathrm{Tr}\Big(\hat{\rho}\Pi^{\hat{B}}_{b'}S^\dagger_{a,b}\Big)}{\langle \varphi^{\hat{B}}_{b'}, \hat{\rho}\varphi^{\hat{B}}_{b'}\rangle}\Pi^{\hat{B}}_{b'},$$

for any $\hat{\rho} \in D_{\hat{B}}$, from which follows

$$\delta_{b,b'}\,\mathrm{Tr}\Big(\hat{\rho}S^\dagger_{a,b}\Big) = \mathrm{Tr}\Big(\hat{\rho}\Pi^{\hat{B}}_{b'}S^\dagger_{a,b}\Big)$$

for any $\hat{\rho} \in D_{\hat{B}}$ and any $a \in \sigma(\hat{A})$, $b, b' \in \sigma(\hat{B})$. Since both sides of this equation are continuous in the variable $\hat{\rho}$, by Lemma 3.4 it must in fact hold for any $\hat{\rho} \in D(\mathcal{H})$. We deduce that

$$\delta_{b,b'}S^\dagger_{a,b} = \Pi^{\hat{B}}_{b'}S^\dagger_{a,b},$$

or equivalently

$$\delta_{b,b'}S^\dagger_{a,b'} = \Pi^{\hat{B}}_{b'}S^\dagger_{a,b}$$

for any $a \in \sigma(\hat{A})$ and $b, b' \in \sigma(\hat{B})$. Summing over the indices b yields

$$S^\dagger_{a,b'} = \Pi^{\hat{B}}_{b'}\Pi^{\hat{A}}_a$$

for any $(a, b') \in \sigma(\hat{A}) \times \sigma(\hat{B})$, and thus :

$$S_{a,b} = \Pi^{\hat{A}}_a\Pi^{\hat{B}}_b$$

for any $(a, b) \in \sigma(\hat{A}) \times \sigma(\hat{B})$, so $(Q, \tilde{Q})$ is indeed the left Kirkwood-Dirac representation ($\mathrm{Q}^{\mathrm{KD},\ell}$,$\tilde{\mathrm{Q}}^{\mathrm{KD},\ell}$) associated with $\hat{A}$ and $\hat{B}$. □

5 Conclusion and discussion

We have considered all Born-compatible quasiprobability representations associated to two incompatible observables $\hat{A}$ and $\hat{B}$. We have shown that among those, the "left" and "right" Kirkwood-Dirac representations are unique, in the sense that they are the only ones whose conditional expectations have a simple interpretation as best estimators. In particular, this distinguishes them from the Wigner representation, when it exists, such as for odd dimensional qudits.

We have limited ourselves to the simplest case of nondegenerate observables on finite dimensional Hilbert spaces. Generalizations of the notion of Born-compatible quasiprobability representations, in particular on infinite dimensional Hilbert spaces, or with positive operator-valued rather than projection-valued measures can easily be envisaged at the cost of a greater technical complexity.

Acknowledgments: This work was supported in part by the CNRS through the MITI interdisciplinary programs. S. De Bièvre, C. Langrenez and M. Spriet acknowledge the support of the CDP C2EMPI, as well as the French State under the France-2030 programme, the University of Lille, the Initiative of Excellence of the University of Lille, the European Metropolis of Lille for their funding and support of the R-CDP-24-004-C2EMPI project. SDB and CL thank David Arvidsson-Shukur for many stimulating and insightful discussions on the subject of this paper.

References

[1] G. Ludwig, *Foundations of Quantum Mechanics I.* Springer, 1983.

[2] P. Busch, "Quantum states and generalized observables: A simple proof of gleason's theorem," *Physical Review Letters*, vol. 91, no. 12, p. 120403, 2003.

[3] C. Ferrie, R. Morris, and J. Emerson, "Necessity of negativity in quantum theory," *Physical Review A*, vol. 82, no. 4, p. 044103, 2010.

[4] M. Lostaglio, A. Belenchia, A. Levy, S. Hernández-Gómez, N. Fabbri, and S. Gherardini, "Kirkwood-Dirac quasiprobability approach to the statistics of incompatible observables," *Quantum*, vol. 7, p. 1128, 2023.

[5] E. Wigner, "On the Quantum Correction For Thermodynamic Equilibrium," *Physical Review*, vol. 40, no. 5, pp. 749–759, 1932.

[6] H. Weyl, "Quantenmechanik und gruppentheorie," *Zeitschrift für Physik*, vol. 46, no. 1, pp. 1–46, 1927.

[7] J. G. Kirkwood, "Quantum statistics of almost classical assemblies," *Physical Review*, vol. 44, no. 1, p. 31, 1933.

[8] P. A. M. Dirac, "On the analogy between classical and quantum mechanics," *Rev. Mod. Phys.*, vol. 17, pp. 195–199, 1945.

[9] J. E. Moyal, "Quantum mechanics as a statistical theory," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 45, no. 1, pp. 99–124, 1949.

[10] K. Cahill and R. J. Glauber, "Ordered expansions in boson amplitude operators," *Physical Review*, vol. 177, no. 5, p. 1857, 1969.

[11] K. Cahill and R. J. Glauber, "Density operators and quasi-probability distributions," *Physical Review*, vol. 177, no. 5, p. 1882, 1969.

[12] D. Gross, "Hudson's theorem for finite-dimensional quantum systems," *Journal of Mathematical Physics*, vol. 47, no. 12, p. 122107, 2006.

[13] C. Bény, J. Crann, H. H. Lee, S.-J. Park, and S.-G. Youn, "Gaussian quantum information over general quantum kinematical systems i: Gaussian states," *Letters in Mathematical Physics*, vol. 115, no. 2, p. 32, 2025.

[14] S. De Bièvre, C. Langrenez, and D. Radchenko, "The Kirkwood-Dirac Representation Associated to the Fourier Transform for Finite Abelian Groups: Positivity," *Annales Henri Poincaré*, 2025.

[15] M. Spriet, "Characterizing the Kirkwood-Dirac positivity on second countable LCA groups," 2025. arXiv:2507.23628.

[16] H. G. Feichtinger and W. Kozek, "Quantization of TF lattice-invariant operators on elementary LCA groups," in *Gabor Analysis and Algorithms: Theory and Applications* (H. G. Feichtinger and T. Strohmer, eds.), pp. 233–266, Birkhäuser, 1998.

[17] F. Nicola and F. Riccardi, "The Hudson theorem in LCA groups and infinite quantum spin systems," 2025. arXiv:2507.13154.

[18] J. R. Klauder and B.-S. Skagerstam, *Coherent States: Applications in Physics and Mathematical Physics*. World Scientific, 1985.

[19] D. Robert and M. Combescure, *Coherent States and Applications in Mathematical Physics*. Theoretical and Mathematical Physics, Springer International Publishing, 2021.

[20] J.-P. Gazeau, *Coherent States in Quantum Physics*. Wiley-VCH Verlag GmbH & Co. KGaA, 2009.

[21] A. Perelomov, *Generalized Coherent States and Their Applications*. Springer, 1986.

[22] D. R. M. Arvidsson-Shukur, W. F. Braasch Jr, S. De Bièvre, J. Dressel, A. N. Jordan, C. Langrenez, M. Lostaglio, J. S. Lundeen, and N. Yunger Halpern, "Properties and applications of the Kirkwood–Dirac distribution," *New Journal of Physics*, vol. 26, p. 121201, 2024.

[23] A. M. Steinberg, "Conditional probabilities in quantum theory and the tunneling-time controversy," *Physical Review A*, vol. 52, no. 1, pp. 32–42, 1995.

[24] L. M. Johansen, "Quantum theory of successive projective measurements," *Physical Review A*, vol. 76, no. 1, p. 012119, 2007.

[25] J. Dressel, "Weak values as interference phenomena," *Physical Review A*, vol. 91, no. 3, p. 032116, 2015.

[26] R. Brummelhuis, "Conditional expectations in quantum mechanics and causal interpretations: the bohm momentum as a best predictor," 2024. arXiv:2411.08532.

[27] R. Brummelhuis, "Conditional expectations of quantum observables: I. Bohm momentum as best predictor of momentum given position," *Journal of Physics A: Mathematical and Theoretical*, to appear (2025).

[28] R. Brummelhuis, "Conditional expectations of quantum observables: II. A causal model for the Pauli equation ," *Journal of Physics A: Mathematical and Theoretical*, to appear (2025).

[29] S. De Bièvre, "Complete Incompatibility, Support Uncertainty, and Kirkwood-Dirac Nonclassicality," *Physical Review Letters*, vol. 127, no. 19, p. 190404, 2021.

[30] S. De Bièvre, "Relating incompatibility, noncommutativity, uncertainty, and Kirkwood–Dirac nonclassicality," *Journal of Mathematical Physics*, vol. 64, 02 2023. 022202.

[31] C. Ferrie, "Quasi-probability representations of quantum theory with applications to quantum information science," *Reports On Progress in Physics*, vol. 74, p. 116001, 2011.

[32] O. Christensen, *An Introduction to Frames and Riesz Bases.* Springer Science & Business Media, 2002.