

Quasi-probabilities in Conditioned Quantum Measurement and a Geometric/Statistical Interpretation of Aharonov's Weak Value

Jaeha Lee¹, Izumi Tsutsui²

Theory Center, Institute of Particle and Nuclear Studies, High Energy Accelerator Research Organization (KEK), 1-1 Oho, Tsukuba, Ibaraki 305-0801, Japan

¹*jlee@post.kek.jp*

²*izumi.tsutsui@kek.jp*

.....
 We show that the joint behaviour of an arbitrary pair of (generally non-commuting) quantum observables can be described by quasi-probabilities, which are an extended version of the standard probabilities used for describing the outcome of measurement for a single observable. The physical situations that require these quasi-probabilities arise when one considers quantum measurement of an observable conditioned by some other variable, with the notable example being the weak measurement employed to obtain Aharonov's weak value. Specifically, we present a general prescription for the construction of quasi-joint-probability (QJP) distributions associated with a given combination of observables. These QJP distributions are introduced in two complementary approaches: one from a bottom-up, strictly operational construction realised by examining the mathematical framework of the conditioned measurement scheme, and the other from a top-down viewpoint realised by applying the results of spectral theorem for normal operators and its Fourier transforms. It is then revealed that, for a pair of simultaneously measurable observables, the QJP distribution reduces to the unique standard joint-probability distribution of the pair, whereas for a non-commuting pair there exists an inherent indefiniteness in the choice of such QJP distributions, admitting a multitude of candidates that may equally be used for describing the joint behaviour of the pair. In the course of our argument, we find that the QJP distributions furnish the space of operators in the underlying Hilbert space with their characteristic geometric structures such that the orthogonal projections and inner products of observables can, respectively, be given statistical interpretations as 'conditionings' and 'correlations'. The weak value A_w for an observable A is then given a geometric/statistical interpretation as either the orthogonal projection of A onto the subspace generated by another observable B , or equivalently, as the conditioning of A given B with respect to the QJP distribution under consideration.

arXiv:1607.06406v1 [quant-ph] 21 Jul 2016

Contents		PAGE
1	Introduction	4
2	Unconditioned Measurement I: In Terms of Expectations	10
2.1	Reference Materials	10
2.1.1	A Crash-Course into Measure and Integration Theory	10
2.1.2	Rudimentary Techniques in handling the CCR	15
2.1.3	Tensor Product of Hilbert Spaces and Self-adjoint Operators	19
2.2	Unconditioned Measurement	21
2.3	Recovery of the Target Profile	25
2.3.1	Strong Unconditioned Measurement	26
2.3.2	Weak Unconditioned Measurement	26
2.3.3	Discussion	26
3	Unconditioned Measurement II: In Terms of Probabilities	28
3.1	Reference Materials	28
3.1.1	The Space of Complex Measures	28
3.1.2	The Space of Density Functions	30
3.1.3	Product Measures	33
3.1.4	Measure on Topological Spaces	34
3.1.5	Spectral Theorem and its Consequences	35
3.1.6	Observables admitting a Description by Density Functions	38
3.1.7	Simultaneously measurable Observables	39
3.2	Unconditioned Measurement	42
3.3	Recovery of the Target Profile	47
3.3.1	Strong Unconditioned Measurement	47
3.3.2	Weak Unconditioned Measurement	52
4	Conditioned Measurement I: In Terms of Conditional Expectations	59
4.1	Reference Materials	59
4.1.1	Conditioning	59
4.2	Conditioned Measurement	64
4.2.1	Topic: ‘Amplification Technique’ by Conditioning	64
4.2.2	Topic: ‘Limit of Amplification’ in Terms of Essential Suprema	65
4.3	Recovery of the Target Profile	70
4.3.1	Preliminary Observation	70
4.3.2	Conditional Quasi-expectations of Quantum Observables	74
4.3.3	Weak Conditioned Measurement	78
5	Conditioned Measurement II: In Terms of Conditional Probabilities	81
5.1	Reference Materials	83
5.1.1	Conditional Probabilities	83
5.1.2	Fourier Transformation	87
5.1.3	Wigner-Ville Distribution	88
5.2	Conditioned Measurement	89
5.2.1	Conditioning over Quasi-probabilities	89
5.2.2	Conditioned Measurement via the WV Distributions	93
5.3	Recovery of the Target Profile	97

5.3.1	Strong Conditioned Measurement	98
5.3.2	Weak Conditioned Measurement	99
5.4	Profile of the Target System	100
6	Quasi-probabilities of Quantum Observables	105
6.1	Reference Materials	105
6.1.1	Fourier Transform of Complex Measures	105
6.2	Quasi-joint-probabilities of a Combination of Quantum Observables	107
6.2.1	Preliminary Observations	107
6.2.2	Quasi-joint-probability Distributions	110
6.3	Complex-parametrised Sub-families	113
6.3.1	Additive Sub-family	113
6.3.2	Convolutive Sub-family	114
6.3.3	Qualification as Quasi-joint-probability Distributions	117
6.3.4	Relation to other known Proposals	118
6.4	Some General Properties	119
6.4.1	Complex Conjugate	119
6.4.2	Realness of the QJP Distributions	121
6.5	Conditioned Measurement Revisited	121
6.5.1	Short Introduction	121
6.5.2	Conditioned Measurement Revisited	122
7	Application: Interpretation of Aharonov's Weak Value	125
7.1	Reference Materials	125
7.1.1	L^2 -Theory of Conditional Expectations	125
7.1.2	Conditioning as Optimal Approximation	126
7.2	Statistical Interpretation of Geometric Structures	126
7.2.1	Preliminary Observation	127
7.2.2	Geometry induced by a specific Sub-family of QJSD	128
7.2.3	The Hilbert Space of Bounded Operators given a fixed State	130
7.3	Interpretation of Conditional Quasi-expectations	133
7.3.1	Geometric Interpretation of Conditional Quasi-expectations	134
7.3.2	Statistical Interpretation of Conditional Quasi-expectations	136
8	Summary and Discussion	138
8.1	Summary	138
8.1.1	QJP: Heuristic Construction	138
8.1.2	QJP: Formal Definition	139
8.1.3	Application	140
8.2	Discussion	140
A	Post-selected Measurement	145
A.1	Example: Analytic Model	145
A.2	Computation of the Gaussian Example	148

1. Introduction

Since the discovery of quantum mechanics in the beginning of the last century, our classical understanding of the concept of observables has undergone a drastic change. It is by now widely accepted that, in the microscopic world, measured values of a physical quantity, termed ‘observable’ in quantum mechanics, are intrinsically random, and that certain combinations of quantum observables do not admit coexistence, as exemplified typically by the pair of observables corresponding to the position and the momentum of a particle.

Such remarkable characteristics of quantum observables impose a strong limitation to the mathematical framework to be employed for describing their probabilistic behaviour; namely, it is no longer possible, in general, to assign probability spaces for the description of the joint behaviour of their arbitrary combinations in the classical sense. Nonetheless, various attempts have been made to construct a proper mathematical framework for the probabilistic description of the combination of quantum observables that resembles the Kolmogorovian style of formulation of classical probability theory. Extending the notion of probability has since been one of the major trends, which yielded the extended notion of probability which goes generally by the name of ‘quasi-probability’ or ‘pseudo-probability’. Among the most celebrated proposal is the Wigner-Ville (WV) distribution [1, 2], commonly known as the Wigner function in the physics community, which is primarily considered for a canonically conjugate pair of quantum observables to describe their joint behaviour. Another, though less known, example is the Kirkwood-Dirac (KD) distribution [3, 4], which is structured differently but is meant to serve a similar purpose for arbitrary pairs.

Historically, those proposals including the WV and KD distributions have been made more or less in a heuristic manner, and as such, the general mathematical framework for the study, including the prescription for the concrete construction of such distributions to a pair of arbitrary quantum observables, which may comprehensively be termed ‘quasi-joint-probability’ (QJP) distributions, is still underdeveloped, not to mention a transparent overview of the relations among the QJPs. We know, for instance, that both the WV and KD distributions retain similar properties to the standard joint-probability distributions defined for a pair of classical random variables, but they exhibit their own outstanding queerness in that the former admits negative numbers to be assigned whereas the latter takes even complex numbers. However, we still do not know whether the peculiar properties of joint-probability including those of the WV and KD distributions, which have occasionally been considered a serious impediment to their physical interpretation, are a norm of QJP distributions, or there can be other types of examples which share classical properties of joint-probability in different aspects. The theme of this paper revolves around the concept of QJP distributions of quantum observables, with the first objective being to present a mathematically solid framework to address some of their problems in a more systematic and lucid manner.

Another motivation of this paper comes from the recent rise of interest in the novel quantum observable called the *weak value*, which has been put forward by Aharonov and co-workers [5] based on their time-symmetric formulation of quantum mechanics [6] proposed more than a half century ago. In simple terms, the weak value

$$A_w := \frac{\langle \psi', A\psi \rangle}{\langle \psi', \psi \rangle} \quad (1.1)$$

is a physical quantity that supposedly characterises the value of the observable A in the process specified by an initial state $|\psi\rangle$ and a final state $|\psi'\rangle$ both specified in advance. Unlike the standard physical value which is given by one of the eigenvalues of an observable A , the weak value admits a definite value for any A , and is envisaged to be meaningful even for a set of non-commutable observables simultaneously.

This inspired a new insight for analysing the quantum nature of the system as well as for understanding various counter-intuitive phenomena in quantum mechanics based on the weak value. For instance, the complex-valued nature of A_w allows for a direct measurement of the wave function, offering a novel technique to rival the existing technology of quantum tomography. This in turn alludes us to contemplate on the possible trajectory of a particle [7, 8], a notion which has conventionally been deemed untenable due to the incompatibility of measuring the position and the momentum simultaneously. The weak value also admits novel physical interpretations on such fundamental aspects of quantum mechanics as the wave-particle duality and the local existence of the physical quantity itself, offering us a possible resolution to some of the quantum paradoxes, including the three-box paradox [9], Hardy's paradox [10] and the Cheshire cat paradox [11].

Despite its growing attention, the status of the weak value in quantum mechanics is still not solid, and especially its physical interpretation is still open to debate. One of the recent strategies in addressing this question has been to investigate its relations to quasi-probabilities, specifically those to the KD distribution [12, 13]. In this paper, we shall follow this line of study and show, among others, that a novel geometric/statistical interpretation emerges from these distributions. This necessitates a sound mathematical basis of QJP distributions, which we will provide in the course of our discussions.

The main theme of this paper is thus to obtain a more coherent understanding of the formalism of QJP distributions of quantum observables, and subsequently to apply the results in some areas of the foundational problems of quantum mechanics. In view of this, the key problems regarding QJP distributions may be to

- (i) provide a reasonably solid mathematical framework for the study of QJP distributions based on measure and integration theory, and possibly on the theory of generalised functions,
- (ii) present a viable scheme to address the inherent indefiniteness/arbitrariness to the possible candidates for QJP distributions of non-commuting pairs of quantum observables, a methodical way for their constructions, and the relation between each of the candidates, and
- (iii) devise a procedure for measuring such various candidates of QJP distributions in a systematic manner.

We shall address these problems from two complementary approaches: one from a bottom-up, strictly operational construction realised by carefully reviewing the mathematical description of the conditioned measurement scheme, and the other from a top-down viewpoint realised by applying the results of spectral theorem for normal operators and its Fourier transforms.

The results of the study shall be subsequently applied to the analysis for the physical interpretation of the weak value. To this end, we first concentrate on the L^2 structures which the QJP distributions naturally induce, and observe that they furnish a statistical interpretation of the geometric structures introduced on the space of observables in the underlying

Hilbert space, analogously to those introduced in the space of random variables in classical probability theory. Geometric concepts such as orthogonal projections and inner products are accordingly endowed with statistical interpretations as ‘conditionings’ and ‘correlations’, respectively, and in addition the representation of linear operators by functions provides us with a convenient tool for evaluating statistical quantities involved. These observations form a basis to perform further study on the weak value in general. As a result, the weak value A_w is given a geometric/statistical interpretation: either as the orthogonal projection of an observable A on the subspace generated by another observable B which is determined by one of the predetermined states entering in the weak value, or equivalently, as the conditioning of A given B with respect to the QJP distribution under consideration. Although we shall not discuss it here, we mention that this interpretation also leads to a set of novel and remarkable inequalities of uncertainty relations for approximation/estimation which are capable of treating both the standard position-momentum inequality and the time-energy inequality [14].

As for the practical outcomes of our argument laid out for QJP distributions, we mentioned earlier the systematic construction of QJP distributions and the geometric/statistical interpretation of the weak value, but each of these can be made more explicit as follows. First, for the systematic construction of QJP distributions, we furnish a general prescription which ensures that it can describe the joint behaviour of an arbitrary pair of quantum observables. Specifically, inspired by the observations made on the Fourier transform of the product spectral measure of two simultaneously measurable observables A and B , we introduce a mixture $\#(s, t)$ of the disintegrated components of e^{-isA} and e^{-itB} with real parameters s, t for arbitrary pairs of (generally non-commuting) observables A and B , and thereby define the QJP distribution of the pair by the inverse Fourier transform of the distribution $(s, t) \mapsto \langle \psi, \#(s, t)\psi \rangle / \|\psi\|^2$ to a given quantum state $|\psi\rangle$. Each of the QJP distributions is then found to possess reasonable properties to be qualified as what its name suggests to be, and one can confirm that both the WV distribution and the KD distribution do belong to this class. The inherent arbitrariness observed to the candidates for QJP distributions is then understood as the possible variety of the way one could mix the disintegrated components of the unitary operators, which originates directly from the non-commutative nature of the pair of the observables A and B . A concrete measurement scheme for members of a specific subfamily of QJP distributions is further proposed.

For the geometric/statistical interpretation of the weak value, on the other hand, we start by noting that, as distributions, each QJP distribution naturally induces an L^2 structure. We will then find that the QJP distributions provide convenient methods of representing geometric structures in terms of the inner products of the form

$$\langle\langle B, A \rangle\rangle_{\psi, \alpha} := \frac{1 + \alpha}{2} \cdot \frac{\langle B\psi, A\psi \rangle}{\|\psi\|^2} + \frac{1 - \alpha}{2} \cdot \frac{\langle A\psi, B\psi \rangle}{\|\psi\|^2}, \quad -1 \leq \alpha \leq 1, \quad (1.2)$$

which can be introduced on the space of operators in the underlying Hilbert space by integration of functions. With this inner product, we are allowed to consider orthogonal projections onto the subspaces $\mathfrak{E}_\psi(B)$ of operators generated by self-adjoint operators B , and find that the orthogonal projections can be interpreted as conditioning given B with respect to the

QJP distributions under consideration. The projection

$$P_\alpha(A|B; \psi) = \int_{\mathbb{R}} \left(\frac{1 + \alpha}{2} \cdot \frac{\langle b, A\psi \rangle}{\langle b, \psi \rangle} + \frac{1 - \alpha}{2} \cdot \frac{\langle \psi, Ab \rangle}{\langle \psi, b \rangle} \right) dE_B(b) \quad (1.3)$$

of the observable A on the subspace $\mathfrak{E}_\psi(B)$ is further found to be described by the weak value¹, providing us with its proper geometric/statistical interpretation (Proposition 7.7).

Having furnished a general introduction to the topic of QJP distributions of quantum observables and the weak value along with a brief summary of the content, we now give the outline of the present paper. After this introductory section, we organize the main body, Section 2 to Section 7, into the following three logical groups of mutually interrelated topics:

- (A) **QJP: Heuristic Construction** Four sections starting from Section 2 to 5 are devoted to a heuristic and bottom-up construction of QJP distributions of a pair of quantum observables. This is accomplished by a thorough analysis on the mathematical formalism of two measurement schemes. One is the standard scheme, which we call the ‘unconditioned measurement (UM) scheme’, in which we measure an observable A under a given state as conventionally done (Section 2 to 3). The other is what we call the ‘conditioned measurement (CM) scheme’, in which under a given state we measure an observable A along with another observable B whose outcome is used for conditioning (Section 4 to 5). Each of these analyses will be conducted on the level of (conditional) expectations and (conditional) probabilities.
- (i) **UM I** We start by reviewing, in Section 2, the UM scheme by a standard operator-centric approach, and investigate how one could reclaim the information of the target system by that means.
 - (ii) **UM II** Subsequently, in Section 3, we take a closer look on the UM scheme in the level of probabilities, where the quantity of interest is now not only the statistical average, but also the ‘raw’ probability measure describing the probabilistic behaviour of the measurement outcomes of the meter observable, and discuss how one could recover the probability measure describing the outcomes of the target observable.
 - (iii) **CM I** From Section 4 onward, we turn our attention to the CM scheme. In Section 4, we first conduct, in a parallel manner as we have done in the preceding Section 2, an analysis in the operator level, where now the quantity of interest becomes the conditional expectation of the meter observable given another conditioning observable B of the target system.
 - (iv) **CM II** In Section 5, the study of the CM scheme is given a probabilistic approach, where the quantity of interest is the Wigner-Ville distribution of a pair of canonically conjugate observables on the meter system conditioned by the outcome of the conditioning observable B of the target system. We then see that this implies the existence of the concept of QJP distributions of pairs of generally non-commuting observables.

¹ Here, E_B denotes the unique spectral measure associated to the self-adjoint operator B (more on this in Section 3.1). Intuitively, $E_B(b) = |b\rangle\langle b|$ is the projection associated with each of the eigenvalues b of B , and thus $dE_B(b) = |b\rangle\langle b|db$ in a laxer expression.

- (B) **QJP: Formal Definition** Inspired by the heuristic arguments employed in the operational analyses over the preceding four sections, we devote Section 6 to the top-down construction of QJP distributions for arbitrary pairs of generally non-commuting quantum observables. We shall then summarise our findings obtained through Section 2 to Section 5 from a rather aerial viewpoint, discussing where the heuristic arguments and observations in the preceding sections find their places in this relatively general framework.
- (C) **Application to the Interpretation of Weak Values** As an application of the mathematical formalism provided so far, in Section 7 we conduct a study on the quantum analogue of correlations, which can be defined even for a pair of non-commuting observables. This leads us to the aforementioned geometric/statistical interpretation of the weak value as conditional quasi-expectations.

We shall finally summarise our results and give some concluding remarks in the last Section 8.

Prior to our main discussions, however, we wish to say a few words about the mathematical preliminaries we supposed for the readers in preparing this paper. The formalism that we intend to provide necessarily requires, on top of the mandatory functional analysis, moderate acquaintance to measure and integration theory, preferably some familiarity with the basic terminologies in general topology, and ideally insight into the basic ideas of the theory of generalised functions. The obvious difficulty is then to find a decent balance between rigour and generality on one side, and accessibility on the other. To achieve this balance as much as possible, and assure our entire arguments to be fully accessible without any prior knowledge of advanced mathematics, we have included at the beginning of each section a subsection entitled Reference Materials containing a rather lengthy introduction of mathematical concepts that are used in the subsequent discussions. While the authors took care in introducing these mathematical concepts and their results in a self-contained manner to respect their logical sequence, these Reference Materials are primarily intended to serve as a convenient place to summarise the basic concepts and results in a crash-course, and as such, the mathematical theories presented there are not intended to be learned from scratch. For those who are interested in the mathematics itself are advised to be referred to standard textbooks on the respective topics, *e.g.*, for general topology [15–17], measure and integration theory [18–23], functional analysis [24, 25], and also those specifically targeting the audience from the physics community [26–29]. Naturally, those who are already familiar with the preparatory materials may safely skip them and directly go to the main arguments that follow.

Admittedly, the style of discussion found in this paper is heavily oriented toward mathematical rigorousness and logical clarity rather than brevity and physical intuition, especially compared to those found in the majority of the literature in physics. However, in spite of the possible initial hesitation that may be expected for the general readers due to the unfamiliarity of the style, the authors decided to adopt it in the belief that this way of presentation has its own merit, and that the costs will outweigh the rewards in the end. In fact, several important concepts and results from the branches of mathematics mentioned above (specifically, measure and integration theory and functional analysis) are quite indispensable in understanding some of the interesting results obtained in this paper. This is so, for instance, in defining the conditional quasi-expectations (to which Aharonov’s weak value belongs as

a special case) in terms of the Radon-Nikodým derivative to understand their properties (Section 4.3.2), in formulating the problem of the ‘limit of amplification’ by conditioning in terms of essential suprema (Section 4.2.1), in defining a family of QJP distributions of a combination of generally non-commuting quantum observables by the method of hashing (Section 6), and in providing geometric and ‘statistical’ interpretation of conditional quasi-expectations (Section 7). The authors hope that the readers will not be discouraged by these mathematical materials, but rather enjoy them to go through the discussions and reach the fruit of the physical results they finally brings forth.

Mathematical Notations Employed. Throughout this paper, we denote by \mathbb{K} either the real field \mathbb{R} or the complex field \mathbb{C} , and define $\mathbb{K}^\times := \mathbb{K} \setminus \{0\}$. In order to avoid confusion, we denote the collection of all natural numbers including 0 by \mathbb{N}_0 , and $\mathbb{N}^\times := \mathbb{N}_0 \setminus \{0\}$. Since our primary interest is on quantum mechanics, Hilbert spaces are always assumed to be complex. Conforming to the convention in physical literature, we denote the complex conjugate of a complex number $c \in \mathbb{C}$ by c^* , and an inner product $\langle \cdot, \cdot \rangle$ defined on a complex linear space is anti-linear in its first argument and linear in the second. For simplicity, we adopt the natural units where we specifically have $\hbar = 1$, unless stated otherwise.

2. Unconditioned Measurement I: In Terms of Expectations

We start by providing a brief review on the archetype of the indirect measurement scheme widely known as the von Neumann measurement scheme. The scheme will be referred to as the *unconditioned measurement* (UM) scheme in generic terms throughout this paper, primarily in order to contrast it with the *conditioned measurement* (CM) scheme (which includes the *post-selected measurement* scheme as a special case) discussed later.

2.1. Reference Materials

As a preamble to this section, we here include three introductory topics that form the basis of our study. We start by collecting some of the basic terminologies and results of measure and integration theory, based on which modern probability theory was established by Kolmogorov *et al.* Subsequently, we provide a brief note on both the *Schrödinger representation* and the *Weyl representation* of the canonical commutation relations (CCR), which will be extensively employed in describing the meter system in our measurement scheme. We finally close this subsection by providing a short summary on the precise definition of tensor products of Hilbert spaces and that of self-adjoint operators. Since these materials are included just to make our presentation self-contained, those who are already familiar with the subject may safely skip the contents and proceed directly to Section 2.2.

2.1.1. A Crash-Course into Measure and Integration Theory. We begin by presenting some of the most basic concepts and results of measure and integration theory, starting from the definition of measure spaces up to the construction of the Lebesgue integration, followed by the definition of L^p spaces.

σ -algebras and Measurable Spaces. Let X be any set, and let $\mathfrak{P}(X)$ denote the power set² of X , *i.e.*, the collection of all subsets of X . A family $\mathfrak{A} \subset \mathfrak{P}(X)$ of subsets of X is called a *σ -algebra* over X , if it satisfies the following conditions:

- (i) $X \in \mathfrak{A}$.
- (ii) $A \in \mathfrak{A}$ implies $A^c := X \setminus A \in \mathfrak{A}$.
- (iii) For any sequence $(A_n)_{n \geq 1}$ of subsets of X , $\bigcup_{n=1}^{\infty} A_n \in \mathfrak{A}$ holds.

Given a σ -algebra \mathfrak{A} over X , each element $A \in \mathfrak{A}$ is called a *measurable set*, and the ordered pair (X, \mathfrak{A}) is called a *measurable space*.

Generator of a σ -algebra. A trivial, but important property of σ -algebras is that, for any collection $(\mathfrak{A}_i)_{i \in I}$ of σ -algebras over X indexed by an index set I , the intersection $\bigcap_{i \in I} \mathfrak{A}_i = \{A \in \mathfrak{P}(X) : A \in \mathfrak{A}_i, \forall i \in I\}$ is itself a σ -algebra over X . This leads to the following basic fact: For any collection $\mathfrak{E} \subset \mathfrak{P}(X)$ of subsets of X , there exists a smallest (with respect to the set inclusion) σ -algebra encompassing \mathfrak{E} , namely, the intersection of all σ -algebras that encompass \mathfrak{E} . The intersection is called the *σ -algebra generated by \mathfrak{E}* , denoted as $\sigma(\mathfrak{E})$, and \mathfrak{E} is in turn called the *generator* of $\sigma(\mathfrak{E})$.

²The symbol \mathfrak{A} is the capital letter of the Fraktur typeface of ‘A’ as in ‘Algebra’, \mathfrak{B} for ‘B’ as in ‘Borel’, \mathfrak{E} for ‘E’ as in ‘Erzeuger (generator)’, \mathfrak{O} for ‘O’ as in ‘offen (open)’ and \mathfrak{P} for ‘P’ as in ‘Potenz (power)’ (some of them introduced shortly after).

Borel σ -algebras. Let X be a metric (or, in general, a topological) space, and let \mathfrak{D} denote the collection of all open sets of X . We call the σ -algebra generated by \mathfrak{D} , the *Borel σ -algebra of X* , and denote it by $\mathfrak{B}(X) := \sigma(\mathfrak{D})$. We prepare a special symbol for the special case $X = \mathbb{R}^n$ ($n \in \mathbb{N}^\times$), in which we denote the Borel σ -algebra of \mathbb{R}^n by $\mathfrak{B}^n := \mathfrak{B}(\mathbb{R}^n)$, which is among the most well-known examples of σ -algebras that, incidentally, also plays an important role in quantum theory. For simplicity, we occasionally denote $\mathfrak{B} := \mathfrak{B}^1$ whenever there is no risk of confusion.

Measures and Measure Spaces. Let (X, \mathfrak{A}) be a measurable space. A map $\mu : \mathfrak{A} \rightarrow \overline{\mathbb{R}}$ from the σ -algebra \mathfrak{A} to the *extended real line* $\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$ is called a *measure*, if μ satisfies the following conditions:

- (i) $\mu(\emptyset) = 0$.
- (ii) $\mu \geq 0$.
- (iii) For any sequence $(A_n)_{n \geq 1}$ of pairwise disjoint subsets of X , the *countable additivity*

$$\mu \left(\bigcup_{n=1}^{\infty} A_n \right) = \sum_{n=1}^{\infty} \mu(A_n) \quad (2.1)$$

holds.

Given a measure μ over a measurable space (X, \mathfrak{A}) , the ordered triple (X, \mathfrak{A}, μ) is called a *measure space*.

Lebesgue-Borel Measure. As a concrete example, we make notes on the n -dimensional *Lebesgue-Borel measure* β^n ($n \in \mathbb{N}^\times$) defined on the measurable space $(\mathbb{R}^n, \mathfrak{B}^n)$, which is among the most well-known and important examples of measure spaces. To this end, we first recall that a measure μ on $(\mathbb{R}^n, \mathfrak{B}^n)$ is called *translation invariant*, if

$$\mu(B + a) = \mu(B), \quad B \in \mathfrak{B}^n, \quad (2.2)$$

holds for any $a \in \mathbb{R}^n$, where $B + a := \{x + a : x \in B\}$. The Lebesgue-Borel measure β^n is then specified as the unique translation invariant measure on $(\mathbb{R}^n, \mathfrak{B}^n)$ that satisfies the normalisation condition $\beta^n([0, 1]^n) = 1$, where

$$[0, 1]^n := \{x \in \mathbb{R}^n : 0 < x_i \leq 1 \text{ for } 1 \leq i \leq n, x_i \text{ is the } i\text{th coordinate of } x\}. \quad (2.3)$$

This is the measure which is implicitly assumed for the most case in performing the usual integration by the symbol

$$\int_{-\infty}^{\infty} f(x) dx := \int_{\mathbb{R}} f(x) d\beta(x), \quad (2.4)$$

which is a common practice in the physics community (the precise definition of the integral on the r. h. s. will be presented shortly after). The proof of the existence and uniqueness of the Lebesgue-Borel measure will be found in most elementary textbooks on the topic.

Measurable Functions. Let (X, \mathfrak{A}) and (X', \mathfrak{A}') be measurable spaces. A map $f : X \rightarrow X'$ is called *\mathfrak{A} - \mathfrak{A}' measurable* (or just *measurable* for short, whenever the measure spaces concerned are obvious by context), if $f^{-1}(\mathfrak{A}') \subset \mathfrak{A}$ holds. In particular, we call a map $f : X \rightarrow X'$ from a metric (or a topological) space X to another metric (or a topological) space

X' *Borel-measurable* if it is $\mathfrak{B}(X)$ - $\mathfrak{B}(Y)$ measurable. An important fact to note is that a continuous map $f : X \rightarrow X'$ is necessarily Borel-measurable.

Numerical Functions. In integration theory, it proves fruitful to consider not only real functions $f : X \rightarrow \mathbb{R}$, but also functions that take values in the extended real line $\overline{\mathbb{R}}$, which is called a *numerical function*. One naturally equips $\overline{\mathbb{R}}$ with the ordering $-\infty < a < +\infty$, $a \in \mathbb{R}$, and may also define agreeable operations of addition, subtraction and multiplication, where most of them should be self-evident, except for the following rather arbitrary definition

$$0 \cdot (\pm\infty) := (\pm\infty) \cdot 0 := 0, \quad \infty - \infty := -\infty + \infty := 0. \quad (2.5)$$

We then define the σ -algebra on $\overline{\mathbb{R}}$ by

$$\overline{\mathfrak{B}} := \{B \cup E : B \in \mathfrak{B}, E \subset \{-\infty, +\infty\}\}, \quad (2.6)$$

where, in particular, its restriction on the real line gives $\overline{\mathfrak{B}}|_{\mathbb{R}} = \mathfrak{B}$. We then say that a numerical function $f : (X, \mathfrak{A}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathfrak{B}})$ is measurable, if it is \mathfrak{A} - $\overline{\mathfrak{B}}$ measurable. Throughout this paper, we denote by $\mathcal{M}^+(\mathfrak{A})$ (or occasionally by \mathcal{M}^+ , whenever the σ -algebra concerned is evident by context) the collection of all measurable non-negative numerical functions.

Lebesgue Integration. In introducing the concept of integration, we proceed in three steps: We first define the integration for non-negative step functions, then extend the treatment to functions belonging to \mathcal{M}^+ , and finally discuss the integrability of measurable numerical or complex functions.

- (i) *Integration of Step Functions.* Let (X, \mathfrak{A}, μ) be a measure space. A measurable function $f : (X, \mathfrak{A}) \rightarrow (\mathbb{R}, \mathfrak{B})$ is called a *step function* (staircase function, simple function), if it takes only finite distinct values in \mathbb{R} . The collection of all measurable non-negative step functions will be denoted by \mathcal{T}^+ . One readily sees that a non-negative step function $f \in \mathcal{T}^+$ admits an expression

$$f = \sum_{k=1}^m a_k \chi_{A_k}, \quad (2.7)$$

where $a_1, \dots, a_m \geq 0$ are non-negative real numbers, $A_1, \dots, A_m \in \mathfrak{A}$ are measurable sets, and χ_A denotes the characteristic function

$$\chi_A(x) = \begin{cases} 1, & x \in A, \\ 0, & x \notin A, \end{cases} \quad (2.8)$$

of the subset $A \subset X$. We then define the (μ -)integral of f (over X) as

$$\int_X f \, d\mu := \sum_{k=1}^m a_k \mu(A_k), \quad (2.9)$$

whose value lies in $[0, \infty]$. Note that, although the expression (2.7) is non-unique due to the possible choice of the measurable sets used, the definition (2.9) is well-defined since the outcome of the integral is independent of the choice.

(ii) *Integration of Functions in \mathcal{M}^+ .* Now that we have defined the Lebesgue integral of non-negative step functions, we next define the integral of non-negative measurable numerical functions. For $f \in \mathcal{M}^+$, the Lebesgue integral of f is defined as

$$\int_X f \, d\mu := \sup \left\{ \int_X s \, d\mu : 0 \leq s \leq f, s \in \mathcal{T}^+ \right\}. \quad (2.10)$$

The above definition (2.10) is consistent with that for step functions (2.9) introduced earlier, for one readily checks that the integral coincides for $f \in \mathcal{T}^+ \subset \mathcal{M}^+$.

Before we move on to the final step, we introduce some useful notations. We let \mathbb{K} denote either the real field \mathbb{R} or the complex field \mathbb{C} , and we understand them to be respectively equipped with the Borel σ -algebra \mathfrak{B} or \mathfrak{B}^2 . Analogously, we let

$$\hat{\mathbb{K}} := \overline{\mathbb{R}} \text{ or } \mathbb{C}, \text{ respectively equipped with the } \sigma\text{-algebra } \hat{\mathfrak{B}} := \overline{\mathfrak{B}} \text{ or } \mathfrak{B}^2$$

for later convenience. For a numerical function $f : X \rightarrow \overline{\mathbb{R}}$, we define its positive and negative parts as

$$f^\pm(x) := \max(\pm f(x), 0). \quad (2.11)$$

One then sees that a function $f : X \rightarrow \hat{\mathbb{K}}$ is measurable if and only if all the positive and negative parts of both the real and imaginary parts $(\operatorname{Re}f)^\pm, (\operatorname{Im}f)^\pm$ of f are measurable. Given the necessary preparations, we finally obtain the following definition:

Definition (Lebesgue Integral). *Under the assumptions above, a function $f : X \rightarrow \hat{\mathbb{K}}$ is called μ -integrable (or simply integrable) over X if f is measurable, and all the four integrals*

$$\int_X (\operatorname{Re}f)^\pm \, d\mu, \quad \int_X (\operatorname{Im}f)^\pm \, d\mu \quad (2.12)$$

are finite. The value

$$\int_X f \, d\mu := \int_X (\operatorname{Re}f)^+ \, d\mu - \int_X (\operatorname{Re}f)^- \, d\mu + i \int_X (\operatorname{Im}f)^+ \, d\mu - i \int_X (\operatorname{Im}f)^- \, d\mu \quad (2.13)$$

is then called the (μ) -integrable of f (over X) or the Lebesgue integral of f (over X with respect to μ).

By definition, linearity

$$\int_X (af(x) + bg(x)) \, d\mu(x) = a \int_X f(x) \, d\mu(x) + b \int_X g(x) \, d\mu(x), \quad (2.14)$$

of the integration naturally follows as expected. For a measurable set $A \in \mathfrak{A}$, the use of the shorthand

$$\int_A f \, d\mu := \int_X \chi_A \cdot f \, d\mu \quad (2.15)$$

is common, where χ_A is the characteristic function of the measurable set.

Probability Spaces and Expectation Values. A measure space (X, \mathfrak{A}, μ) is called a *probability space*, if the measure is normalised by unity $\mu(X) = 1$. Given a probability space (X, \mathfrak{A}, μ) and a μ -integrable function f , the total integration of f is occasionally denoted by

$$\mathbb{E}[f; \mu] := \int_X f \, d\mu, \quad (2.16)$$

and called the *expectation value* of f under μ .

Dominated Convergence Theorem. The advantage of the Lebesgue integration (over the familiar Riemann counterpart) especially manifests itself when dealing with convergence. For later use throughout this paper, we make a note of one of the most powerful and oft-used theorems regarding the interchange of limit and integration. To this end, we first furnish some terminologies. Let (X, \mathfrak{A}, μ) be a measure space, and let a statement E be defined on each element $x \in X$. We say that the statement E holds $(\mu-)$ *almost everywhere* (abbreviation: (μ) -a.e.), if there exists a measurable set $N \in \mathfrak{A}$ with $\mu(N) = 0$ such that the statement E holds for $x \in X \setminus N$.

Theorem (Dominated Convergence Theorem). *Let (X, \mathfrak{A}, μ) be a measure space, and let $f, f_n : X \rightarrow \hat{\mathbb{K}}$ ($n \in \mathbb{N}^\times$) be measurable. If the sequence of the functions converge point-wise $\lim_{n \rightarrow \infty} f_n = f$ μ -a.e., and if moreover there exists an μ -integrable function $g \in \mathcal{M}^+$ such that $|f_n| \leq g$ holds μ -a.e. for all $n \in \mathbb{N}^\times$, then*

$$\lim_{n \rightarrow \infty} \int_X |f_n - f| d\mu = 0 \quad (2.17)$$

holds, which in particular implies

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu. \quad (2.18)$$

L^p Spaces. Having provided the definition of the Lebesgue integration, we close this subsection by introducing an important class of function spaces: L^p . Let $\mathcal{L}^p(\mu)$, $1 \leq p < \infty$, denote the space of all measurable functions $f : X \rightarrow \mathbb{K}$ for which its L^p -norm

$$\|f\|_p := \left(\int_X |f|^p d\mu \right)^{1/p} \quad (2.19)$$

is finite. For $p = \infty$, we let $\mathcal{L}^\infty(\mu)$ denote the space of all f for which its *essential supremum*

$$\|f\|_\infty := \inf\{\lambda \in [0, \infty] : |f| \leq \lambda \text{ } \mu\text{-a.e.}\} \quad (2.20)$$

is finite (such a function is called *essentially bounded*). The term essential supremum is justified by the fact that the evaluation $|f| \leq \|f\|_\infty$ μ -a.e. universally holds (to see this, observe that if $\|f\|_\infty < \infty$ is given, $\{x : |f| > \|f\|_\infty\} = \bigcup_{n=1}^\infty \{|f| > \|f\|_\infty + 1/n\}$ is a set of measure zero). Now, by identifying two functions $f, g \in \mathcal{L}^p(\mu)$ by the equivalence relation $f \sim g \Leftrightarrow f = g$ μ -a.e., we obtain a quotient space $L^p(\mu) := \mathcal{L}^p(\mu) / \sim$. For simplicity, it is customary to denote an element of $L^p(\mu)$ by its representative $f \in \mathcal{L}^p(\mu)$ whenever there is no risk of confusion. For $f \in L^p(\mu)$, one finds that the quantity $\|f\|_p$, $1 \leq p \leq \infty$ is well-defined (irrespective of the choice of the representative), and that this in fact provides a norm on $L^p(\mu)$, called the L^p -norm. The norm $\|\cdot\|_p$ is also known to be complete and hence makes $L^p(\mu)$ into a Banach space. The case $p = 2$ is of particular interest in the context of quantum mechanics, where the integration,

$$\langle g, f \rangle := \int g^* f d\mu, \quad (2.21)$$

defines an inner product that satisfies $\langle f, f \rangle = \|f\|_2^2$, making $L^2(\mu)$ into a Hilbert space. As a special case, we are mostly interested in the choice $(\mathbb{R}^n, \mathfrak{B}^n, \beta^n)$ of the measure space. Conforming to convention in physical literature, we prepare a special symbol for the L^p spaces of it and denote $L^p(\mathbb{R}^n) := L^p(\beta^n)$.

Hölder's Inequality. Among the most important inequality regarding L^p -spaces is the Hölder's inequality.

Theorem 2.1 (Hölder's Inequality). *Let $1 \leq p, q \leq \infty$, $\frac{1}{p} + \frac{1}{q} = 1$, where we understand $1/\infty := 0$, and let $f, g : X \rightarrow \hat{\mathbb{K}}$ be measurable. Then,*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q \quad (2.22)$$

holds.

For the specific choice $p, q = 2$, the resulting inequality has its own name as the Cauchy-Schwarz Inequality.

2.1.2. Rudimentary Techniques in handling the CCR. While the contents of the following topics are widely known, we include this material mainly for reader's convenience, and also for self-consistency and reference.

Schrödinger Representation of the CCR. We start by recalling the definition of the Schwartz space. A function $f : \mathbb{R}^n \rightarrow \mathbb{K}$ is called *rapidly decreasing* when

$$\lim_{|x| \rightarrow \infty} x^\gamma f(x) = 0 \quad (2.23)$$

holds for any $\gamma := (\gamma_1, \dots, \gamma_n)$ with $\gamma \in \mathbb{N}_0^n$. Here, the multi-index symbol $\gamma \in \mathbb{N}_0^n$ is understood to be used as

$$x^\gamma := x_1^{\gamma_1} \cdots x_n^{\gamma_n}, \quad D^\gamma := (D_1)^{\gamma_1} \cdots (D_n)^{\gamma_n}, \quad (2.24)$$

where $D_i := \partial/\partial x_i$ is the partial differentiation operator with respect to the variable x_i . The space

$$\mathcal{S}(\mathbb{R}^n) := \{f \in C^\infty(\mathbb{R}^n) : D^\gamma f \text{ is rapidly decreasing, } \gamma \in \mathbb{N}_0^n\}, \quad (2.25)$$

is then called the *Schwartz space*, and its elements are in turn called *Schwartz functions*. The Schwartz space is known to be a dense subspace $\mathcal{S}(\mathbb{R}^n) \subset L^p(\mathbb{R}^n)$ for $1 \leq p < \infty$. A well-known example of Schwartz functions is provided by the form,

$$x^\gamma e^{-a|x|^2} \in \mathcal{S}(\mathbb{R}^n), \quad \gamma \in \mathbb{N}_0^n, \quad a > 0. \quad (2.26)$$

Specifically, the Gaussian wave-functions, which also appear later in our analysis, are among the most oft-used members of the Schwartz space belonging to this class.

Now that we have the necessary definitions, we return to the main topic of this subsection and, for simplicity, confine ourselves to the case $n = 1$ without loss of generality. We start by introducing a pair of important operators \hat{x} and \hat{p} on the Hilbert space $L^2(\mathbb{R})$. Among these, $\hat{x} : \text{dom}(\hat{x}) \rightarrow L^2(\mathbb{R})$ is an operator on $L^2(\mathbb{R})$ defined by the multiplication of x on a function f ,

$$\hat{x} : f(x) \mapsto xf(x), \quad (2.27)$$

with its domain,

$$\text{dom}(\hat{x}) := \{f \in L^2(\mathbb{R}) : xf \in L^2(\mathbb{R})\}. \quad (2.28)$$

The operator \hat{x} is known to be self-adjoint and is called the (one-dimensional) *position operator*.

Next, consider the operator $-iD$ defined on $\mathcal{S}(\mathbb{R})$ with $D := d/dx$ being the usual differential operator in our case $n = 1$. The operator $-iD : \mathcal{S}(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ is known to be essentially self-adjoint, which allows us to define the (one-dimensional) *momentum operator* by its self-adjoint extension³,

$$\hat{p} := \overline{-iD}. \quad (2.32)$$

Here, the overline on a closable operator denotes its closure, which in the case of an essentially self-adjoint operator is equivalent to its (unique) self-adjoint extension.

One then verifies that the pair $\{\hat{x}, \hat{p}\}$ satisfies the familiar (one-dimensional) *canonical commutation relations (CCR)*,

$$[\hat{x}, \hat{p}] = iI, \quad (2.33)$$

$$[\hat{x}, \hat{x}] = 0, \quad [\hat{p}, \hat{p}] = 0, \quad (2.34)$$

on the subspace $\mathcal{S}(\mathbb{R}) \subset L^2(\mathbb{R})$, where I denotes the identity operator and $[X, Y] := XY - YX$ denotes the commutator for operators X, Y , whose domain is understood to be $\text{dom}([X, Y]) := \text{dom}(XY) \cap \text{dom}(YX)$.

In general, let $\{\mathcal{H}, \mathcal{D}, \{Q, P\}\}$ be a combination consisting of a Hilbert space \mathcal{H} , its dense subspace $\mathcal{D} \subset \mathcal{H}$, and a pair of self-adjoint operators $\{Q, P\}$ on \mathcal{H} . We say that $\{\mathcal{H}, \mathcal{D}, \{Q, P\}\}$ is a (one-dimensional) *representation of the CCR*, if the CCR

$$[Q, P] = iI, \quad (2.35)$$

$$[Q, Q] = 0, \quad [P, P] = 0 \quad (2.36)$$

hold on the domain \mathcal{D} fulfilling

$$\mathcal{D} \subset \text{dom}(QQ) \cap \text{dom}(QP) \cap \text{dom}(PQ) \cap \text{dom}(PP). \quad (2.37)$$

One then concludes from the above argument that the combination,

$$\{L^2(\mathbb{R}), \mathcal{S}(\mathbb{R}), \{\hat{x}, \hat{p}\}\}, \quad (2.38)$$

gives a concrete example for the representation of the CCR, called the (one-dimensional) *Schrödinger representation of the CCR*.

³ While the explicit identification of the domain of the operator \hat{p} is not quite straightforward, we mention that it is given by

$$\text{dom}(\hat{p}) = \left\{ f \in L^2(\mathbb{R}) : f|_J \in \text{AC}(J) \text{ for all compact sub-intervals } J \subset \mathbb{R}, \frac{df}{dx} \in L^2(\mathbb{R}) \right\}, \quad (2.29)$$

where $f|_J$ denotes the restriction of the function f on the interval J , and $\text{AC}(J)$ denotes the space of all absolutely continuous functions on J . Here, a function $f : [a, b] \rightarrow \mathbb{K}$ is called *absolutely continuous*, if for every $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\sum_{k=1}^n (b_k - a_k) < \delta \quad \Rightarrow \quad \sum_{k=1}^n |f(b_k) - f(a_k)| < \epsilon \quad (2.30)$$

holds for arbitrary partitions $a \leq a_1 < b_1 \leq a_2 < b_2 \leq \dots \leq a_n < b_n \leq b$, $n \in \mathbb{N}^\times$ of the interval $[a, b]$. It is known that a function $f : [a, b] \rightarrow \mathbb{K}$ is absolutely continuous if and only if f is differentiable almost everywhere (hence df/dx in (2.29) is well-defined), its derivative is Lebesgue integrable $df/dx \in L^1(J)$, and that

$$f(t) - f(s) = \int_s^t \frac{df}{dx} dx, \quad s, t \in [a, b], \quad s \leq t \quad (2.31)$$

holds (*cf.* fundamental theorem of calculus).

Weyl Representation of the CCR. We call a combination $\{\mathcal{H}, \{Q, P\}\}$ consisting of a Hilbert space \mathcal{H} and a pair of self-adjoint operators $\{Q, P\}$, a (one-dimensional) *Weyl representation of the CCR*, if $\{Q, P\}$ satisfies the *Weyl relations*:

$$e^{isQ}e^{itP} = e^{-istI}e^{itP}e^{isQ}, \quad (2.39)$$

$$e^{isQ}e^{itQ} = e^{itQ}e^{isQ}, \quad e^{isP}e^{itP} = e^{itP}e^{isP}, \quad (2.40)$$

for $s, t \in \mathbb{R}$. One of the advantages of the Weyl relations, as compared to the CCR, is that they deal only with unitary operators, for which no particular consideration for the domain of the involved operators is necessary because of their boundedness. Fortunately, in the present case one can actually prove that the pair $\{\hat{x}, \hat{p}\}$ of the position and momentum operators introduced earlier satisfy the Weyl relations (2.39) and (2.40) on $L^2(\mathbb{R})$. This implies that the Schrödinger representation of the CCR $\{L^2(\mathbb{R}), \mathcal{S}(\mathbb{R}), \hat{x}, \hat{p}\}$ furnishes an example of the Weyl representation of the CCR, at least in the case of the configuration space \mathbb{R} . One also finds that this is true for the Euclidean configuration space \mathbb{R}^n .

One may naturally be interested in how the Weyl representation of the CCR relates to the standard representation of the CCR. To this end, we first begin by collecting some of the necessary definitions and basic theorems. Recall that a vector-valued map $F : U \rightarrow V$ from an open subset $U \subset \mathbb{R}$ to a normed space V is called *strongly continuous* at $t_0 \in U$ if

$$\lim_{u \rightarrow 0} \|F(u + t_0) - F(t_0)\| = 0 \quad (2.41)$$

with respect to the norm $\|\cdot\|$ on V , and in turn, strongly continuous on U if it is strongly continuous at every point of U . The map F is then called *strongly differentiable* at $t_0 \in U$ with strong derivative $F'(t_0) \in V$ if

$$\lim_{u \rightarrow 0} \left\| \frac{F(u + t_0) - F(t_0)}{u} - F'(t_0) \right\| = 0 \quad (2.42)$$

holds, and accordingly strongly differentiable on U if it is strongly differentiable at every point of U . We will occasionally write its strong derivative in either of the notations,

$$\frac{dF(t_0)}{dt} = \frac{dF}{dt}(t_0) = \left. \frac{d}{dt}F(t) \right|_{t=t_0} := F'(t_0). \quad (2.43)$$

Now, let $A : \mathcal{H} \supset \text{dom}(A) \rightarrow \mathcal{H}$ be a self-adjoint operator on a Hilbert space \mathcal{H} , and consider a one-parameter unitary group $\{e^{itA}\}_{t \in \mathbb{R}}$ (defined by means of functional calculus). Then, Stone's theorem on one-parameter unitary groups states that, on account of the boundedness of the unitary operator, for a fixed $|\phi\rangle \in \mathcal{H}$ the unitary group yields a strongly continuous vector-valued map,

$$F : t \mapsto e^{itA}|\phi\rangle, \quad t \in \mathbb{R}, \quad (2.44)$$

for any self-adjoint operator A . However, consideration of the domain $\text{dom}(A)$ becomes necessary when differentiation of the map is considered. In fact, the map is strongly differentiable

on \mathbb{R} if and only if $|\phi\rangle \in \text{dom}(A)$, in which case the derivative reads

$$\frac{dF}{dt}(t) = ie^{itA}A|\phi\rangle = iAe^{itA}|\phi\rangle. \quad (2.45)$$

Returning to our main topic, we rewrite the r. h. s. of (2.39) to obtain

$$e^{isQ}e^{itP} = e^{itP}e^{is(Q-tI)}, \quad s, t \in \mathbb{R}. \quad (2.46)$$

Considering the vector-valued map,

$$s \mapsto e^{isQ}e^{itP}|\psi\rangle = e^{itP}e^{is(Q-tI)}|\psi\rangle, \quad (2.47)$$

for a fixed $|\psi\rangle \in \text{dom}(Q) = \text{dom}(Q - tI)$, $t \in \mathbb{R}$, one concludes from the above argument that the r. h. s. of (2.47) is strongly differentiable at all $s \in \mathbb{R}$ with the derivative

$$\begin{aligned} \frac{d}{ds} \left(e^{itP}e^{is(Q-tI)}|\psi\rangle \right) &= e^{itP} \left(\frac{d}{ds} e^{is(Q-tI)}|\psi\rangle \right) \\ &= ie^{itP}e^{is(Q-tI)}(Q-tI)|\psi\rangle, \quad s, t \in \mathbb{R}. \end{aligned} \quad (2.48)$$

Note here that the first equality follows from the linearity and boundedness (hence, continuity) of the unitary operator e^{itP} . Turning to the l. h. s. of (2.47), differentiability implies that $e^{itP}|\psi\rangle \in \text{dom}(Q)$, whereby one has

$$\frac{d}{ds} (e^{isQ}e^{itP}|\psi\rangle) = ie^{isQ}Qe^{itP}|\psi\rangle. \quad (2.49)$$

Combining the two results, one duly obtains

$$e^{isQ}Qe^{itP}|\psi\rangle = e^{itP}e^{is(Q-tI)}(Q-tI)|\psi\rangle, \quad s, t \in \mathbb{R}. \quad (2.50)$$

Taking $s = 0$, one finds the validity of the operator identity $Qe^{itP} = e^{itP}(Q - tI)$, or equivalently

$$e^{-itP}Qe^{itP} = Q - tI, \quad t \in \mathbb{R}, \quad (2.51)$$

on the subspace $\text{dom}(Q)$. This shows how the unitary adjoint action generated by P results in a parallel translation $Q \mapsto Q - tI$ on its conjugate operator Q^4 . Now, if one further considers the vector-valued map by rewriting (2.51),

$$Qe^{itP}|\psi\rangle = e^{itP}Q|\psi\rangle - te^{itP}|\psi\rangle, \quad t \in \mathbb{R}, \quad (2.52)$$

one proves the differentiability of the r. h. s. for the choice of the initial state $|\psi\rangle \in \text{dom}(PQ) \cap \text{dom}(P)$, which yields

$$\frac{d}{dt} (e^{itP}Q|\psi\rangle - te^{itP}|\psi\rangle) = (ie^{itP}PQ - e^{itP} - ite^{itP}P)|\psi\rangle, \quad t \in \mathbb{R}. \quad (2.53)$$

Turning to the l. h. s., differentiability also leads to

$$\begin{aligned} \frac{d}{dt} (Qe^{itP}|\psi\rangle) &= Q \left(\frac{d}{dt} e^{itP}|\psi\rangle \right) \\ &= iQe^{itP}P|\psi\rangle, \quad t \in \mathbb{R}, \end{aligned} \quad (2.54)$$

where, in particular, $e^{itP}P|\psi\rangle \in \text{dom}(Q)$ is implied, and the first equality is due to the closedness of the operator Q (recall that a self-adjoint operator is necessarily closed). By

⁴Note that what we are discussing here is something more than just proving the Campbell-Baker-Hausdorff formula.

combining the above two results, one has

$$(iQe^{itP}P)|\psi\rangle = (ie^{itP}PQ - e^{itP} - ite^{itP}P)|\psi\rangle, \quad t \in \mathbb{R}. \quad (2.55)$$

Taking $t = 0$, we learn that this in particular leads to the operator identity,

$$QP = PQ + iI \quad \Leftrightarrow \quad [Q, P] = iI, \quad (2.56)$$

on the subspace $\text{dom}(PQ) \cap \text{dom}(P)$. One also sees from this result that the choice $|\psi\rangle \in \text{dom}(PQ) \cap \text{dom}(P)$ automatically implies $|\psi\rangle \in \text{dom}(QP)$.

Proceeding further from (2.40) by analogous reasoning, one eventually obtains the CCR (2.35) and (2.36) on the domain (2.37). In the case where \mathcal{D} is dense, one sees that a Weyl representation of the CCR $\{\mathcal{H}, \{Q, P\}\}$ together with the subspace \mathcal{D} indeed gives a representation of the CCR. In fact, in the case where \mathcal{H} is separable, \mathcal{D} is known to be dense.

In passing, we mention that the importance of the Weyl relations becomes evident when one considers configuration spaces, other than the Euclidean space \mathbb{R}^n , where no reasonable counterpart of the CCR can be defined. For instance, when the configuration space is given by a coset space G/H where G is a Lie group and H its subgroup (typical examples being the spheres $S^n \simeq O(n+1)/O(n)$), one can readily adopt the inherent group theoretic structure of the configuration space to define the Weyl relations extended to the space. Unlike the Euclidean case, such extended Weyl relations are known to admit a multiple of inequivalent representations.

2.1.3. Tensor Product of Hilbert Spaces and Self-adjoint Operators. We finally provide a brief review on tensor products of Hilbert spaces and those of self-adjoint operators. Although the topic is elementary, we find it beneficial to give a summary of its precise definition in consideration of its extensive use due to the nature of this paper focusing on indirect measurement schemes.

Algebraic Tensor Products. Let V, W be \mathbb{K} -vector spaces. We call an ordered pair

$$(V \otimes W, \otimes) \quad (2.57)$$

consisting of a vector space $V \otimes W$ and a bilinear map $\otimes : V \times W \rightarrow V \otimes W$, an (algebraic) *tensor product* of vector spaces V and W , if for any \mathbb{K} -vector space Z and a bilinear map $T : V \times W \rightarrow Z$, there exists a unique linear map $\tilde{T} : V \otimes W \rightarrow Z$ for which the diagram

$$\begin{array}{ccc} V \times W & \xrightarrow{\otimes} & V \otimes W \\ & \searrow T & \downarrow \tilde{T} \\ & & Z \end{array} \quad (2.58)$$

commutes⁵ (universal property of (algebraic) tensor products). Each element of $V \otimes W$ is called a *tensor*, and the bilinear map \otimes is called the *tensor map*, the image of which is

⁵ We say that a diagram is a *commutative diagram*, or more casually, *the diagram commutes*, if all directed paths in the diagram with the same start and endpoints lead to the same result by composition.

denoted by

$$v \otimes w := \otimes(v, w). \quad (2.59)$$

The thus defined tensor products are in fact unique up to isomorphism. Indeed if $(V \otimes W, \otimes)$ and $(V' \otimes' W', \otimes')$ were two of such, then by first letting $Z = V' \otimes W'$ and $T = \otimes'$ in the above diagram, and then subsequently by changing roles of $(V \otimes W, \otimes)$ and $(V' \otimes' W', \otimes')$, one concludes that $\widetilde{\otimes}$ and $\widetilde{\otimes}'$ are linear bijections with $\widetilde{\otimes}' \circ \widetilde{\otimes} = I$. In this sense, we may refer to $(V \otimes W, \otimes)$ as *the* tensor product of V and W , and forget about the way how it is constructed⁶. One of the basic facts worth of special note is that, given two bases $\{e_i\}_{i \in I}$ and $\{f_j\}_{j \in J}$ of V and W , respectively, the tensors $\{e_i \otimes f_j\}_{i \in I, j \in J}$ form a basis of $V \otimes W$.

Tensor Product of Hilbert Spaces. We are specifically interested in tensor products of Hilbert spaces. For a pair of Hilbert spaces $(\mathcal{H}_1, \langle \cdot, \cdot \rangle_{\mathcal{H}_1})$ and $(\mathcal{H}_2, \langle \cdot, \cdot \rangle_{\mathcal{H}_2})$, we denote by

$$(\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2, \widehat{\otimes}) \quad (2.60)$$

their algebraic tensor product defined from their purely algebraic structures described as above. We then introduce

$$\langle \phi_1 \widehat{\otimes} \phi_2, \psi_1 \widehat{\otimes} \psi_2 \rangle := \langle \phi_1, \psi_1 \rangle_{\mathcal{H}_1} \langle \phi_2, \psi_2 \rangle_{\mathcal{H}_2}, \quad \phi_i, \psi_i \in \mathcal{H}_i \quad (2.61)$$

defined for pairs of all tensors of the form $D := \{v \otimes w : v \in V, w \in W\}$, and let it extend linearly on whole $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2 = \text{Span}(D)$. Here,

$$\text{span}(S) := \{k_1 v_1 + \cdots + k_n v_n : k_i \in \mathbb{K}, v_i \in S, n \in \mathbb{N}\} \quad (2.62)$$

denotes the subspace of a \mathbb{K} -vector space V spanned by a nonempty set $S \subset V$, *i.e.* the set of all finite linear combinations of vectors belonging to S . It is routine to check that the thus defined extension $\langle \cdot, \cdot \rangle_{\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2}$ is well-defined, and one moreover proves that the extension in fact makes itself an inner product on $\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2$, making the pair $(\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2, \langle \cdot, \cdot \rangle_{\mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2})$ into a pre-Hilbert space (*i.e.*, an inner product space). The tensor map $\widehat{\otimes}$ can be also shown to be continuous with respect to the topology that the inner product generates. We then finally define the completion of the pre-Hilbert space, and denote it by

$$(\mathcal{H}_1 \otimes \mathcal{H}_2, \langle \cdot, \cdot \rangle_{\mathcal{H}_1 \otimes \mathcal{H}_2}). \quad (2.63)$$

The new space $(\mathcal{H}_1 \otimes \mathcal{H}_2, \langle \cdot, \cdot \rangle_{\mathcal{H}_1 \otimes \mathcal{H}_2})$ is a Hilbert space by construction, and together with the continuous extension \otimes of the bilinear map $\widehat{\otimes}$, is called the (topological) *tensor product of the Hilbert spaces* \mathcal{H}_1 and \mathcal{H}_2 . The map \otimes is called the *tensor map* and the elements of $\mathcal{H}_1 \otimes \mathcal{H}_2$ are called *tensors*.

⁶ One finds several concrete constructions of tensor products in various literatures. See, for example [30].

Tensor Product of Linear Operators. A pair of linear operators $A_i : \mathcal{H}_i \supset \text{dom}(A_i) \rightarrow \mathcal{H}_i$, $i = 1, 2$, defines a natural bilinear map

$$A_1 \times A_2 : \text{dom}(A_1) \times \text{dom}(A_2) \rightarrow \mathcal{H}_1 \times \mathcal{H}_2, \quad (|\phi_1\rangle, |\phi_2\rangle) \mapsto (A_1|\phi_1\rangle, A_2|\phi_2\rangle). \quad (2.64)$$

From the universal property of the algebraic tensor product mentioned above, one readily sees the existence of a unique linear map

$$A_1 \widehat{\otimes} A_2 : \text{dom}(A_1) \widehat{\otimes} \text{dom}(A_2) \rightarrow \mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2 \quad (2.65)$$

that makes the diagram

$$\begin{array}{ccc} \text{dom}(A_1) \times \text{dom}(A_2) & \xrightarrow{\widehat{\otimes}} & \text{dom}(A_1) \widehat{\otimes} \text{dom}(A_2) \\ A_1 \times A_2 \downarrow & \searrow \circlearrowright & \downarrow A_1 \widehat{\otimes} A_2 \\ \mathcal{H}_1 \times \mathcal{H}_2 & \xrightarrow{\widehat{\otimes}} & \mathcal{H}_1 \widehat{\otimes} \mathcal{H}_2 \end{array} \quad (2.66)$$

commute. Note in particular that the diagram implies

$$A_1 \widehat{\otimes} A_2(|\phi_1\rangle \widehat{\otimes} |\phi_2\rangle) = |A_1\phi_1\rangle \widehat{\otimes} |A_2\phi_2\rangle, \quad |\phi_i\rangle \in \text{dom}(A_i), \quad i = 1, 2. \quad (2.67)$$

Extending both the domain and the range of (2.65), we can think of

$$A_1 \widehat{\otimes} A_2 : \mathcal{H}_1 \otimes \mathcal{H}_2 \supset \text{dom}(A_1) \widehat{\otimes} \text{dom}(A_2) \rightarrow \mathcal{H}_1 \otimes \mathcal{H}_2 \quad (2.68)$$

as an operator on the Hilbert space $\mathcal{H}_1 \otimes \mathcal{H}_2$.

Tensor Product of Self-Adjoint Operators. Now, for a pair of densely defined closable operators $A_i : \mathcal{H}_i \supset \text{dom}(A_i) \rightarrow \mathcal{H}_i$, $i = 1, 2$, the operator (2.68) itself is known to be closable, whereby we define the *tensor product*

$$A_1 \otimes A_2 := \overline{A_1 \widehat{\otimes} A_2} \quad (2.69)$$

of the pair by its closure. Specifically, since self-adjoint operators are densely defined and closed, the tensor product (2.69) is always well-defined. Although self-adjointness is not preserved in general by taking (2.68), its essential self-adjointness is at least known to be guaranteed. As the closure of an essentially self-adjoint operator, this makes the tensor product (2.69) itself self-adjoint, which is precisely the definition of the *tensor product of self-adjoint operators*.

2.2. Unconditioned Measurement

Now that we have reviewed the necessary materials, we begin our study on the unconditioned measurement scheme. Suppose that the experimenter wishes to extract information of the combination of a given but unknown observable A and a state $|\phi\rangle \in \mathcal{H}$ of the target system, without direct access to it. To accomplish this, one first arranges an auxiliary meter system \mathcal{K} equipped with a pair of observables $\{Q, P\}$ for which $\{\mathcal{K}, \{Q, P\}\}$ gives a Weyl representation of the CCR. As we have seen above, the choice $\mathcal{K} = L^2(\mathbb{R})$, $Q = \hat{x}$ and $P = \hat{p}$ gives a concrete example. One then prepares the meter system in a certain initial state represented by the vector $|\psi\rangle \in \mathcal{K}$, and combines the two systems into the direct product state $|\phi \otimes \psi\rangle \in \mathcal{H} \otimes \mathcal{K}$.

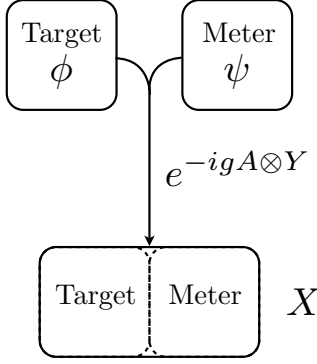


Fig. 1 A graphical illustration of the unconditioned measurement scheme. The figure is to be read from top to bottom. The initial state preparation stage of both the target and the meter systems is depicted in the top part, and the manner in which the two quantum systems undergoes a von Neumann type interaction is illustrated in the middle part. The composite system after the interaction, which is depicted in the bottom part, generally becomes entangled. One finally performs a measurement of an observable X on the meter system.

Choosing an observable Y of the meter system \mathcal{K} either by $Y = Q$ or $Y = P$, the composite system is subjected to a von Neumann type interaction,

$$|\Psi^g\rangle := e^{-igA\otimes Y}|\phi\otimes\psi\rangle, \quad g \in \mathbb{R}, \quad (2.70)$$

i.e., a unitary evolution on the composite system parametrised by a real number g , which is often interpreted as the intensity, its time duration, or the combination thereof, of the interaction between the two systems. Finally, the experimenter performs local measurement of an observable X of the meter system \mathcal{K} by choosing either by $X = Q$ or $X = P$ (chosen independently of Y), or equivalently $I \otimes X$ on the generally entangled composite state $|\Psi^g\rangle$ after the interaction (see figure 1).

As a preparation for further analysis, we first introduce the reduced density operator

$$\psi^g := \text{Tr}_{\mathcal{H}}[|\Psi^g\rangle\langle\Psi^g|] \quad (2.71)$$

representing the state of the meter system \mathcal{K} after the measurement⁷ obtained by taking the partial trace of the composite state $|\Psi^g\rangle$ with respect to the target system \mathcal{H} . The quantity of interest for our measurement is thus the expectation value

$$\begin{aligned} \mathbb{E}[I \otimes X; \Psi^g] &:= \frac{\langle\Psi^g, (I \otimes X)\Psi^g\rangle}{\|\Psi^g\|^2} \\ &= \frac{\text{Tr}_{\mathcal{K}}[X \text{Tr}_{\mathcal{H}}[|\Psi^g\rangle\langle\Psi^g|]]}{\text{Tr}_{\mathcal{K}}[\text{Tr}_{\mathcal{H}}[|\Psi^g\rangle\langle\Psi^g|]]} \\ &= \frac{\text{Tr}_{\mathcal{K}}[X\psi^g]}{\text{Tr}_{\mathcal{K}}[\psi^g]} \\ &=: \mathbb{E}[X; \psi^g], \end{aligned} \quad (2.72)$$

of the observable $I \otimes X$ on the composite state $|\Psi^g\rangle$ after the interaction, which can interchangeably be written in terms of that of the local observable X on the density matrix $|\psi^g\rangle$ of the meter system.

⁷ Here we are adopting, instead of the more common usage ρ^g , a slightly unusual notation ψ^g to denote the generically mixed state of the meter. This we do because we wish to reserve the letter ρ for the density of some absolutely continuous complex measures (see Section 3.1.2). However, our notation has an advantage on its own in that, if we also write the state as $|\psi^g\rangle$ when it is pure as we usually do, the correspondence between the two, ψ^g and $|\psi^g\rangle$ (both represent the same state), becomes obvious.

Main Objective of this Subsection. The main objective of this subsection is to demonstrate the following basic proposition, which provides the sufficient condition for its well-definedness and its explicit evaluations. For definiteness, we shall from now on fix $Y = P$ without loss of generality.

Proposition 2.2 (Unconditioned Measurement I). *In the context of the UM scheme, let $Y = P$ for definiteness. Given the right choices*

- (i) *If $X = Q$: $|\phi\rangle \in \text{dom}(A)$, $|\psi\rangle \in \text{dom}(X)$,*
- (ii) *If $X = P$: $|\phi\rangle \in \mathcal{H}$, $|\psi\rangle \in \text{dom}(X)$,*

of the initial states of both the target and the meter systems, depending on the choice of the observable X on the meter system to be measured, the composite state after the interaction lies in $|\Psi^g\rangle \in \text{dom}(I \otimes X)$, $g \in \mathbb{R}$. The expectation value (2.72) thus remains finite for all range of the interaction parameter, which reads

$$\mathbb{E}[X; \psi^g] = \begin{cases} \mathbb{E}[Q; \psi] + g \mathbb{E}[A; \phi], & (X = Q) \\ \mathbb{E}[P; \psi], & (X = P) \end{cases} \quad (2.73)$$

for each of the choice of X .

Some Operator Identities. Before we move on to the proof, we make notes on some important operator identities that will be extensively used throughout this paper. Our analysis is based on the following operator identities on the composite Hilbert space $\mathcal{H} \otimes \mathcal{K}$, similar to those of (2.39) and (2.40).

Lemma 2.3. *Let H and K be Hilbert spaces, and let A be a self-adjoint operator on \mathcal{H} , and $\{Q, P\}$ be a pair of self-adjoint operators on \mathcal{K} for which $\{\mathcal{K}, \{Q, P\}\}$ defines a Weyl representation of the CCR. Then, the operator equalities*

$$e^{isI \otimes Q} e^{itA \otimes P} = e^{-istA \otimes I} e^{itA \otimes P} e^{isI \otimes Q}, \quad s, t \in \mathbb{R}, \quad (2.74)$$

$$e^{isI \otimes P} e^{itA \otimes P} = e^{itA \otimes P} e^{isI \otimes P}, \quad s, t \in \mathbb{R}, \quad (2.75)$$

hold.

Proof. Since (2.75) is trivial, we only need to prove (2.74). To this end, we first consider the special case where the self-adjoint operator A on the target system \mathcal{H} has a spectrum $\sigma(A)$ of finite cardinality. Letting $\sigma(A) = \{a_1, \dots, a_N\}$, $N \in \mathbb{N}^\times$ be any enumeration of its eigenvalues, the spectral decomposition of A reads

$$A = \sum_{n=1}^N a_n \Pi_{a_n}, \quad (2.76)$$

where Π_{a_n} is the projection on the eigenspace associated with the eigenvalue a_n . In the case where the eigenspace is one-dimensional (or non-degenerate), one may write $\Pi_{a_n} = |a_n\rangle\langle a_n|$ with the eigenstate $|a_n\rangle$ for which $A|a_n\rangle = a_n|a_n\rangle$ holds (more on the topic of spectral decomposition in Section 3.1.5). Now, for an arbitrary self-adjoint operator Z on the meter system \mathcal{K} , one may expect from the defining property $\Pi_{a_n}^2 = \Pi_{a_n}$ of projections that the

formal computation

$$\begin{aligned}
e^{itA \otimes Z} &= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} (A \otimes Z)^k \\
&= \sum_{k=0}^{\infty} \frac{(it)^k}{k!} \left(\left(\sum_{n=1}^N a_n^k \Pi_{a_n} \right) \otimes Z^k \right) \\
&= \sum_{k=0}^{\infty} \sum_{n=1}^N \left(\Pi_{a_n} \otimes \frac{(ita_n Z)^k}{k!} \right) \\
&= \sum_{n=1}^N (\Pi_{a_n} \otimes e^{ita_n Z}), \quad t \in \mathbb{R},
\end{aligned} \tag{2.77}$$

is legitimate. This in fact turns out to be correct as an operator identity on $\mathcal{H} \otimes \mathcal{K}$ with full rigour, which can be proven in a fairly straightforward manner by means of rudimentary techniques of functional calculus. One then has

$$\begin{aligned}
e^{isI \otimes Q} e^{itA \otimes P} &= (I \otimes e^{isQ}) \left(\sum_{n=1}^N (\Pi_{a_n} \otimes e^{ita_n P}) \right) \\
&= \sum_{n=1}^N (\Pi_{a_n} \otimes (e^{isQ} e^{ita_n P})) \\
&= \sum_{n=1}^N (\Pi_{a_n} \otimes (e^{ista_n} e^{ita_n P} e^{isQ})) \\
&= \sum_{n=1}^N (\Pi_{a_n} \otimes e^{ita_n (P-sI)}) (I \otimes e^{isQ}) \\
&= e^{itA \otimes (P-sI)} e^{isI \otimes Q} \\
&= e^{-istA \otimes I} e^{itA \otimes P} e^{isI \otimes Q}, \quad s, t \in \mathbb{R},
\end{aligned} \tag{2.78}$$

which proves (2.74) for our special case, where we have used (2.39) in the third step. Returning to the general case in which A is now an arbitrary self-adjoint operator, one observes that the well-definedness of both the left-most and right-most hand sides of the above equality remains valid. From this, one may expect that the same result also holds for the general case, which indeed turns out to be true (as usual, one may prove this without much difficulty through rudimentary techniques of functional calculus). \blacksquare

Measurement Outcomes. We now return to the main problem of this subsection. We are interested in finding the condition for which (2.72) is well-defined, and subsequently in obtaining an explicit formula in terms of the components of both the target and the meter system. Since most of the techniques employed here is the same as those introduced in Section 2.1, we shall proceed by sketching the proofs.

Proof of Proposition 2.2. Let us begin by choosing the operator $X = Q$ for the measurement of the meter system, and thereby rewrite the r. h. s. of (2.74) to obtain

$$e^{isI \otimes Q} e^{itA \otimes P} = e^{itA \otimes P} e^{is(\overline{I \otimes Q - tA \otimes I})}, \quad s, t \in \mathbb{R} \tag{2.79}$$

for better usability⁸. By differentiating both sides of the above equality and taking $s = 0$, an analogous argument given earlier for obtaining (2.51) leads to the operator identity,

$$e^{-itA \otimes P}(I \otimes Q)e^{itA \otimes P} = \overline{I \otimes Q - tA \otimes I}, \quad t \in \mathbb{R}, \quad (2.80)$$

on the subspace $\text{dom}(\overline{I \otimes Q - tA \otimes I})$. This ensures that, if $|\Phi\rangle \in \text{dom}(\overline{I \otimes Q - tA \otimes I})$, then one has

$$e^{itA \otimes P}|\Phi\rangle \in \text{dom}(I \otimes Q). \quad (2.81)$$

Here, it may be worthwhile to note the analogy between (2.51) and (2.80). To put this in our context, let $t = -g$ above. If one chooses $|\psi\rangle \in \text{dom}(Q)$ as the meter state, and likewise assumes $|\phi\rangle \in \text{dom}(A)$ as the system state prepared prior to the interaction, one has in particular $|\phi \otimes \psi\rangle \in \text{dom}(\overline{I \otimes Q + gA \otimes I})$. Then, equating $|\Phi\rangle = |\phi \otimes \psi\rangle$ in (2.81), we find

$$|\Psi^g\rangle = e^{-igA \otimes P}|\phi \otimes \psi\rangle \in \text{dom}(I \otimes Q), \quad g \in \mathbb{R}. \quad (2.82)$$

This guarantees that the expectation value (2.72) of the observable $I \otimes Q$ on the composite state $|\Psi^g\rangle$ remains finite and is given by

$$\begin{aligned} \mathbb{E}[I \otimes Q; \Psi^g] &:= \frac{\langle \Psi^g, (I \otimes Q)\Psi^g \rangle}{\|\Psi^g\|^2} \\ &= \frac{\langle \phi \otimes \psi, (e^{igA \otimes P}(I \otimes Q)e^{-igA \otimes P})\phi \otimes \psi \rangle}{\|\phi\|^2\|\psi\|^2} \\ &= \frac{\langle \phi \otimes \psi, \overline{I \otimes Q + gA \otimes I}\phi \otimes \psi \rangle}{\|\phi\|^2\|\psi\|^2} \\ &= \mathbb{E}[Q; \psi] + g\mathbb{E}[A; \phi], \quad g \in \mathbb{R}, \end{aligned} \quad (2.83)$$

for any such combination of the initial states.

Evidently, for the choice $X = P$, one finds the validity of the operator identity

$$e^{itA \otimes P}(I \otimes P)e^{-itA \otimes P} = I \otimes P, \quad t \in \mathbb{R}, \quad (2.84)$$

on the subspace $\text{dom}(I \otimes P)$ by analogous reasoning. From this, one readily concludes that the expectation value of $I \otimes P$ reads

$$\mathbb{E}[I \otimes P; \Psi^g] = \mathbb{E}[P; \psi], \quad g \in \mathbb{R}, \quad (2.85)$$

which is well-defined for any choice of the state $|\psi\rangle \in \text{dom}(P)$ of the meter system and $g \in \mathbb{R}$, irrespective of the initial choice of the state $|\phi\rangle \in \mathcal{H}$ of the target system. \blacksquare

2.3. Recovery of the Target Profile

Now that we have revealed the explicit behaviour of the measurement outcomes of the meter, we are thus interested in recovering the information of the target system from it. As one may expect from the statement in Proposition 2.2, the information of the target system (which should essentially consist of the specification of the pair of A and $|\phi\rangle$) manifests

⁸Note here that the sum of two (possibly unbounded) self-adjoint operators is not necessarily self-adjoint. Fortunately, essential self-adjointness is at least assured for the sum of $I \otimes Q$ and $A \otimes I$ for our case. We may thus take the self-adjoint extension of their sum in order to ensure its self-adjointness (more to this in Section 3.1.7).

itself in the form of the expectation value $\mathbb{E}[A; \phi]$. In recovering the desired information, one subsequently recognises from (2.73) that it fully suffices to examine only the outcomes of the measurement of the observable X conjugate to Y , and there is no use for that of the choice $X = Y$ (this is to be contrasted with the conditional measurement we discuss later). Specifically, one finds below that there are two typical techniques in obtaining the desired information: one is to investigate the behaviour of the measurement outcome (2.73) in the strong region $g \rightarrow \pm\infty$ of the interaction parameter, and the other is to examine the local behaviour of it around $g = 0$, which shall be respectively called the *strong unconditioned measurement* and the *weak unconditioned measurement* in this paper.

2.3.1. Strong Unconditioned Measurement. Our result (2.73) shows that the expectation value of the measurement of $X = Q$ behaves linearly with respect to g , and that its growth is proportional to the expectation value $\mathbb{E}[A; \phi]$ of the target observable. The experimenter would thus divide the measurement outcomes of Q by g and then take the limit of the strong coupling $g \rightarrow \pm\infty$ (or equivalently $g^{-1} \rightarrow 0$):

$$\begin{aligned} \lim_{g^{-1} \rightarrow 0} \frac{\mathbb{E}[Q; \psi^g]}{g} &= \mathbb{E}[A; \phi] + \lim_{g^{-1} \rightarrow 0} \frac{\mathbb{E}[Q; \psi]}{g} \\ &= \mathbb{E}[A; \phi]. \end{aligned} \tag{2.86}$$

allowing the recovery of the desired information of the target system $\mathbb{E}[A; \phi]$ in the form of expectation values⁹.

2.3.2. Weak Unconditioned Measurement. The same information may be obtained by examining the weak region ($g \rightarrow 0$) of the interaction. Indeed, one trivially finds from (2.73) that

$$\left. \frac{d^n}{dg^n} \mathbb{E}[Q; \psi^g] \right|_{g=0} = \begin{cases} \mathbb{E}[Q; \psi], & n = 0, \\ \mathbb{E}[A; \phi], & n = 1, \\ 0, & n \geq 2, \end{cases} \tag{2.89}$$

which implies that the expectation value $\mathbb{E}[A; \phi]$ of our interest may also be obtained as the first differential coefficient ($n = 1$) of the measured outcome at $g = 0$.

2.3.3. Discussion. While this whole section consisted of rather trivial results, the line of arguments presented here serves as the baseline of our analysis throughout this paper. Namely, we first examine the full behaviour of the target of our measurement (for this section, it was the expectation value (2.72) of the observable X of the meter) and intend to obtain

⁹ Alternatively, one may consider the shift of the expectation value,

$$\Delta_X(g) := \mathbb{E}[X; \psi^g] - \mathbb{E}[X; \psi^0] = \begin{cases} g\mathbb{E}[A; \phi], & (X = Q), \\ 0, & (X = P), \end{cases} \tag{2.87}$$

for $g \in \mathbb{R}$ as a quantity directly related to the observable A of the system. For the choice $X = Q$, one then simply has

$$\frac{\Delta_Q(g)}{g} = \mathbb{E}[A; \phi], \quad g \in \mathbb{R}^\times, \tag{2.88}$$

which might be a more straight-forward way to be employed practically.

an explicit description of how the profile of the initial configuration of the target system gets mixed into that of the meter system through the interaction (which, for the current case, is the result (2.73)). We then intend to extract the information of the target system (for this section, it is the expectation value $\mathbb{E}[A; \phi]$) by separating it from the measurement outcomes. Specifically, we find that examining either the strong or the weak region of the interaction parameter g reveals itself useful for this purpose, and this should be the strategy that we take in the subsequent sections.

In the next section, the UM scheme is analysed in depth in terms of probabilities, following the same line as described above. Specifically, while the distinction between the strong and the weak measurements looked rather vague at the operator level, we shall see shortly that these two strategies are recognised to be qualitatively different from the viewpoint of probabilities.

3. Unconditioned Measurement II: In Terms of Probabilities

We have so far conducted an analysis of the UM scheme on the operator level, where the quantity of interest is the *expectation value* of an observable. However, one may be interested in the raw information that the measurement provides, *i.e.*, the *probability* describing the behaviour of each measurement outcomes, which is the target of our study in this section.

3.1. Reference Materials

To prepare for our discussion, we here provide a concise summary on the topic of complex measures and integration with respect to them. We next make a brief review on the spectral theorem for self-adjoint operators and recall the general framework for describing the ideal measurement of a quantum observable. Subsequently, we expound on density functions and see how this relates to the description by measures.

3.1.1. The Space of Complex Measures. As a preparation in dealing with the spectral theorem for self-adjoint operators, we collect below the basic definitions and results regarding complex measures and integration with respect to them.

Signed Measures, Jordan Decomposition and Total Variation. Let (X, \mathfrak{A}) be a measurable space. A map $\nu : \mathfrak{A} \rightarrow \overline{\mathbb{R}}$ is called a *signed measure*, if it satisfies the following properties:

- (i) $\nu(\emptyset) = 0$.
- (ii) $\nu(\mathfrak{A}) \subset]-\infty, +\infty]$ or $\nu(\mathfrak{A}) \subset [-\infty, +\infty[$.
- (iii) Countable additivity (2.1) holds for any sequence $(A_n)_{n \geq 1}$ of pairwise disjoint subsets of X .

They are, in a sense, generalisations of the concept of the standard measures by allowing negative numbers to be assigned to each measurable sets. A signed measure ν is called *finite* if $\nu(\mathfrak{A}) \subset \mathbb{R}$. One of the most important properties of a signed measure is described by the *Jordan decomposition theorem*, which states that every signed measure ν has the *Jordan decomposition*, *i.e.*, a unique decomposition of ν into a difference

$$\nu = \nu^+ - \nu^- \tag{3.1}$$

of two measures ν^+ and ν^- , respectively called the *positive* and *negative variation* of ν , and at least one of which being finite. Here, the positive and negative variations are *singular* to one another, denoted as $\nu^+ \perp \nu^-$, in the sense there exists a decomposition of $X = P \cup N$ into two measurable sets such that $\nu^+(N) = 0$ and $\nu^-(P) = 0$ holds. The Jordan decomposition is minimal in the following sense: Given any decomposition $\nu = \rho - \sigma$ of ν into two measures ρ, σ , at least one of which being finite, then $\nu^+ \leq \rho, \nu^- \leq \sigma$ holds.

Let $\mathbf{M}(\mathfrak{A})$ denote the collection of all finite signed measures. One readily sees that $\mathbf{M}(\mathfrak{A})$ becomes an \mathbb{R} -linear space, equipped with the natural addition $(\mu + \nu)(A) := \mu(A) + \nu(A)$ and scalar multiplication $(c\mu)(A) := c\mu(A)$ for $\mu, \nu \in \mathbf{M}(\mathfrak{A})$ and $c \in \mathbb{R}$. Now, let $\nu = \nu^+ - \nu^-$ be the Jordan decomposition of $\nu \in \mathbf{M}(\mathfrak{A})$, and define a new measure by their sum

$$|\nu| := \nu^+ + \nu^-, \tag{3.2}$$

called the *variation* of ν . We then define its *total variation* by $\|\nu\| := |\nu|(X)$, which is nothing but the evaluation of the whole space X by the non-negative measure $|\nu|$. One proves that

the total variation defines a norm on $\mathbf{M}(\mathfrak{A})$, and in fact makes $(\mathbf{M}(\mathfrak{A}), \|\cdot\|)$ into a real Banach space.

Complex Measures. Let (X, \mathfrak{A}) be a measurable space. A map $\nu : \mathfrak{A} \rightarrow \mathbb{C}$ is called a *complex measure*, when it is countably additive (2.1). One sees that ν is a complex measure if and only if both its real and imaginary parts $\operatorname{Re}[\nu]$, $\operatorname{Im}[\nu]$ are finite signed measures. Analogous to the case of signed measures, the collection $\mathbf{M}_{\mathbb{C}}(\mathfrak{A})$ of all complex measures on (X, \mathfrak{A}) becomes a \mathbb{C} -linear space, equipped with the natural addition and scalar multiplication. For a complex measure $\nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{A})$, we define the *variation* of a measurable set $A \in \mathfrak{A}$ by

$$|\nu|(A) := \sup \left\{ \sum_{j=1}^{\infty} |\nu(A_j)| : A_j \in \mathfrak{A} \text{ disjoint for } j \geq 1, A = \bigcup_{j=1}^{\infty} A_j \right\}, \quad (3.3)$$

and also its *total variation*,

$$\|\nu\| := |\nu|(X). \quad (3.4)$$

The definition coincides with the previous definition when ν happens to be a signed measure. The total variation $\|\nu\|$ of ν is known to be the smallest positive measure μ on (X, \mathfrak{A}) satisfying $|\nu(A)| \leq \mu(A)$, $A \in \mathfrak{A}$. In parallel to the case of signed measures, one finds that the total variation defines a norm on the linear space $\mathbf{M}_{\mathbb{C}}(\mathfrak{A})$ and makes $(\mathbf{M}_{\mathbb{C}}(\mathfrak{A}), \|\cdot\|)$ into a complex Banach space.

Integration over Complex Measures. It is now tempting to define integration with respect to complex measures, as a natural extension to that defined for (standard) measures. For a complex measure $\nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{A})$, we let $\rho := \operatorname{Re}[\nu]$, $\sigma := \operatorname{Im}[\nu]$ and consider the intersection of the spaces

$$\mathcal{L}^1(\nu) := \mathcal{L}^1(\rho^+) \cap \mathcal{L}^1(\rho^-) \cap \mathcal{L}^1(\sigma^+) \cap \mathcal{L}^1(\sigma^-), \quad (3.5)$$

where ρ^{\pm} and σ^{\pm} respectively being the positive and negative variations of ρ and σ . We then define the Lebesgue integral of $f \in \mathcal{L}^1(\nu)$ with respect to ν by

$$\int_X f d\nu := \int_X f d\rho^+ - \int_X f d\rho^- + i \int_X f d\sigma^+ - i \int_X f d\sigma^-. \quad (3.6)$$

Linearity of the Lebesgue integral with respect to the complex measure follows naturally as expected.

New Measure from Old. There are several ways to construct a new (complex) measure from a given measure. We mention below two of the most important manners that are frequently employed throughout this paper.

(A) *Measure with Density.* Let (X, \mathfrak{A}, μ) be a measure space. Given a μ -integrable function $f : X \rightarrow \mathbb{C}$, one may define a complex measure by

$$\nu(A) := \int_A f d\mu, \quad A \in \mathfrak{A}. \quad (3.7)$$

The complex measure constructed in this manner is occasionally called the *complex measure with the density f with respect to μ* , and we write it as $\nu = f \odot \mu$. A measurable function $g : X \rightarrow \mathbb{K}$ is known to be $(f \odot \mu)$ -integrable, if and only if the product

$g \cdot f$ is μ -integrable, in which case the equality

$$\int_X g \, d(f \odot \mu) = \int_X g \cdot f \, d\mu \quad (3.8)$$

holds.

(B) *Image Measure.* Let (X, \mathfrak{A}, μ) be a measure space. Given another measurable space (Y, \mathfrak{B}) and a measurable map $f : X \rightarrow Y$, one may construct a new measure on (Y, \mathfrak{B}) by

$$f(\mu)(B) := \mu(f^{-1}(B)), \quad B \in \mathfrak{B}, \quad (3.9)$$

called the *image measure (push-forward measure)* of μ with respect to f . A measurable function $g : Y \rightarrow \widehat{\mathbb{K}}$ is known to be $f(\mu)$ -integrable, if and only if the composition $g \circ f$ is μ -integrable, in which case the *change of variables formula*

$$\int_Y g \, df(\mu) = \int_X g \circ f \, d\mu \quad (3.10)$$

holds.

Measure Algebra. The space of complex measures has an additional well-known structure regarding convolutions. The *convolution* of the two complex measures $\mu, \nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ is defined by

$$(\mu * \nu)(B) := \int_{\mathbb{R}^n} \mu(B - x) \, d\nu(x), \quad B \in \mathfrak{B}^n. \quad (3.11)$$

One can easily confirm that the convolution is a bilinear operation, and is moreover shown to be associative $\mu * (\nu * \rho) = (\mu * \nu) * \rho$ and commutative $\mu * \nu = \nu * \mu$. Together with the evaluation $\|\mu * \nu\| \leq \|\mu\| \|\nu\|$ based on the total variation norm (3.4), one sees that the convolution makes the complex Banach space $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ into a complex commutative Banach algebra, called the *measure algebra* of \mathfrak{B}^n . The measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ has a multiplicative identity e given by the delta measure $e = \delta_0$ centred at the origin, that is,

$$\mu * \delta_0 = \delta_0 * \mu = \mu \quad (3.12)$$

holds for all $\mu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$. Here, the *delta measure* (or the *Dirac measure*) δ_a is a finite measure centred at $a \in \mathbb{R}^n$ defined by

$$\delta_a(B) = \begin{cases} 1, & a \in B, \\ 0, & a \notin B, \end{cases} \quad B \in \mathfrak{B}^n, \quad (3.13)$$

characterised by the integral

$$\int_{\mathbb{R}^n} f(x) \, d\delta_a(x) = f(a), \quad (3.14)$$

whenever the integration is well-defined. It is essentially the same object as the delta distribution that appears in the theory of generalised functions.

3.1.2. The Space of Density Functions. For later use, we are particularly interested in the special subspace of the space $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ of complex measures, namely, the space of *absolutely continuous* complex measures with respect to the Lebesgue-Borel measure β^n . We shall provide a concise review on its definition, make comments on its relation to the space of complex *density functions*, and see that the subspace reveals itself to be a sub-algebra of the measure algebra.

Absolute Continuity and Density Functions. Let μ and ν be signed (or complex) measures on a measurable space (X, \mathfrak{A}) . We say that ν is μ -continuous or *absolutely continuous with respect to μ* , written as $\nu \ll \mu$, if $\mu(A) = 0$ implies $\nu(A) = 0$ for all $A \in \mathfrak{A}$. A signed measure μ is called σ -finite if there exists a sequence $(A_n)_{n \geq 1}$ of disjoint measurable sets $A_n \in \mathfrak{A}$ satisfying $X = \bigcup_{n=1}^{\infty} A_n$ and $|\mu(A_n)| < \infty$ ($n \in \mathbb{N}^\times$). By definition, finite measures are always σ -finite. The Lebesgue-Borel measure β^n is among the most important examples of σ -finite measures. The following theorem is of great importance.

Theorem (Radon-Nikodým Theorem for Complex Measures). *Let μ be a σ -finite measure and $\nu \ll \mu$ be a complex measure. Then, ν has a density with respect to μ , that is, there exists a μ -integrable function $\rho: X \rightarrow \mathbb{C}$ such that $\nu = \rho \odot \mu$, and ρ is unique μ -a.e. If ν happens to be positive, then one may choose $\rho \geq 0$.*

In the above situation of the Radon-Nikodým theorem, the function ρ satisfying $\nu = \rho \odot \mu$ is called the *Radon-Nikodým derivative* (or more casually, the *density*), and is denoted by

$$\rho =: \frac{d\nu}{d\mu}. \quad (3.15)$$

This is nothing but to say that

$$\nu(A) = \int_A \frac{d\nu}{d\mu}(x) d\mu(x), \quad A \in \mathfrak{A}, \quad (3.16)$$

holds, if explicitly written out. For a ν -integrable function f , a direct application of (3.8) leads to

$$\int_{\mathbb{X}} f(x) d\nu(x) = \int_X f(x) \frac{d\nu}{d\mu}(x) d\mu(x), \quad (3.17)$$

in which the notation for the Radon-Nikodým derivative (which might at first seems strange) reveals its advantage.

Absolute Continuity with respect to the Lebesgue-Borel Measure. We are particularly interested in the sub-family $L^1(\mathfrak{B}^n) \subset \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ consisting of complex measures that are absolutely continuous with respect to the Lebesgue-Borel measure β^n on $(\mathbb{R}^n, \mathfrak{B}^n)$. Whenever there is no risk of confusion, members of $L^1(\mathfrak{B}^n)$ shall occasionally be referred to as *absolutely continuous measures*, simply without reference to the base measure β^n . One readily finds that the collection $L^1(\mathfrak{B}^n)$ forms a linear subspace of $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$. Now, uniqueness β^n -a.e. of the Radon-Nikodým derivative allows us to define a linear map

$$L^1(\mathfrak{B}^n) \rightarrow L^1(\mathbb{R}^n), \quad \mu \mapsto \frac{d\mu}{d\beta^n}, \quad (3.18)$$

which maps an absolutely continuous complex measure to its density. Conversely, one may construct a new complex measure given an integrable function $f \in L^1(\mathbb{R}^n)$ by $\nu := f \odot \beta^n$. From this, one obtains a bijective linear map between the space of absolutely continuous complex measures $L^1(\mathfrak{B}^n)$ and the space of integrable functions $L^1(\mathbb{R}^n)$, associating an absolutely continuous complex measure $\nu \in L^1(\mathfrak{B}^n)$ to its density $d\nu/d\beta^n \in L^1(\mathbb{R}^n)$. In this manner, one may identify a specific subspace of the space of complex measures with that of integrable functions as

$$L^1(\mathfrak{B}^n) \cong L^1(\mathbb{R}^n), \quad (3.19)$$

and may translate and interpret various properties of complex measures in terms of density functions. To discuss how this works, let $d\nu/d\beta^n \in L^1(\mathbb{R}^n)$ be the density of $\nu \in L^1(\mathfrak{B}^n)$ with

respect to the Lebesgue-Borel measure. One confirms from (3.17) that, for any measurable function g , the equality

$$\int_{\mathbb{R}^n} g(x) d\nu(x) = \int_{\mathbb{R}^n} g(x) \frac{d\nu}{d\beta^n}(x) d\beta^n(x) \quad (3.20)$$

holds whenever the integration exists. In this manner, one may replace the Lebesgue integration of g with respect to the complex measure ν (the l. h. s.) by that with respect to the Lebesgue-Borel measure with the help of the (possibly more familiar notion of) density function $d\nu/d\beta^n$ (the r. h. s.).

Convolution Algebra. The space $L^1(\mathfrak{B}^n)$ of absolutely continuous complex measures is readily shown to be a topologically closed subset (with respect to the topology induced by the total variation norm $\|\cdot\|$ in (3.4)) of the Banach space $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$. This implies that the subspace $L^1(\mathfrak{B}^n)$ is itself a Banach space. One then finds that the linear bijection (3.18) between the two Banach spaces actually defines an *isometric (linear) isomorphism*, which is to say that

$$\|\nu\| = \left\| \frac{d\nu}{d\beta^n} \right\|_1 \quad (3.21)$$

holds for all $\nu \in L^1(\mathfrak{B}^n)$, where the l. h. s. is the total variation norm (3.4) of the complex measure ν and the r. h. s. is the L^1 -norm (2.19) of its density function.

We next see how this bijection plays with convolution. To this end, we first recall that a linear subspace \mathfrak{I} of a commutative algebra \mathfrak{A} is called an *ideal* if it ‘absorbs’ multiplication by elements of \mathfrak{A} , *i.e.*,

$$i \in \mathfrak{I}, a \in \mathfrak{A} \quad \Rightarrow \quad i \cdot a = a \cdot i \in \mathfrak{I}. \quad (3.22)$$

In fact, it is known that the subspace $L^1(\mathfrak{B}^n)$ forms an ideal of the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$, which is to say that

$$\mu \in L^1(\mathfrak{B}^n), \nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n) \quad \Rightarrow \quad \mu * \nu = \nu * \mu \in L^1(\mathfrak{B}^n). \quad (3.23)$$

In passing, the density of the convolution $\mu * \nu$ above is given by the convolution of the density of μ and the complex measure ν as

$$\frac{d(\mu * \nu)}{d\beta^n} = \frac{d\mu}{d\beta^n} * \nu, \quad (3.24)$$

in which we understand the convolution of an integrable function $f \in L^1(\mathbb{R}^n)$ and a complex measure $\mu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ to be

$$(f * \mu)(x) := \int_{\mathbb{R}^n} f(x - y) d\mu(y), \quad (3.25)$$

where the integral is well-defined β^n -a.e. for $x \in \mathbb{R}^n$.

In particular, being an ideal trivially implies that the space $L^1(\mathfrak{B}^n)$ of absolutely continuous complex measures is closed under the operation of convolution, *i.e.*, it forms a *sub-algebra* of the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$. Applying (3.17) to (3.24), one concludes that the density of the convolution of two absolutely continuous complex measures $\mu, \nu \in L^1(\mathfrak{B}^n)$ is given by

the convolution of their densities as

$$\frac{d(\mu * \nu)}{d\beta^n} = \frac{d\mu}{d\beta^n} * \frac{d\nu}{d\beta^n}, \quad (3.26)$$

in which we understand the familiar convolution of two integrable functions $f, g \in L^1(\mathbb{R}^n)$ to be

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x-y)g(y) d\beta^n(y), \quad (3.27)$$

where the integral is well-defined β^n -a.e. for $x \in \mathbb{R}^n$. Equality (3.26) implies that, equipped with the convolution (3.27), the space $L^1(\mathbb{R}^n)$ of integrable functions becomes a Banach algebra that is isomorphically mapped to the sub-algebra $L^1(\mathfrak{B}^n) \subset \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ by the *isometric algebra isomorphism* (3.18). Incidentally, the sub-algebra $L^1(\mathfrak{B}^n) \cong L^1(\mathbb{R}^n)$ of the measure algebra is given its own name, and is occasionally called the *convolution algebra*.

At this point, we note that the convolution algebra $L^1(\mathfrak{B}^n)$ is a proper sub-algebra of the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ in general, *i.e.*, not every complex measure may be represented by integrable functions. This can be readily seen by observing that the delta measure δ_a centred at $a \in \mathbb{R}^n$ (3.13) does not admit a description by density functions. Intuitively, such a density function, if existed, would be given by the ‘delta function’ centred at a , but it is actually a distribution and not a member of $L^1(\mathbb{R}^n)$ as required. This leads to the basic fact that the convolution algebra $L^1(\mathbb{R}^n)$ is non-unital, *i.e.*, it lacks a multiplicative identity in the sense that there is no element $e \in L^1(\mathbb{R}^n)$ for which

$$e * f = f \quad (3.28)$$

holds for all $f \in L^1(\mathbb{R}^n)$. This should be contrasted to the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$, which always possesses a multiplicative identity.

3.1.3. Product Measures. Given two measure spaces (X, \mathfrak{A}, μ) and (Y, \mathfrak{B}, ν) , we intend to construct a ‘product measure’ on the product space $X \times Y$ so that $\rho(A \times B) = \mu(A)\nu(B)$ holds for all $A \in \mathfrak{A}, B \in \mathfrak{B}$. As its domain of definition, we let

$$\mathfrak{A} * \mathfrak{B} := \{A \times B : A \in \mathfrak{A}, B \in \mathfrak{B}\} \quad (3.29)$$

and define

$$\mathfrak{A} \otimes \mathfrak{B} := \sigma(\mathfrak{A} * \mathfrak{B}) \quad (3.30)$$

to be the *product- σ -algebra* of \mathfrak{A} and \mathfrak{B} . The following fact and definition is of importance.

Definition (Product Measure). *Given two measure spaces (X, \mathfrak{A}, μ) and (Y, \mathfrak{B}, ν) , let both μ and ν be σ -finite. Then there exists a unique measure $\mu \otimes \nu : \mathfrak{A} \otimes \mathfrak{B} \rightarrow \overline{\mathbb{R}}$ such that*

$$\mu \otimes \nu(A \times B) = \mu(A)\nu(B), \quad A \in \mathfrak{A}, B \in \mathfrak{B} \quad (3.31)$$

holds. The measure $\mu \otimes \nu$ is σ -finite and is called the product measure of μ and ν .

The integration with respect to the product measure $\mu \otimes \nu$ of two σ -finite measures μ and ν can be performed by iterated integration of each of the respective variables. This is the essence of the following *Fubini’s Theorem*, which belongs to one of the most oft-used theorems of integration theory.

Theorem (Fubini’s Theorem). *Let μ and ν be σ -finite. Then, the following statements hold:*

(i) If $f : X \otimes Y \rightarrow \hat{\mathbb{K}}$ is $\mu \otimes \nu$ -integrable, then $f(x, \cdot)$ is ν -integrable for almost all $x \in X$.

Moreover

$$A := \{x \in X : f(x, \cdot) \text{ is not } \nu\text{-integrable}\} \in \mathfrak{A}; \quad (3.32)$$

and likewise

$$B := \{y \in Y : f(\cdot, y) \text{ is not } \mu\text{-integrable}\} \in \mathfrak{B}. \quad (3.33)$$

The functions

$$x \mapsto \int_Y f(x, y) d\nu(y) \quad x \mapsto \int_X f(x, y) d\mu(x) \quad (3.34)$$

are respectively μ -integrable on A^c and ν -integrable on B^c , and the equalities

$$\begin{aligned} \int_{X \times Y} f d\mu \otimes \nu &= \int_X \left(\int_Y f(x, y) d\nu(y) \right) d\mu(x) \\ &= \int_Y \left(\int_X f(x, y) d\mu(x) \right) d\nu(y) \end{aligned} \quad (3.35)$$

hold.

(ii) If $f : X \otimes Y \rightarrow \hat{\mathbb{K}}$ is $\mu \otimes \nu$ -integrable, and one of the integrals

$$\int_{X \times Y} |f| d\mu \otimes \nu, \quad \int_X \left(\int_Y |f(x, y)| d\nu(y) \right) d\mu(x), \quad \int_Y \left(\int_X |f(x, y)| d\mu(x) \right) d\nu(y) \quad (3.36)$$

is finite, then all three of them are finite and agree, f is $\mu \otimes \nu$ -integrable, and the statements under (i) hold.

3.1.4. Measure on Topological Spaces. Let X be a metric space (or a topological space). One may naturally be interested in how the topology relates to the complex measures defined on the Borel σ -algebra $\mathfrak{B} := \mathfrak{B}(X)$ generated by it. To this end, we briefly review one of the prominent results in the study of this realm, namely the famous *Riesz-Markov-Kakutani Representation Theorem*. In order to avoid complexity, we shall only deal with the case where the given measurable space is $(\mathbb{R}^n, \mathfrak{B}^n)$. Observing now that a complex measure $\nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ generates an (algebraic) linear map $f \mapsto \int_{\mathbb{R}^n} f d\nu$ that maps a function to a complex number, the opposite question is then our interest, namely: what class of linear functionals admits representation by integration with respect to some complex measure?

Riesz-Markov-Kakutani Representation Theorem. Let $C_0(\mathbb{R}^n)$ be the space of all continuous functions $f : \mathbb{R}^n \rightarrow \mathbb{C}$ that vanish at infinity, in the sense for every $\epsilon > 0$ there exists a compact subset $K \subset \mathbb{R}^n$ for which $|f|_{K^c} < \epsilon$ holds. The space $C_0(\mathbb{R}^n)$ equipped with the supremum norm $\|f\|_{\infty} := \sup\{|f(x)| : x \in \mathbb{R}^n\}$ is known to be a Banach space. Now for each $\nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$, the map

$$I_{\nu} : f \mapsto \int_{\mathbb{R}^n} f d\nu, \quad f \in C_0(\mathbb{R}^n) \quad (3.37)$$

gives rise to a continuous (*i.e.*, bounded) \mathbb{C} -linear functional from $C_0(\mathbb{R}^n)$ to \mathbb{C} , for indeed the evaluation

$$|I_{\nu}(f)| \leq \|\nu\| \cdot \|f\|_{\infty} \quad (3.38)$$

holds. The Riesz-Markov-Kakutani representation theorem is a classical theorem in measure and integration theory stating that the converse is also true, which is to say that, for any continuous \mathbb{C} -linear functional $I \in C'_0(\mathbb{R}^n)$, there exists a unique complex measure $\nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$

for which

$$I(f) = \int_{\mathbb{R}^n} f \, d\nu, \quad f \in C_0(\mathbb{R}^n) \quad (3.39)$$

holds. The precise statement is given as follows.

Theorem (Riesz-Markov-Kakutani Representation Theorem for Euclidian Spaces). *The correspondence*

$$\begin{aligned} \Phi : \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n) &\rightarrow C'_0(\mathbb{R}^n), \\ \Phi(\nu)(f) &:= \int_{\mathbb{R}^n} f \, d\nu, \quad \nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n), f \in C_0(\mathbb{R}^n), \end{aligned} \quad (3.40)$$

that maps a complex measure to a continuous linear functional on C_0 is a bijection, which moreover satisfies

$$\|\Phi(\nu)\| = \|\nu\|. \quad (3.41)$$

In other words, the space of complex measures $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ is isomorphic to the topological dual of $C_0(\mathbb{R}^n)$, and can be mapped to each other by an isometric isomorphism.

Here, the norm on $\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ on the r. h. s. of (3.41) is naturally the total variation norm, and the norm on the topological dual $C'_0(\mathbb{R}^n)$ (the l. h. s.) is the operator norm defined by

$$\|I\| := \sup_{\|f\|_{\infty} \leq 1} |I(f)|, \quad I \in C'_0(\mathbb{R}^n). \quad (3.42)$$

In this sense we identify

$$\mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n) \cong C'_0(\mathbb{R}^n), \quad (3.43)$$

and may interchangeably interpret a continuous \mathbb{C} -linear functional on the space $C_0(\mathbb{R}^n)$ as a complex measure on the measurable space $(\mathbb{R}^n, \mathfrak{B}^n)$, and vice versa.

3.1.5. Spectral Theorem and its Consequences. We next provide a concise review on some of the basic facts regarding the spectral theorem for self-adjoint operators, which is just the generalisation of the familiar eigendecomposition theorem for Hermitian matrices on finite-dimensional vector spaces to the arbitrary dimensional case. In order to avoid confusion with operators, Borel sets on \mathbb{R}^n shall occasionally be denoted by $\Delta \in \mathfrak{B}^n$ in place of B , especially when we are working in the context of quantum mechanics.

Spectral Measures. Closely associated to the notion of complex measures is that of spectral measures on a Hilbert space \mathcal{H} . Let $L(\mathcal{H})$ denote the space of all bounded operators on \mathcal{H} , and recall that a map

$$E : \mathfrak{B}^n \rightarrow L(\mathcal{H}), \quad \Delta \mapsto E(\Delta) \quad (3.44)$$

is called an n -dimensional *spectral measure* (or *projection-valued measure*), if each $E(\Delta)$, $\Delta \in \mathfrak{B}^n$ is an orthogonal projection on \mathcal{H} and satisfies

- (i) $E(\emptyset) = 0$, $E(\mathbb{R}^n) = I$,
- (ii) for pairwise disjoint $\Delta_1, \Delta_2, \dots \in \mathfrak{B}^n$,

$$\sum_{i=1}^{\infty} E(\Delta_i)|\phi\rangle = E\left(\bigcup_{i=1}^{\infty} \Delta_i\right)|\phi\rangle, \quad \forall |\phi\rangle \in \mathcal{H}. \quad (3.45)$$

The support of a spectral measure E on \mathfrak{B}^n is defined as the smallest Borel set $\Delta \in \mathfrak{B}^n$ that satisfies $E(\Delta) = I$. An important point is that a spectral measure E and a pair of vectors $|\phi\rangle, |\phi'\rangle \in \mathcal{H}$ induce a complex measure on \mathfrak{B}^n given by

$$\Delta \mapsto \langle \phi', E(\Delta)\phi \rangle, \quad \Delta \in \mathfrak{B}^n. \quad (3.46)$$

Spectral Theorem of Self-adjoint Operators. Having recalled the necessary definitions, we now state the *spectral theorem for self-adjoint operators*, which constitutes one of the most important mathematical ingredients in quantum mechanics.

Theorem (Spectral decomposition of self-adjoint operators). *Let $A : \mathcal{H} \supset \text{dom}(A) \rightarrow \mathcal{H}$ be self-adjoint. Then there exists a unique one-dimensional spectral measure E_A supported on the spectrum $\sigma(A) \subset \mathbb{R}$ of A satisfying*

$$\langle \phi', A\phi \rangle = \int_{\sigma(A)} a \, d\langle \phi', E_A(a)\phi \rangle, \quad \forall |\phi\rangle \in \text{dom}(A), \forall |\phi'\rangle \in \mathcal{H}, \quad (3.47)$$

where the r. h. s. of the equality is understood as the Lebesgue integral with respect to the complex measure $\Delta \mapsto \langle \phi', E_A(\Delta)\phi \rangle$ induced from E_A and the pair of vectors $|\phi\rangle$ and $|\phi'\rangle$.

Under the situation above, the self-adjoint operator A is occasionally written symbolically as

$$A = \int_{\sigma(A)} a \, dE_A(a), \quad (3.48)$$

in terms of integration with respect to its spectral measure.

Finite-dimensional Case. To see the meaning of the above formula, we make a brief note on how the familiar eigendecomposition theorem for Hermitian matrices appears as a special case of the general statement. Let A be a Hermitian matrix on an N -dimensional complex Hilbert space $\mathcal{H} := \mathbb{C}^N$, $N \in \mathbb{N}^\times$. The eigendecomposition theorem states that, there exists an orthonormal basis $\mathcal{B}_A := \{|a_1\rangle, \dots, |a_N\rangle\}$ of \mathcal{H} with real numbers $a_1, \dots, a_N \in \mathbb{R}$ such that

$$A|a_i\rangle = a_i|a_i\rangle, \quad i = 1, \dots, N, \quad (3.49)$$

hold. For each eigenvalue $a \in \sigma(A) = \{a_1, \dots, a_N\}$ of A , we have the projection Π_a onto the subspace,

$$\mathcal{H}_a := \text{span}\{|a\rangle \in \mathcal{B}_A : A|a\rangle = a|a\rangle\} \quad (3.50)$$

spanned by the collection of all eigenvectors associated with a . As we noted before, when the eigenstate $|a\rangle$ is non-degenerate for a , or the subspace \mathcal{H}_a is one-dimensional, we may write $\Pi_a = |a\rangle\langle a|$. With the projection Π_a in hand, the spectral measure of A is defined by

$$E_A(\Delta) := \sum_{a \in \sigma(A) \cap \Delta} \Pi_a, \quad \Delta \in \mathfrak{B}, \quad (3.51)$$

with the convention $\sum_{a \in \emptyset} \Pi_a := 0$. One readily verifies that E_A is indeed a spectral measure supported on its spectrum $\sigma(A)$, and subsequently sees that the projection $\Pi_a = E_A(\{a\})$ is nothing but the image of the spectral measure E_A on the Borel set $\{a\} \in \mathfrak{B}$ consisting of

a single eigenvalue $a \in \sigma(A)$ of the observable A . One then finds

$$A = \sum_{a \in \sigma(A)} a \Pi_a = \sum_{a \in \sigma(A)} a E_A(\{a\}) \quad (3.52)$$

in accordance with (2.76), and subsequently proves

$$\langle \phi', A\phi \rangle = \sum_{a \in \sigma(A)} a \langle \phi', E_A(\{a\})\phi \rangle, \quad \forall |\phi'\rangle, |\phi\rangle \in \mathcal{H}. \quad (3.53)$$

The spectral decomposition formula (3.47) and the formal expression (3.48) are respectively just the generalisations of the finite dimensional versions (3.53) and (3.52).

Functional Calculus. By means of the spectral decomposition of a self-adjoint operator, one may create a new set of operators from it. Let $A : \mathcal{H} \supset \text{dom}(A) \rightarrow \mathcal{H}$ be a self-adjoint operator on a Hilbert space \mathcal{H} , and let E_A the unique one-dimensional spectral measure associated with it. Given a measurable complex function $f : \mathbb{R} \rightarrow \mathbb{C}$, the integral

$$\langle \phi', f(A)\phi \rangle = \int_{\sigma(A)} f(a) d\langle \phi', E_A(a)\phi \rangle \quad (3.54)$$

defines a unique linear operator $f(A)$ on \mathcal{H} , where

$$|\phi\rangle \in \text{dom}(f(A)) := \left\{ |\phi\rangle \in \mathcal{H} : \int_{\sigma(A)} |f(a)|^2 d\langle \phi, E_A(a)\phi \rangle < \infty \right\} \quad (3.55)$$

is any vector belonging to its domain, and $|\phi'\rangle \in \mathcal{H}$. The operator $f(A)$ is occasionally written symbolically as

$$f(A) = \int_{\sigma(A)} f(a) dE_A(a), \quad (3.56)$$

in terms of integration with respect to its spectral measure.

Born Rule and Quantum Measurement. The axiom of quantum mechanics states that a quantum observable is represented by a self-adjoint operator $A : \mathcal{H} \supset \text{dom}(A) \rightarrow \mathcal{H}$ on a Hilbert space \mathcal{H} , and that the probabilistic behaviour of the outcomes of an ideal measurement of A on the state $|\phi\rangle \in \mathcal{H}$ is described by the probability measure,

$$\Delta \mapsto \mu_A^\phi(\Delta) := \frac{\langle \phi, E_A(\Delta)\phi \rangle}{\|\phi\|^2}, \quad \Delta \in \mathfrak{B}. \quad (3.57)$$

Here, the spectral measure E_A is induced from A by the spectral theorem, and the Born rule proclaims that the measurement outcome be given by one of the elements in the spectrum $\sigma(A)$ and that $\mu_A^\phi(\Delta)$ provides the probability of finding the measurement in the measurable set $\Delta \in \mathfrak{B}$. Given $|\phi\rangle \in \text{dom}(A)$, one then realises from the spectral theorem (3.47) that the statistical average of the measurement outcomes of A gives the expectation value,

$$\int_{\mathbb{R}} a d\mu_A^\phi(a) = \frac{\langle \phi, A\phi \rangle}{\|\phi\|^2} =: \mathbb{E}[A; \phi], \quad (3.58)$$

where the l. h. s of the first equality is understood to be the Lebesgue integral with respect to the probability measure (3.57).

3.1.6. *Observables admitting a Description by Density Functions.* While the analysis based on probability measures provides an adequately general framework to work with, we find it useful to prepare a terminology for a special class of observables for which probability density functions, not just probability measures, are available to fully describe the behaviour of the measurement outcomes.

Observable admitting a description by probability density functions. In this paper, we simply say that an observable A admits a description by probability density functions, if the probability measure (3.57) induced by the spectral measure of A is absolutely continuous with respect to the Lebesgue-Borel measure for every choice of the quantum state $|\phi\rangle \in \mathcal{H}$, which is to say that, if for every $|\phi\rangle \in \mathcal{H}$, there exists an integrable function $\rho_A^\phi \in L^1(\mathbb{R})$ such that

$$\mu_A^\phi(\Delta) = \int_\Delta \rho_A^\phi(a) d\beta(a), \quad \Delta \in \mathfrak{B}, \quad (3.59)$$

holds.

A well-known example of it is provided by the one-dimensional position operator \hat{x} on $L^2(\mathbb{R})$ defined in (2.27). Indeed, one proves that the spectral measure of \hat{x} is given by the multiplication of the characteristic function (2.8) as

$$E_{\hat{x}}(\Delta) : \psi(x) \mapsto \chi_\Delta(x)\psi(x), \quad \psi \in L^2(\mathbb{R}), \quad (3.60)$$

for each $B \in \mathfrak{B}$, so that

$$\begin{aligned} \langle \psi_1, E_{\hat{x}}(\Delta)\psi_2 \rangle &= \int_{\mathbb{R}} \psi_1^*(x)\chi_\Delta(x)\psi_2(x) d\beta(x) \\ &= \int_\Delta \psi_1^*(x)\psi_2(x) d\beta(x), \quad \Delta \in \mathfrak{B}, \end{aligned} \quad (3.61)$$

holds. Specifically, this implies that

$$\mu_{\hat{x}}^\psi(\Delta) = \int_\Delta \frac{|\psi(x)|^2}{\|\psi\|_2^2} d\beta(x), \quad \Delta \in \mathfrak{B}, \quad (3.62)$$

where the denominator of the integrand of the r. h. s. denotes the square of the L^2 -norm of $\psi \in L^2(\mathbb{R})$ (see (2.19)). One thus concludes that the density of the probability measure $\mu_{\hat{x}}^\psi$ is provided by

$$\rho_{\hat{x}}^\psi(x) = \frac{|\psi(x)|^2}{\|\psi\|_2^2}. \quad (3.63)$$

Incidentally, it is known that each member of the pair of observables $\{Q, P\}$ that satisfies the Weyl relations (2.39) and (2.40) admits descriptions in terms of density functions.

However, it should be noted that this is not always the case in general: an observable with the spectrum consisting of a finite number of discrete eigenvalues (such as spin) provides a simple counterexample. To see this, let A be such an observable with $N \in \mathbb{N}^\times$ distinct eigenvalues, and let $\sigma(A) = \{a_1, \dots, a_N\}$ be any enumeration of its spectrum. A straightforward application of (3.51) leads to

$$\mu_A^\phi = \sum_{n=1}^N \mu_A^\phi(\{a_n\}) \cdot \delta_{a_n}, \quad (3.64)$$

in which one sees that the probability measure μ_A^ϕ is given by the weighted sum of delta measures centred at each eigenvalue. Obviously, since each of the delta measures is not

absolutely continuous, the resultant probability measure does not admit a description by density functions.

For later use, we also note that, once the observable A admits a description in terms of probability density functions, then the complex measure (3.46) is also absolutely continuous for an arbitrary pair of vectors $|\phi'\rangle, |\phi\rangle \in \mathcal{H}$. That this is the case can be seen by a straightforward application of the polarisation identity

$$\begin{aligned} \langle \phi', T\phi \rangle &= \frac{1}{4} \{ \langle \phi' + \phi, T(\phi' + \phi) \rangle - \langle \phi' - \phi, T(\phi' - \phi) \rangle \\ &\quad + i \langle \phi' + i\phi, T(\phi' + i\phi) \rangle - i \langle \phi' - i\phi, T(\phi' - i\phi) \rangle \} \end{aligned} \quad (3.65)$$

with respect to the operator T valid for any pair of vectors $|\phi\rangle, |\phi'\rangle \in \text{dom}(T)$, where we simply replace $T = E_A(\Delta)$ for each $\Delta \in \mathfrak{B}$.

3.1.7. Simultaneously measurable Observables. For reference, we briefly review the basic mathematical definitions and facts involved in describing measurements of simultaneously measurable observables, including the simultaneous measurement of local observables on the tensor product of Hilbert spaces.

Strong Commutativity of Self-adjoint Operators. Let A and B be self-adjoint operators on a Hilbert space \mathcal{H} , and let E_A and E_B be their respective spectral measures. We say that the pair of operators A and B *strongly commutes*, if

$$E_A(\Delta_A) E_B(\Delta_B) = E_B(\Delta_B) E_A(\Delta_A), \quad \Delta_A, \Delta_B \in \mathfrak{B} \quad (3.66)$$

holds as an operator equality. Note that the strong commutativity of A and B implies its (familiar) commutativity $AB = BA$. On the other hand, it is known that the converse is in general not true in the case where either (or both) of the operators happens to be unbounded. The term strong commutativity is named after this fact, for it indicates a stronger condition than mere commutativity.

Product Spectral Measures. It is a basic result of functional analysis that, given such a pair of A and B of strongly commuting self-adjoint operators, there exists a unique two-dimensional spectral measure $E_{A,B}$ called the *product spectral measure* of A and B , for which

$$E_{A,B}(\Delta_A \times \Delta_B) = E_A(\Delta_A) E_B(\Delta_B) = E_B(\Delta_B) E_A(\Delta_A), \quad \Delta_A, \Delta_B \in \mathfrak{B} \quad (3.67)$$

holds. This is a straightforward operator-valued analogue of product measures in measure theory. With a pair of vectors $|\phi\rangle, |\phi'\rangle \in \mathcal{H}$ being specified, this gives rise to a complex measure on $(\mathbb{R}^2, \mathfrak{B}^2)$, defined by

$$\Delta \mapsto \langle \phi', E_{A,B}(\Delta) \phi \rangle, \quad \Delta \in \mathfrak{B}^2. \quad (3.68)$$

In the context of quantum mechanics, for a given pair of simultaneously measurable quantum observables represented by strongly commuting self-adjoint operators A and B , the probabilistic behaviour of the outcomes of an ideal simultaneous measurement of both the

observables on the state $|\phi\rangle \in \mathcal{H}$ is described by the *joint-probability distribution*

$$\Delta \mapsto \mu_{A,B}^\phi(\Delta) := \frac{\langle \phi, E_{A,B}(\Delta) \phi \rangle}{\|\phi\|^2}, \quad \Delta \in \mathfrak{B}^2, \quad (3.69)$$

of the pair of observables A and B on the state $|\phi\rangle$, which is a two-dimensional probability measure on the measurable space $(\mathbb{R}^2, \mathfrak{B}^2)$. Here, the r. h. s. of (3.69) is interpreted as the probability of finding the outcomes of a simultaneous measurement of both observables in the Borel set $\Delta \in \mathfrak{B}^2$. Note that the measurement outcomes of A and B may not be independent, *i.e.*, the equality,

$$\mu_{A,B}^\phi(\Delta_A \times \Delta_B) = \mu_A^\phi(\Delta_A) \cdot \mu_B^\phi(\Delta_B), \quad \Delta_A, \Delta_B \in \mathfrak{B}, \quad (3.70)$$

may not necessarily hold, or in other words, the joint-probability distribution is not necessarily the product measure $\mu_{A,B}^\phi \neq \mu_A^\phi \otimes \mu_B^\phi$ of each of the respective measurements, in general.

Functional Calculus regarding simultaneously measurable Observables. Given a pair of strongly commuting self-adjoint observables A and B , one readily confirms

$$A = \int_{\mathbb{R}} a \, dE_{A,B}(a, b), \quad B = \int_{\mathbb{R}} b \, dE_{A,B}(a, b). \quad (3.71)$$

As for the sum and product of the observables, we first note the following basic fact.

Lemma 3.1. *Let a pair of self-adjoint operators A and B strongly commute. Then,*

- (i) *The operators A and B commute with each other on $\text{dom}(AB) \cap \text{dom}(BA)$, and the anti-commutator¹⁰*

$$\{A, B\} := AB + BA \quad (3.72)$$

is essentially self-adjoint.

- (ii) *$A + B$ is essentially self-adjoint.*

As a direct consequence, we thus have the operator equalities

$$\overline{A + B} = \int_{\mathbb{R}^2} (a + b) \, dE_{A,B}(a, b) \quad (3.73)$$

$$\frac{\overline{AB + BA}}{2} = \int_{\mathbb{R}^2} ab \, dE_{A,B}(a, b) \quad (3.74)$$

worth of special notice. As above, overlines on closable operators denote their closures, and specifically for essentially self-adjoint operators, their self-adjoint extensions.

Composite Systems. We comment on the special case of the above situation in which the Hilbert space of our interest is the tensor product $\mathcal{H} \otimes \mathcal{K}$ of the target system \mathcal{H} and the meter system \mathcal{K} , and the operators involved are (local) self-adjoint operators A_1 and A_2 on

¹⁰ Here, the domain of the anti-commutator $\{X, Y\} := XY + YX$ of the pair of operators X, Y are understood to be $\text{dom}(\{X, Y\}) := \text{dom}(XY) \cap \text{dom}(YX)$.

the respective Hilbert spaces. Observing that the operators

$$\tilde{A}_1 := A_1 \otimes I, \quad \tilde{A}_2 := I \otimes A_2 \quad (3.75)$$

strongly commute with each other on the composite Hilbert space $\mathcal{H} \otimes \mathcal{K}$, and that their spectral measures respectively read

$$E_{\tilde{A}_1} := E_{A_1} \otimes I, \quad E_{\tilde{A}_2} := I \otimes E_{A_2}, \quad (3.76)$$

the previous argument leads to the existence of a unique two-dimensional product spectral measure $E_{A_1} \otimes E_{A_2} := E_{\tilde{A}_1, \tilde{A}_2}$ satisfying the operator equality

$$\begin{aligned} (E_{A_1} \otimes E_{A_2})(\Delta_1 \times \Delta_2) &= E_{\tilde{A}_1}(\Delta_1)E_{\tilde{A}_2}(\Delta_2) \\ &= (E_{A_1}(\Delta_1) \otimes I)(I \otimes E_{A_2}(\Delta_2)) \\ &= E_{A_1}(\Delta_1) \otimes E_{A_2}(\Delta_2), \quad \Delta_1, \Delta_2 \in \mathfrak{B}. \end{aligned} \quad (3.77)$$

Here, the left-most hand side denotes the two-dimensional spectral measure defined as in (3.67), while the right-most hand side denotes the tensor product of the self-adjoint operators $E_{A_1}(\Delta_1)$ and $E_{A_2}(\Delta_2)$ for each $\Delta_1, \Delta_2 \in \mathfrak{B}$. As we have seen in the previous argument, this gives rise to a complex measure,

$$\Delta \mapsto \langle \Phi', (E_{A_1} \otimes E_{A_2})(\Delta) \Phi \rangle, \quad \Delta \in \mathfrak{B}^2, \quad (3.78)$$

for a given selection of a pair $|\Psi\rangle, |\Phi\rangle \in \mathcal{H} \otimes \mathcal{K}$ of vectors of the composite system, and the map,

$$\Delta \mapsto \mu_{A_1, A_2}^{\Phi}(\Delta) := \frac{\langle \Phi, (E_{A_1} \otimes E_{A_2})(\Delta) \Phi \rangle}{\|\Phi\|^2}, \quad \Delta \in \mathfrak{B}^2, \quad (3.79)$$

(here, we have slightly abused the notation on the l. h. s. by writing A_n in place of \tilde{A}_n for each $n = 1, 2$) provides a probability measure describing the probabilistic behaviour of the outcomes of the ideal local measurements simultaneously performed on each system in the state $|\Phi\rangle \in \mathcal{H} \otimes \mathcal{K}$.

In passing, we note that in the case where the state $|\Phi\rangle$ happens to be a direct product state $|\Phi\rangle = |\phi_1 \otimes \phi_2\rangle$, the induced joint-probability distribution of the two local observables (3.79) becomes the product measure of the two probability measures associated with A_1 and A_2 ,

$$\mu_{A_1, A_2}^{\phi_1 \otimes \phi_2}(\Delta_1 \times \Delta_2) = \mu_{A_1}^{\phi_1}(\Delta_1) \cdot \mu_{A_2}^{\phi_2}(\Delta_2), \quad \Delta_1, \Delta_2 \in \mathfrak{B}, \quad (3.80)$$

indicating that the measurement outcomes of each local measurement A_1 and A_2 are statistically independent (*i.e.*, $\mu_{A_1, A_2}^{\phi_1 \otimes \phi_2} = \mu_{A_1}^{\phi_1} \otimes \mu_{A_2}^{\phi_2}$). On the other hand, if one chooses the state $|\Phi\rangle$ to be an entangled state (*i.e.*, those states in $\mathcal{H} \otimes \mathcal{K}$ that are not direct product states), the joint-probability distribution (3.79) is no more a product measure of those associated to the local observables in general. In the language of physics, this implies that the local measurements performed on each remote system may have some correlation if the state of the composite system happens to be entangled, and this is widely considered to be one of the most intriguing properties of quantum mechanics. Of course, statistical independence between the target and the meter systems is useless for the purpose of our measurement, and we naturally need an entangled state $|\Phi\rangle$ in order to retrieve any meaningful information of the former system out of the measurement of the latter.

Sum and Product of Local Observables. As for the sum and product of a pair of local observables, we note that a direct application of Lemma 3.1 leads to

$$A_1 \otimes I = \int_{\mathbb{R}} a_1 d(E_{A_1} \otimes E_{A_2})(a_1, a_2), \quad I \otimes A_2 = \int_{\mathbb{R}} a_2 d(E_{A_1} \otimes E_{A_2})(a_1, a_2), \quad (3.81)$$

and subsequently

$$\overline{A_1 \otimes I + I \otimes A_2} = \int_{\mathbb{R}^2} (a_1 + a_2) d(E_{A_1} \otimes E_{A_2})(a_1, a_2), \quad (3.82)$$

$$A_1 \otimes A_2 = \int_{\mathbb{R}^2} a_1 a_2 d(E_{A_1} \otimes E_{A_2})(a_1, a_2), \quad (3.83)$$

as expected.

3.2. Unconditioned Measurement

Now that we have recalled the necessary mathematical concepts and results, we shall embark on our main analysis. The target of our analysis is the probability measure describing the behaviour of the outcome of the composite observable $I \otimes X$ on $|\Psi^g\rangle$, which may be rewritten in terms of that of the local observable X on the mixed state ψ^g as

$$\mu_{I \otimes X}^{\Psi^g}(\Delta) = \mu_X^{\psi^g}(\Delta), \quad \Delta \in \mathfrak{B}, \quad (3.84)$$

where the last definition $\mu_X^{\psi^g}(\Delta) := \text{Tr}[E_X(\Delta)\psi^g]/\text{Tr}[\psi^g]$ is merely a straightforward extension of probability measures (3.57) for density operators (for the proof of the equality (3.84), just replace X with $E_X(\Delta)$ in (2.72)).

Main Objective of this Subsection. The primary interest of our study is now to investigate how the information of the target system is encoded into the profile of the outcome of the meter system (3.84) through the interaction. As in the previous subsection, we assume without loss of generality that the meter observable Y coupled with the target observable A to yield the von Neumann interaction (2.70) is given by $Y = P$. The main objective of the passage is to demonstrate the following proposition as an answer to this question. The results, which shall be shortly demonstrated, form the bases we rely on in conducting our further study.

Proposition 3.2 (Unconditioned Measurement II.a). *In the context of the UM scheme, let $Y = P$ be fixed for definiteness, and let $|\phi\rangle \in \mathcal{H}$ and $|\psi\rangle \in \mathcal{K}$ respectively be the initial states of the target and the meter systems. Then, the probability measure (3.84) for both the choice $X = Q, P$ reads*

$$\begin{aligned} \mu_Q^{\psi^g} &= \mu_Q^\psi * \mu_{(gA)}^\phi, \\ \mu_P^{\psi^g} &= \mu_P^\psi, \end{aligned} \quad g \in \mathbb{R}, \quad (3.85)$$

in which the resultant profile of the measurement outcomes of X after the interaction can be exclusively written by the convolution of the initial profiles of both the target and the meter systems.

Specifically, the interaction causes the change only in the profile of the outcome of the observable X conjugate to Y , in which the initial profile of the target system acts upon that of the meter system through convolution of measures. On the other hand, the profile of X

for the same choice as Y is left untouched. The proposition can be readily demonstrated by observing that the change of the spectral measure of the measuring observables $(I \otimes X)$ with respect to the unitary operator $U(g) := e^{-igA \otimes P}$ is provided by

$$U(-g) E_{I \otimes Q}(\Delta) U(g) = E_{\overline{(I \otimes Q + gA \otimes I)}}(\Delta), \quad \Delta \in \mathfrak{B}, \quad (3.86)$$

$$U(-g) E_{I \otimes P}(\Delta) U(g) = E_{(I \otimes P)}(\Delta), \quad \Delta \in \mathfrak{B}, \quad (3.87)$$

in the Heisenberg picture (they are respectively direct consequences of (2.74) and (2.75)), and that the probability distribution dictating the probabilistic behaviour of the sum of two simultaneously measurable observables is described by the convolution of both the individual profiles of the observables involved (which is in parallel to the well-known result for random variables in classical probability theory). However, in the main passages that follow, we intend to provide a more elementary and straightforward demonstration. As a corollary to this, one equivalently has:

Corollary 3.3 (Unconditioned Measurement II.b). *Under the same condition as above, the result (3.85) can also be rewritten as*

$$\begin{aligned} \mu_{(g^{-1}Q)}^{\psi^g} &= \mu_{(g^{-1}Q)}^{\psi} * \mu_A^{\phi}, \\ \mu_{(gP)}^{\psi^g} &= \mu_{(gP)}^{\psi}, \end{aligned} \quad g \in \mathbb{R}^\times, \quad (3.88)$$

by rescaling the outcome by the interaction parameter.

The two different manners (3.85) and (3.88) of describing the effect of the interaction correspond to the two possible ways of combining the interaction parameter g in the unitary group as

$$U(g) := e^{-i(gA) \otimes P} = e^{-iA \otimes (gP)}. \quad (3.89)$$

Combining the interaction parameter g and the target observable A (the former) corresponds to the scaling of the target observable $A \rightarrow gA$, whereas combining g and the meter observable P (the latter) corresponds to the scaling of the pair of the meter observables $\{Q, P\} \rightarrow \{g^{-1}Q, gP\}$. Note that the pair of scaled observables $\{g^{-1}Q, gP\}$ for $g \in \mathbb{R}^\times$ still satisfies the Weyl relations (2.39) and (2.40).

Later on, we shall be investigating how one could recover the information of the target system μ_A^{ϕ} based on the results that we obtained here. Incidentally, one finds that probing either the strong or the weak region of the interaction parameter proves itself useful for this purpose, and the equalities (3.85) and (3.88) shall serve as the respective starting points for analysing the weak and the strong UM schemes.

Preliminary Observation. For our purpose, we first consider the case where the target observable A has a finite point spectrum $\sigma(A) = \{a_1, \dots, a_N\}$, $N \in \mathbb{N}^\times$. Writing the spectral decomposition of A as (2.76) and applying (2.77), one finds that the composite state after the interaction reads

$$\begin{aligned} |\Psi^g\rangle &= \sum_{n=1}^N (\Pi_{a_n} \otimes e^{-iga_n P}) |\phi \otimes \psi\rangle \\ &= \sum_{n=1}^N (|\Pi_{a_n} \phi\rangle \otimes |e^{-iga_n P} \psi\rangle), \quad g \in \mathbb{R}. \end{aligned} \quad (3.90)$$

It then follows that

$$\begin{aligned}
\mu_Q^{\psi^g}(\Delta) &= \frac{\|(I \otimes E_Q(\Delta))\Psi^g\|^2}{\|\Psi^g\|^2} \\
&= \sum_{m=1}^N \sum_{n=1}^N \frac{\langle \phi, \Pi_{a_m} \Pi_{a_n} \phi \rangle}{\|\phi\|^2} \cdot \frac{\langle e^{-ig a_m P} \psi, E_Q(\Delta) e^{-ig a_n P} \psi \rangle}{\|\psi\|^2} \\
&= \sum_{n=1}^N \frac{\|\Pi_{a_n} \phi\|^2}{\|\phi\|^2} \cdot \frac{\|E_Q(\Delta) e^{-ig a_n P} \psi\|^2}{\|\psi\|^2} \\
&= \sum_{n=1}^N \mu_A^\phi(\{a_n\}) \cdot \mu_Q^\psi(\Delta - g a_n) \\
&= \sum_{n=1}^N \mu_A^\phi(\{a_n\}) \cdot \int_{\mathbb{R}} \mu_Q^\psi(\Delta - g a) d\delta_{a_n}(a) \\
&= \int_{\mathbb{R}} \mu_Q^\psi(\Delta - g a) d\mu_A^\phi(a), \quad g \in \mathbb{R}, \quad \Delta \in \mathfrak{B},
\end{aligned} \tag{3.91}$$

where we have used the operator equality¹¹ $e^{itP} E_Q(\Delta) e^{-itP} = E_Q(\Delta - t)$, $\Delta \in \mathfrak{B}$ in the third to last equality, and have applied (3.64) to obtain the last equality.

Description of the Measurement Outcome. Returning to the general case, where the target observable A is now arbitrary, we may conjecture from (3.91) that

$$\mu_Q^{\psi^g}(\Delta) = \int_{\mathbb{R}} \mu_Q^\psi(\Delta - g a) d\mu_A^\phi(a), \quad g \in \mathbb{R}, \quad \Delta \in \mathfrak{B} \tag{3.92}$$

generally holds, which indeed turns out to be true; it can be shown straightforwardly in the general framework of functional analysis and measure and integration theory. From (3.92), we see that the probability measure describing the behaviour of the measurement outcome of Q on the (mixed) state ψ^g after the interaction can be explicitly given by those of the initial states of both the meter and the system. Speaking in an intuitive way, each value $a \in \sigma(A)$ of the spectrum of A causes a translation $\mu_Q^\psi(\Delta) \mapsto \mu_Q^\psi(\Delta - g a)$, $\Delta \in \mathfrak{B}$ to the probability measure of the initial meter state while keeping its ‘shape’ of the profile intact, and each of these effects is all added over, weighted by the original probability μ_A^ϕ of the target observable A .

Parallel to this, we remark that the ideal measurement of the observable $X = P$ after the von Neumann interaction would result in

$$\mu_P^{\psi^g} = \mu_P^\psi, \quad g \in \mathbb{R}, \tag{3.93}$$

which states that the interaction does not alter the profile of the measurement of $X = P$ at all. This can be readily shown by changing Q to P in (3.91), and by applying the operator equality $e^{itP} E_P(\Delta) e^{-itP} = E_P(\Delta)$, $\Delta \in \mathfrak{B}$.

Scaling of Measures and Density Functions. For later arguments, it proves convenient to rewrite our previous result (3.92) in terms of convolution of measures after introducing

¹¹ This is a direct result of (2.51).

some notations. Let $\mu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ be a complex measure, and define a parametrised family $\{\mu_t\}_{t \in \mathbb{R}}$ of complex measures by

$$\mu_t(B) := \begin{cases} \mu(t^{-1}\Delta), & t \in \mathbb{R}^\times, \\ \mu(\mathbb{R}^n) \cdot \delta_0(\Delta), & t = 0, \end{cases} \quad \Delta \in \mathfrak{B}^n. \quad (3.94)$$

Note that this definition is well-defined, for the continuity of the map $x \mapsto tx$ implies its Borel-measurability, hence $t^{-1}\Delta \in \mathfrak{B}^n$ for $\Delta \in \mathfrak{B}^n$. The coefficient $\mu(\mathbb{R}^n)$ multiplied to the delta measure for $t = 0$ is to keep the total evaluation $\mu_t(\mathbb{R}^n) = \mu(\mathbb{R}^n)$ constant for all $t \in \mathbb{R}$. Intuitively speaking, this parametrisation allows us to narrow down the profile of a given complex measure μ while keeping its total evaluation $\mu_t(\mathbb{R}^n) = \mu(\mathbb{R}^n)$ intact, so that it ‘tends’ in an intuitive way to the delta measure (weighted by its total evaluation $\mu(\mathbb{R}^n)$) as $t \rightarrow 0$. To help visualise this, suppose that μ is absolutely continuous and write $\rho := d\mu/d\beta^n$ for simplicity. One then finds

$$\begin{aligned} \mu(t^{-1}\Delta) &= \int_{(t^{-1}\Delta)} \rho(x) d\beta^n(x) \\ &= \int_{\mathbb{R}^n} \chi_\Delta(tx) \rho(x) d\beta^n(x) \\ &= \int_{\mathbb{R}^n} \chi_\Delta(x) \cdot |t|^{-n} \rho\left(\frac{x}{t}\right) d\beta^n(x), \\ &= \int_{\Delta} \rho_t(x) d\beta^n(x), \quad \Delta \in \mathfrak{B}^n, t \in \mathbb{R}^\times, \end{aligned} \quad (3.95)$$

where we have introduced the scaling

$$f_t(x) := |t|^{-n} f\left(\frac{x}{t}\right), \quad t \in \mathbb{R}^\times, \quad (3.96)$$

of any given integrable function $f \in L^1(\mathbb{R}^n)$ by $t \in \mathbb{R}^\times$. This implies that μ_t is also absolutely continuous for each $t \in \mathbb{R}^\times$ by definition, and that its density is given by ρ_t , *i.e.*,

$$\frac{d\mu_t}{d\beta^n} = \left(\frac{d\mu}{d\beta^n}\right)_t, \quad t \in \mathbb{R}^\times, \quad (3.97)$$

where the l. h. s. is the density of the scaled probability measure μ_t , and the r. h. s. is the density of the original probability measure μ scaled by t as in (3.96). In the special case where μ is a probability measure, one may intuitively see that the parametrisation (3.96) takes any non-negative integrable function with the total integral of unity (*i.e.*, a probability density function) to the ‘delta function’ in the limit $t \rightarrow 0$.

Von Neumann Interaction and Convolution. Now, note here that for each $t \in \mathbb{R}^\times$, the probability measure μ_t is nothing but the image measure (3.9) of μ with respect to the map $x \mapsto tx$ (*i.e.*, multiplication by t). With the help of the change of variables formula for image measures (3.10), one confirms that the equality

$$\int_{\mathbb{R}^n} f(x) d\mu_t(x) = \int_{\mathbb{R}^n} f(tx) d\mu(x), \quad t \in \mathbb{R} \quad (3.98)$$

holds for all f that is integrable with respect to μ . This allows us to rewrite (3.92) in terms of convolution as

$$\mu_Q^{\psi_g} = \mu_Q^\psi * \left(\mu_A^\phi\right)_g, \quad g \in \mathbb{R}. \quad (3.99)$$

Alternatively, by scaling $\Delta \rightarrow g\Delta$ in (3.92), one finds from the definition that

$$\left(\mu_Q^{\psi^g}\right)_{g^{-1}} = \left(\mu_Q^\psi\right)_{g^{-1}} * \mu_A^\phi, \quad g \in \mathbb{R}^\times, \quad (3.100)$$

which is another way to describe how the von Neumann type interaction causes a change in the profile of the meter observable $X = Q$.

Scaling of Observables. We make a short digression at this point to seek for the physical meaning of the two findings (3.99) and (3.100), which we have just acquired. To prepare for our argument, we first introduce some notations regarding scaling of spectral measures, in parallel to that of complex measures as we have done before. Let $E : \mathfrak{B}^n \rightarrow L(\mathcal{H})$ be an n -dimensional spectral measure on the Hilbert space \mathcal{H} , and define a parametrised family $\{E_t\}_{t \in \mathbb{R}}$ of spectral measures by

$$E_t(\Delta) := \begin{cases} E(t^{-1}\Delta), & t \in \mathbb{R}^\times, \\ E_0(\Delta), & t = 0, \end{cases} \quad \Delta \in \mathfrak{B}^n. \quad (3.101)$$

Here, we have introduced the ‘delta spectral measure’ E_0 centred at $0 \in \mathbb{R}^n$, defined by

$$E_0(\Delta) := \begin{cases} I, & 0 \in \Delta, \\ 0, & 0 \notin \Delta, \end{cases} \quad \Delta \in \mathfrak{B}^n. \quad (3.102)$$

Incidentally, for the one-dimensional case ($n = 1$), the delta spectral measure E_0 centred at the origin is nothing but the spectral measure accompanying the zero operator 0 on \mathcal{H} .

We next confirm some basic facts regarding scaling of observables and their accompanying spectral measures. Let E_A be the spectral measure of a self-adjoint operator $A : \mathcal{H} \supset \text{dom}(A) \rightarrow \mathcal{H}$. The goal is to specify the spectral measure of the scaled self-adjoint operator tA , ($t \in \mathbb{R}$) and to show that

$$E_{(tA)} = (E_A)_t, \quad t \in \mathbb{R}, \quad (3.103)$$

where the l. h. s. is the desired spectral measure accompanying the scaled operator tA , whereas the r. h. s. is the spectral measure accompanying the operator A scaled by t . To see this, first observe the following equality

$$\begin{aligned} \langle \phi, (tA)\phi \rangle &= \int_{\mathbb{R}} ta \, d\mu_A^\phi(a) \\ &= \int_{\mathbb{R}} a \, d(\mu_A^\phi)_t(a) \\ &= \int_{\mathbb{R}} a \, d\langle \phi, (E_A)_t(a)\phi \rangle, \quad t \in \mathbb{R} \end{aligned} \quad (3.104)$$

for the choice $|\phi\rangle \in \text{dom}(A)$, where we have used (3.98) to obtain the second to last equality. Applying the polarisation identity (3.65) for $T = tA$, one then has

$$\langle \phi', (tA)\phi \rangle = \int_{\mathbb{R}} a \, d\langle \phi', (E_A)_t(a)\phi \rangle, \quad t \in \mathbb{R} \quad (3.105)$$

for any $|\phi\rangle, |\phi'\rangle \in \text{dom}(A)$. Observing that the domain of a self-adjoint operator is dense in \mathcal{H} by definition, one may continuously extend the above equality on $|\phi'\rangle \in \mathcal{H}$, based on which the uniqueness of the spectral measure leads to the desired result (3.103).

Returning to our main line of arguments, we first observe that the equality (3.103) leads to

$$\mu_{(tA)}^\phi = \left(\mu_A^\phi \right)_t, \quad t \in \mathbb{R}, \quad (3.106)$$

which states that the probability measure describing the ideal measurement outcome of the scaled observable tA on the state $|\phi\rangle \in \mathcal{H}$ coincides with that of the original observable A scaled by t . Armed with this result, one may reformulate our previous findings (3.99) and (3.100) respectively as

$$\begin{cases} \mu_Q^{\psi^g} = \mu_Q^\psi * \mu_{(gA)}^\phi, \\ \mu_P^{\psi^g} = \mu_P^\psi, \end{cases} \quad g \in \mathbb{R}, \quad (3.107)$$

and

$$\begin{cases} \mu_{(g^{-1}Q)}^{\psi^g} = \mu_{(g^{-1}Q)}^\psi * \mu_A^\phi, \\ \mu_{(gP)}^{\psi^g} = \mu_{(gP)}^\psi, \end{cases} \quad g \in \mathbb{R}^\times, \quad (3.108)$$

where we have also explicitly written down the profile of the outcome of the measurement of $X = P$. This completes our proof for Proposition 3.2 and Corollary 3.3.

3.3. Recovery of the Target Profile

We now consider the inverse problem of what we have discussed so far, that is, we argue how one can recover the probability measure μ_A^ϕ of the target observable A from the probability measure $\mu_Q^{\psi^g}$ obtained through the measurement of Q on the meter system. Following the same line in the previous section, one finds it useful to probe either the strong or the weak region of the interaction for this purpose, which we shall see below one by one.

3.3.1. Strong Unconditioned Measurement. We first concentrate on (3.88) (or equivalently (3.100)), and observe that the problem of recovering the desired probability measure reduces to the problem of ‘deconvolution’, where one wishes to find the solution $\mu := \mu_A^\phi$ of the equation of the form

$$\nu_{\text{out}} = \nu_{\text{in}} * \mu, \quad (3.109)$$

having knowledge and control over both the ‘input’ $\nu_{\text{in}} := \mu_{(g^{-1}Q)}^\psi$ and ‘output’ $\nu_{\text{out}} := \mu_{(g^{-1}Q)}^{\psi^g}$ on their respective sides. Whilst there is rich literature on the topic of deconvolution, we take a specific approach to the solution in order to make our arguments simple.

Main Objective of this Passage. A quick observation leads us to a naïve expectation that, if one could attune the input so that ν_{in} may become a multiplicative identity (in our case, it is the delta measure δ_0 centred at the origin), or in the case where this is impossible, if one gradually approximates the input close enough to it, then, one may obtain the desired solution μ directly as the measured output $\nu_{\text{out}} \rightarrow \delta_0 * \mu = \mu$. One of the typical manners in which we attain such gradual approximation would be to fix the initial state ψ and taking the strong limit $g^{-1} \rightarrow 0$ ($g \rightarrow \pm\infty$) of the interaction parameter, so that $\nu_{\text{in}} = (\mu_Q^\psi)_{g^{-1}}$ ‘tends’ towards the desired identity δ_0 in an intuitive manner (recall (3.94) and (3.96)). The main objective of this passage is to confirm that this idea is indeed valid, and thus to state it in a mathematically rigorous way.

As it becomes apparent through the line of discussions below, there are some certain mathematical hurdles that must be overcome to achieve this objective. In order to avoid much intricacies, we shall impose certain condition to the choice of the target observable, and present our main result in the following way:

Proposition 3.4 (Strong Unconditioned Measurement). *In the context of the UM scheme, suppose that*

- (i) *the target observable A admits description by density functions,*
- (ii) *the initial profile μ_Q^ψ of the meter observable Q on the state $|\psi\rangle$ is compactly supported¹².*

Then, the scaled profile of Q after the interaction converges to the desired target in the strong limit of interaction

$$\lim_{g \rightarrow \pm\infty} \left\| \mu_{(g^{-1}Q)}^{\psi^g} - \mu_A^\phi \right\| = 0 \quad (3.110)$$

with respect to the total variation norm (or, equivalently the L^1 -norm) for any choice of the initial states $|\phi\rangle \in \mathcal{H}$.

The remainder of this passage is devoted to its demonstration.

Preliminary Observations. Let us make a preliminary observation following the above idea. The first thing we realise is that, in general, we cannot prepare the input ν_{in} so that its profile may exactly coincide with the multiplicative identity δ_0 . To see this quickly, first recall that the realisable input probability measures ν_{in} are exactly those that are absolutely continuous with respect to the Lebesgue-Borel measure. Since the delta measure δ_0 does not belong to the space $L^1(\mathfrak{B})$, one concludes that it is impossible to prepare the input in such a way that $\nu_{\text{in}} = \delta_0$ holds. An alternative approach to this problem may be to consider a sequence of inputs $(\nu_{\text{in}})_n$ that tends to the delta measure δ_0 in hope that the resultant sequence of multiplicative products $(\nu_{\text{out}})_n := (\nu_{\text{in}})_n * \mu$ also converges towards the desired solution μ in the limit. Indeed, if one could only construct a sequence $(\nu_{\text{in}})_n$ so that

$$\lim_{n \rightarrow \infty} \|(\nu_{\text{in}})_n - \delta_0\| = 0, \quad (3.111)$$

under the total variation norm, one concludes from the evaluation

$$\|(\nu_{\text{in}})_n * \mu - \mu\| = \|((\nu_{\text{in}})_n - \delta_0) * \mu\| \leq \|(\nu_{\text{in}})_n - \delta_0\| \cdot \|\mu\| \quad (3.112)$$

that the outcome tends to the desired solution

$$\lim_{n \rightarrow \infty} \|(\nu_{\text{out}})_n - \mu\| = 0 \quad (3.113)$$

in the limit. Unfortunately, however, one immediately realises that this idea also fails, since in general there is no such sequence $(\nu_{\text{in}})_n$ that meets the condition (3.111) in the first place, for indeed, since the space $L^1(\mathfrak{B})$ of absolutely continuous complex measures is a topologically closed subset of the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$, a sequence in $L^1(\mathfrak{B})$ never converges to an element outside of $L^1(\mathfrak{B})$ with respect to the total variation norm.

¹² We say that a complex measure ν has a compact support if there exists a compact subset $K \subset \mathbb{R}$ for which the restriction of the variation $|\nu|$ on the complement $|K^c = 0$ is a zero measure.

Discussion on the possible Approaches. From the quick overview of our current situation, we learn that the problem at hand is to do with the *topology* we have given to the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$. Namely, the topology induced from the total variation norm is too strong (fine) for our convenience. A fundamental cure for this would thus be to equip the space with a weaker (coarser) topology on $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$ such that, at least, it may allow us to construct sufficiently abundant sequences (or nets, in general) of the ‘inputs’ in $L^1(\mathfrak{B})$ that converges towards δ_0 , and that the sequence of the resulting ‘outputs’ (*i.e.*, the multiplicative product (3.109)) would subsequently converge towards the desired solution in the limit¹³.

However, since this strategy, while being desirable, presupposes moderate familiarity with the mathematical branch of general topology, which the authors have deemed to be beyond the scope of this paper, an alternative approach to the problem without explicit exposure to it would be favourable (possibly at the cost of generality, while hopefully having the merit of being mathematically less demanding). In this paper, this would be accomplished by introducing an auxiliary concept of ‘approximate identities’, whose definition would be shortly presented. In essence, we focus only on the convergence of the output in the total variation norm, based on the observation that, even though there is *no* sequence of the input that converges to the delta measure (3.111), there are certain conditions in which the sequence of the output *do* converge towards the desired solution (3.113). As a preliminary observation to this approach, observe that the output ν_{out} also necessarily lies in $L^1(\mathfrak{B})$ ¹⁴, and by recalling that $L^1(\mathfrak{B})$ is closed under the topology induced by the total variation norm, one finds that the candidates of the solution μ towards which the sequence of outputs could ever converge are only those that also lie in $L^1(\mathfrak{B})$. Based on this inspection, in what follows, we shall only treat the case in which *the target observable A admits a description by density functions*, which is to say that the solutions $\mu = \mu_A^\phi$ are always guaranteed to lie in $L^1(\mathfrak{B})$, is assumed.

Approximate Identities. The convolution algebra $L^1(\mathfrak{B}) \cong L^1(\mathbb{R}^n)$, contrasted to the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$, is non-unital. In order to compensate the inconvenience arising from the lack of a multiplicative identity, a weaker concept is often used in analysing problems involving algebras. In this paper, we call a family $\{e_t\}_{t>0}$ of elements of $L^1(\mathbb{R}^n)$ an *approximate identity*, if for every element $f \in L^1(\mathbb{R}^n)$, the convolution $e_t * f$ converges to f

¹³ A straightforward candidate for such a topology would be the weak-* topology based on the identification (3.43) by the Riesz-Markov-Kakutani representation theorem, namely, the initial topology with respect to the family of all algebraic linear functionals of the form $\mu \mapsto \int_{\mathbb{R}} f d\mu$, where $f \in C_0(\mathbb{R})$. One eventually finds that the norm topology of the total variation is nothing but the strong topology with respect to the identification, which implies that the weak-* topology is strictly weaker than the topology we currently have at hand. Moreover, direct application of the dominated convergence theorem and Fubini’s theorem reveals that the convergence of a sequence of probability measures $\nu_n \rightarrow \delta_0$ implies $\nu_n * \mu \rightarrow \mu$ (both the convergence is meant in weak-*), which is a much cleaner result than what we have seen in the main paragraphs. As an example of such a sequence (net) of probability measures converging towards δ_0 , one finds that the scaling ν_t (3.94) of a given probability measure ν is typical. In fact, the scaling becomes a continuous parametrisation from \mathbb{R} to $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$ under the topology, which is also a welcome property.

¹⁴ To see this, recall that the output can be written as a multiplicative product of two probability measures with one of which being absolutely continuous, and that the space $L^1(\mathfrak{B})$ of absolutely continuous complex measures is an ideal in $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$.

in the topology induced by the L^1 -norm, *i.e.*,

$$\lim_{t \rightarrow 0} \|e_t * f - f\|_1 = 0, \quad f \in L^1(\mathbb{R}^n). \quad (3.114)$$

Before we move on to the construction of an example, we collect some necessary terminologies. Recall that the support of a function $f : \mathbb{R}^n \rightarrow \mathbb{K}$ is a subset of \mathbb{R}^n defined by

$$\text{supp}(f) := \overline{\{x \in \mathbb{R}^n : f(x) \neq 0\}}, \quad (3.115)$$

where the overline on a set denotes its topological closure. A support of a function $f : \mathbb{R}^n \rightarrow \mathbb{K}$ is said to be compact if $\text{supp}(f)$ is bounded. Now, let $\eta \in L^1(\mathbb{R}^n)$ be any integrable function possessing a compact support with the total integration of unity,

$$\int_{\mathbb{R}^n} \eta(x) dx = 1. \quad (3.116)$$

With this, consider a family $\{\eta_t\}_{t \in \mathbb{R}^\times}$ of scaled functions defined as in (3.96), which preserve the total integration of unity for all $t \in \mathbb{R}^\times$. One may then intuitively expect that η_t tends to the ‘delta function’ in the limit $t \rightarrow 0$ and can be used for an approximate identity,

$$\lim_{t \rightarrow 0} \|\eta_t * f - f\|_1 = 0, \quad (3.117)$$

for all $f \in L^1(\mathbb{R}^n)$. To confirm that this is indeed the case, observe the inequality

$$\begin{aligned} \|\eta_t * f - f\|_1 &:= \int_{\mathbb{R}^n} \left| \left(\int_{\mathbb{R}^n} \eta_t(y) f(x-y) d\beta^n(y) \right) - f(x) \right| d\beta^n(x) \\ &= \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \eta_t(y) (f(x-y) - f(x)) d\beta^n(y) \right| d\beta^n(x) \\ &\leq \int_{\mathbb{R}^n} |\eta(y)| \left(\int_{\mathbb{R}^n} |f(x-ty) - f(x)| d\beta^n(x) \right) d\beta^n(y) \\ &= \int_{\mathbb{R}^n} |\eta(y)| \cdot \|\tau_{(-ty)} f - f\|_1 d\beta^n(y), \end{aligned} \quad (3.118)$$

where τ_a is the translation operator defined by

$$\tau_a f(x) := f(x+a). \quad (3.119)$$

Recalling that $\lim_{a \rightarrow 0} \|\tau_a f - f\|_1 = 0$ for any $f \in L^1(\mathbb{R}^n)$, we see that for any $\epsilon > 0$, there exists a $\delta > 0$ for which $a \in K_\delta(0) := \{x \in \mathbb{R}^n : |x| < \delta\}$ leads to $\|\tau_a f - f\|_1 < \epsilon$. By taking $|t|$ small enough so that $\text{supp}(\eta) \subset K_{t^{-1}\delta}(0)$, we find that the r. h. s of the above inequality is less than ϵ . This shows that the family defined by

$$e_t := \eta_{(\pm t)}, \quad t > 0, \quad (3.120)$$

makes a simple example of approximate identities (here, the meaning of the subscript on both sides of the equation is not to be confused, where the subscript on the l. h. s. indicates an index of the elements of the convolution algebra $L^1(\mathbb{R}^n)$, whereas that on the r. h. s. indicates the scaling parameter of an integrable function η defined in (3.96)). Obviously, the construction of such approximate identities is highly non-unique, and one may attain it in various different ways.

Realisation of Approximate Identities. Our observation so far revealed that, as long as the target profile $\mu \in L^1(\mathfrak{B})$ is absolutely continuous, by considering the family of inputs $\{(\nu_{\text{in}})_t\}_{t>0}$ in such a way that it makes an approximate identity in $L^1(\mathfrak{B})$, the resulting family of outputs $(\nu_{\text{out}})_t := (\nu_{\text{in}})_t * \mu$ would successfully converge to the desired solution

$$\lim_{t \rightarrow 0} \|(\nu_{\text{out}})_t - \mu\| = 0 \quad (3.121)$$

in the L^1 -norm (or equivalently, in the total variation norm)¹⁵. We are now interested in the construction of such approximate identities for our current situation. To this, we first observe that, since the profile of the input $\nu_{\text{in}} = \mu_{(g^{-1}Q)}^\psi$ in our case is exclusively determined by the choice of the interaction parameter g and the initial state $|\psi\rangle \in \mathcal{K}$ of the meter system, the problem reduces to finding a sequence of the pair $(g, |\psi\rangle)_t$, $t > 0$ that makes the input an approximate identity. As an example of such a construction, we first fix the initial state $|\psi\rangle$ and observe that the density of the input is given by

$$\frac{d\mu_{(g^{-1}Q)}^\psi}{d\beta} = \left(\frac{d\mu_Q^\psi}{d\beta} \right)_{g^{-1}}, \quad g \in \mathbb{R}^\times, \quad (3.122)$$

where we have used our previous result (3.97). Then, choosing $|\psi\rangle$ so that the density of μ_Q^ψ may be compactly supported, one realises that taking the strong limit of the interaction $g^{-1} \rightarrow 0$ (or equivalently $g \rightarrow \pm\infty$) yields the desired result. In turn, we fix the interaction parameter $g \in \mathbb{R}^\times$ and choose a sequence of initial states that makes the corresponding probability measures an approximate identity. Since the scaling of an approximate identity by g^{-1} is still an approximate identity, one achieves another example of such a construction.

One thus finds a general guiding principle for the construction of an approximate identity to be the combination of the two manoeuvres, namely, either

- by taking the strong limit of the interaction $g^{-1} \rightarrow 0$,
- by narrowing down the profile of the probability measures to the delta measure (symbolically $\mu_Q^\psi \rightarrow \delta_0$) by changing the meter state $|\psi\rangle \in \mathcal{K}$.

In order to explicitly see how these work together, choose a sequence of initial states $|\psi\rangle$, $|\psi_h\rangle \in \mathcal{K}$, $h > 0$, such that the density of the initial profile μ_Q^ψ is compactly supported and that the parametrisation corresponds to its scaling

$$\frac{d\mu_Q^{\psi_h}}{d\beta} = \left(\frac{d\mu_Q^\psi}{d\beta} \right)_h, \quad (3.123)$$

which makes itself an approximate identity as $h \rightarrow 0$ (one may easily construct such a sequence in the special case in which the meter system is described in the Schrödinger

¹⁵ We note again that the subscripts t used here is meant to be an index, and not to be confused with that denoting scaling of complex measures.

representation of the CCR¹⁶). Then, observing that the scaling of it by g^{-1} is

$$\left(\frac{d\mu_Q^{\psi_h}}{d\beta}\right)_{g^{-1}} = \left(\frac{d\mu_Q^\psi}{d\beta}\right)_{hg^{-1}}, \quad (3.125)$$

one finds that it is indeed an approximate identity that tends to the delta in the limit as $hg^{-1} \rightarrow 0$ together.

Concluding Remarks. In conclusion, we see that the UM scheme allows us to recover the information of the target system and its observable A , not only in the form of expectation values described earlier, but also in the form of probability measures μ_A^ϕ . This is accomplished by taking the limit of either narrowing the profile of the probability measure μ_Q^ψ of the meter system, or intensifying the interaction parameter $g \rightarrow \pm\infty$, or otherwise by appropriately balancing both contributions and having $hg^{-1} \rightarrow 0$ as a whole. In this sense, we may say that intensifying the interaction parameter has an equivalent role to narrowing the profile of the probability measure of the meter. It may thus appear reasonable that, also in this respect, the von Neumann measurement scheme is sometimes referred to as the ‘strong measurement’ or the ‘sharp measurement’.

3.3.2. Weak Unconditioned Measurement. We shall see next how the measurement outcome of the UM scheme behaves locally around $g = 0$ in terms of probability measures. Specifically, we are interested in the (higher-order) derivatives of the map

$$\mathbb{R} \rightarrow L^1(\mathfrak{B}), \quad g \mapsto \mu_Q^{\psi^g}, \quad (3.126)$$

which is now a map from the real line \mathbb{R} to the space of complex measures $L^1(\mathfrak{B}) \subset \mathbf{M}_{\mathbb{C}}(\mathfrak{B})$.

Main Objective of this Passage. The main objective of this passage is to first compute the derivatives of the map (3.126) at the origin $g = 0$, and subsequently argue how one may reconstruct the profile of the probability measure μ_A^ϕ of our interest from the information obtained. However, as one realises in the line of discussion that follows, this involves certain mathematical intricacies. In order to avoid any difficulties and complication that may arise, we impose some restrictions to the configuration of the target and meter systems, and thus obtain the following two propositions, the first of which shall be demonstrated in the main passages below.

Proposition 3.5 (Outcome of the Weak Unconditioned Measurement). *In the context of the UM scheme, suppose that*

- (i) *the target profile μ_A^ϕ is compactly supported,*
- (ii) *the density of μ_Q^ψ belongs to the Schwartz space $d\mu_Q^\psi/d\beta \in \mathcal{S}(\mathbb{R})$.*

¹⁶ One may choose any wave-function $\psi \in L^2(\mathbb{R})$ with compact support, and define

$$\psi_{(h)}(x) := |h|^{-1/2} \psi\left(\frac{x}{h}\right). \quad (3.124)$$

Here, the braces among the subscript h to denote the index is merely employed in order to avoid confusion with that denoting scaling of a function (3.96). One then readily finds that this qualifies as an example of the desired family (3.123).

Then, the map (3.126) is arbitrarily many times strongly differentiable in the L^1 -norm (or, equivalently, in the total variation norm), and its derivatives at $g = 0$ reads

$$\left. \frac{d^n}{dg^n} \mu_Q^{\psi^g} \right|_{g=0} = \mathbb{E}[A^n; \phi] \cdot (-D)^n \mu_Q^\psi, \quad n \in \mathbb{N}_0, \quad (3.127)$$

where D denotes the operation uniquely specified through the relation

$$d(D\nu)/d\beta := D(d\nu/d\beta), \quad (3.128)$$

by differentiating the density of absolutely continuous complex measures $\nu \in L^1(\mathfrak{B})$ whose density $d\nu/d\beta \in \mathcal{S}(\mathbb{R})$ lies in the Schwartz space.

Note that compactness of the support of μ_A^ϕ implies the existence of all the higher-order moments $|\mathbb{E}[A^n; \phi]| < \infty$ of the observable A , and that the Schwartz space is closed under the operation of differentiation (i.e., $D^n(d\nu/d\beta) \in \mathcal{S}(\mathbb{R})$), hence both sides of (3.127) is well-defined. Operationally, the above proposition implies that one may obtain not only the expectation value ($n = 1$) of μ_A^ϕ , as we have found by the operator level analysis (2.89) conducted in the previous section, but also its higher-order moments

$$\mathbb{E}[A^n; \phi] = \int_{\mathbb{R}} a^n d\mu_A^\phi(a), \quad n \in \mathbb{N}_0, \quad (3.129)$$

by probing the local behaviour of the interaction around $g = 0$. Incidentally, one might expect that one could recover the full profile of the original probability measure μ_A^ϕ by knowing enough numbers of its higher-order moments, which in fact turns out to be positive under our assumption.

Proposition 3.6 (Weak Unconditioned Measurement). *Let A be self-adjoint and $|\phi\rangle \in \mathcal{H}$ for which the probability measure μ_A^ϕ is compactly supported. Given another compactly supported probability measure μ on $(\mathbb{R}, \mathfrak{B})$ such that all their higher moments*

$$\mathbb{E}[A^n; \phi] = \int_{\mathbb{R}} a^n d\mu(a), \quad n \in \mathbb{N}_0, \quad (3.130)$$

coincide with those of μ_A^ϕ , then the two probability measures agree $\mu = \mu_A^\phi$. In other words, one may uniquely reconstruct the probability measure μ_A^ϕ of the target system by knowing all the higher moments of A by means of the weak UM.

Proof. In fact, this is one instance of the famous problems collectively called the classical *moment problem* [31, 32]. We provide a sketch of the proof for our specific case at hand, and to this, we first observe that knowing all the higher-order moments (3.129) is equivalent to knowing the integral $\int p(a) d\mu_A^\phi(a)$ of all polynomials $p \in P(K)$ on some compact subset $K \subset \mathbb{R}$ on which μ_A^ϕ is supported. Now, choose a compact subset $K \subset \mathbb{R}$ that contains the support of both μ_A^ϕ and μ , i.e., $\mu|_{K^c} = \mu_A^\phi|_{K^c} = 0$, and observe that the space of continuous functions on K trivially coincide with that of continuous functions on K that vanishes at infinity $C(K) = C_0(K)$. We thus have $C(K)' = C_0(K)' \cong \mathbf{M}_{\mathbb{C}}(\mathfrak{B}|_K)$ by the Riesz-Markov-Kakutani representation theorem. Since the space of polynomials $P(K)$ is dense in $C(K)$ with respect to the supremum norm (cf. Stone-Weierstraß approximation theorem), one concludes that $\int_{\mathbb{R}} p(a) d\mu(a) = \int p(a) d\mu_A^\phi(a)$, $p \in P(K)$ implies $\mu = \mu_A^\phi$. ■

Preliminary Observation. We now begin our analysis. To provide some preliminary observation to this problem, we start by observing that the target of our study would be the following formal expression

$$\left. \frac{d}{dg} \mu_Q^{\psi^g} \right|_{g=0} := \lim_{g \rightarrow 0} \frac{\mu_Q^{\psi^g} - \mu_Q^\psi}{g}, \quad (3.131)$$

in which we leave aside, just for now, all the inherent subtleties that will shortly become apparent regarding the operation of taking the limit. Now, since the numerator of the r. h. s. of the above formula can be written as

$$\mu_Q^{\psi^g} - \mu_Q^\psi = \mu_Q^\psi * \left(\left(\mu_A^\phi \right)_g - \delta_0 \right), \quad (3.132)$$

one finds that the analysis of (3.131) reduces to the study of the formal expression of the form

$$\nu'_{\text{out}}(0) := \left. \frac{d}{dt} \nu_{\text{out}}(t) \right|_{t=0} = \lim_{t \rightarrow 0} \nu_{\text{in}} * \frac{\mu_t - \delta_0}{t}, \quad (3.133)$$

where $\mu, \nu_{\text{in}} \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B})$ are probability measures (the latter being absolutely continuous), $\nu_{\text{out}}(t) := \nu_{\text{in}} * \mu_t$, and the subscript on μ_t denotes the scaling defined in (3.94). In studying (3.133), one might find it a decent starting point to focus on the formal expression (the right component of the above convolution)

$$\lim_{t \rightarrow 0} \frac{\mu_t - \delta_0}{t} =: \mu'_0. \quad (3.134)$$

From this, one realises that our problem is nothing but the differentiability of the map $t \mapsto \mu_t$ at the origin $t = 0$ (recall that we have defined $\mu_0 := \delta_0$ for any probability measure μ), and thus have symbolically written the limit of the above expression by μ'_0 , temporarily leaving aside the question of its existence and well-definedness just as before. It would then be tempting to expect

$$\nu'_{\text{out}}(0) = \nu_{\text{in}} * \mu'_0, \quad (3.135)$$

which should resolve our main problem fairly nicely.

A Formal Computation of the Derivative. Guided by the above naïve observation, we are naturally led to consider what the derivative of the map $t \rightarrow \mu_t$ at $t = 0$ for a given probability measure $\mu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B})$ would look like. As a first step, suppose for simplicity that μ is absolutely continuous, and denote its density by $\eta := d\mu/d\beta$. Armed with our previous findings $\eta_t = d\mu_t/d\beta$, $t \in \mathbb{R}^\times$ regarding scaling of measures and that of its densities (see (3.97)), we then intend to formally obtain

$$\mu'_0 = \lim_{t \rightarrow 0} \mu'_t \quad (3.136)$$

in view of density functions, by first computing its derivative at $t > 0$ and then taking the limit $t \rightarrow 0$. Now, assuming suitable differentiability and integrability conditions for the

density η , one computes the derivative of the map $t \mapsto \eta_t$ at $t > 0$ as

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\eta_{t+h} - \eta_t}{h} &= -\frac{1}{t^2} \eta\left(\frac{x}{t}\right) - \frac{x}{t^3} (D\eta)\left(\frac{x}{t}\right) \\ &= -D\left(\frac{1}{t} \frac{x}{t} \eta\left(\frac{x}{t}\right)\right) \\ &= -D(x\eta)_t, \quad t > 0, \end{aligned} \tag{3.137}$$

where $D := d/dx$ was the usual operation of differentiation. Then, one might be tempted to formally proceed as

$$\begin{aligned} \lim_{t \rightarrow 0} D(x\eta)_t &= D\left[\lim_{t \rightarrow 0} (x\eta)_t\right] \\ &= D\left[\left(\int_{\mathbb{R}} x\eta \, d\beta\right) \cdot \delta_0\right] \\ &= \mathbb{E}[x; \mu] \cdot D\delta_0, \end{aligned} \tag{3.138}$$

where we have used (3.94) in the second equality. The above argument implies that the derivative of the map $t \rightarrow \mu_t$ at the origin would appear as

$$\mu'_0 = \mathbb{E}[x; \mu] \cdot (-D)\delta_0, \tag{3.139}$$

which is the ‘derivative of the delta measure’ weighted by the expectation value of the original probability measure μ . As for the general case in which the original probability measure μ is now not necessarily absolutely continuous, we may conjecture that, since the r. h. s. of (3.139) does not depend on the absolute continuity of the original probability measure μ , the same result should hold even in the general case as well.

Discussion on the possible Approaches. While we have conducted a very formal discussion above, the result in fact turns out to be true and can be made mathematically fully rigorous in the framework of the *theory of generalised functions (distributions)*. In fact, it turns out that the derivative μ'_0 that appears in (3.139) is no longer a member of the space $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$ of complex measures¹⁷, and accordingly the framework in which we have been working so far (*i.e.*, the space of complex measures) is insufficient for our analysis. For further study of the weak UM scheme, a preferable approach would thus be to expand our framework by introducing the space of distributions. While this method has a great merit in being able

¹⁷Incidentally, one may recall that the (higher-order) derivatives of the delta distribution appears in several branches of physics, one of the most familiar of which being presumably the theory of electromagnetism. The derivative of the delta distribution $D\delta_0$ is among the most well-known example of a distribution that cannot be expressed by a complex measure. In order to provide an intuitive reasoning with the tools at hand, let φ be a smooth function with compact support (*i.e.*, a test function) satisfying $(D\varphi)(0) = 1$. As a concrete example, one may take $\varphi(x) := x\varphi_0(x)$ with

$$\varphi_0(x) := \begin{cases} e^{-\frac{1}{1-x^2}} & (|x| < 1) \\ 0 & (|x| \geq 1). \end{cases} \tag{3.140}$$

Defining a sequence of test functions by $\varphi_n(x) := n^{-1}\varphi(nx)$, $n \in \mathbb{N}^\times$, observe that the dominated convergence theorem necessarily implies $\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \varphi_n d\mu = 0$ for any complex measure $\mu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B})$. On the other hand, with the help of an auxiliary smooth density function ρ to symbolically express the delta distribution by the limit of its scaling $\delta_0 = \lim_{t \rightarrow 0} \rho_t$, one may formally compute the integral

to conduct our analysis with decent generality (and in fact, distributions have their role, not just in this subsection, but also later in studying the quasi-joint-probability distributions in Section 5 and 6), at the same time, it has a drawback in that it would be rather mathematically demanding, especially since the theory of distributions is build up on the results of general topology.

In view of this, an alternative approach to the problem without direct exposure to the theory of distributions would be favourable. To this end, recalling the idea employed in the previous subsection, we concentrate only on the differentiability of the multiplicative product (3.133), setting aside the intricacies involving that of the map $t \mapsto \mu_t$ we have seen above. To see what we mean, we first expect, by combining (3.135) and (3.139), that the derivative of the map $t \mapsto \nu_{\text{out}}(t)$ at the origin be written as

$$\nu'_{\text{out}}(0) = \mathbb{E}[x; \mu] \cdot (\nu_{\text{in}} * (-D)\delta_0). \quad (3.142)$$

Now, assuming suitable differentiability condition of the density $\rho_{\text{in}} := d\nu_{\text{in}}/d\beta$ of the input ν_{in} as a starting point, we employ an auxiliary smooth density function η to symbolically express the delta distribution by the limit of its scaling $\delta_0 = \lim_{t \rightarrow 0} \eta_t$ (a similar technique is used in (3.141)) and formally obtain the ‘density’ of the convolution $\nu_{\text{in}} * D\delta_0$ as

$$\begin{aligned} (\rho_{\text{in}} * D\delta_0)(x) &= \left(\rho_{\text{in}} * D \left(\lim_{t \rightarrow 0} \eta_t \right) \right)(x) \\ &= \lim_{t \rightarrow 0} \int_{\mathbb{R}} \rho_{\text{in}}(x-y) (D\eta_t)(y) d\beta(y) \\ &= \lim_{t \rightarrow 0} \int_{\mathbb{R}} (D\rho_{\text{in}})(x-y) \eta_t(y) d\beta(y) \\ &= (D\rho_{\text{in}})(x). \end{aligned} \quad (3.143)$$

Introducing the notation $D\nu_{\text{in}}$ as defined in (3.128), we thus obtain

$$\nu'_{\text{out}}(0) = \mathbb{E}[x; \mu] \cdot (-D)\nu_{\text{in}}. \quad (3.144)$$

The basic idea is that, while we have seen that the distributional derivative of the delta $D\delta_0$ does not allow itself to be expressed by a complex measure, the distributional derivative $D\nu_{\text{in}}$ of some probability measure ν_{in} might belong to the space $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$ of complex measures¹⁸.

of φ_n weighted by the ‘density’ $D\delta_0$ as

$$\begin{aligned} \int_{\mathbb{R}} \varphi_n(x) (D\delta_0)(x) d\beta(x) &= \lim_{t \rightarrow 0} \left(\int_{\mathbb{R}} \varphi_n(x) (D\rho_t)(x) d\beta(x) \right) \\ &= \lim_{t \rightarrow 0} \left(- \int_{\mathbb{R}} (D\varphi_n)(x) \rho_t(x) d\beta(x) \right) \\ &= - \int_{\mathbb{R}} (D\varphi_n)(x) \delta_0(x) d\beta(x), \end{aligned} \quad (3.141)$$

where we have used integration by parts to obtain the second equality. This implies $\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \varphi_n(D\delta_0) d\beta = \lim_{n \rightarrow \infty} -(D\varphi_n)(0) = -1$, which would lead to a contradiction if $(D\delta_0)$ were to be expressed by a complex measure.

¹⁸ As one may expect, the distributional derivative $D\nu$ of an arbitrary complex measure ν can be made well-defined by extending our framework into the theory of generalised functions. In general, the derivative derivative $D\nu$ is a distribution itself (as we have seen for the special case $\nu = \delta_0$), but not necessarily a complex measure anymore.

If we could moreover find a condition for which the differentiability (3.144) is valid with respect to the norm topology of the total variation (*i.e.*, strongly differentiable), we could develop a line of argument that is totally confined in the space $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$, without referring to the theory of distributions at all.

On the Main Results. One finds below that the the above idea is indeed valid. To this end, we assume

- The probability measure μ has compact support.
- The density of ν_{in} belongs to the Schwartz space $d\nu_{\text{in}}/d\beta \in \mathcal{S}(\mathbb{R})$.

Under the above two conditions, we demonstrate below that the map $t \mapsto \nu_{\text{out}}(t)$ is in fact arbitrarily many times strongly differentiable, and that its higher-order derivatives read

$$\nu_{\text{out}}^{(n)}(t) = ((-D)^n \nu_{\text{in}}) * (x^n \odot \mu)_t, \quad t \in \mathbb{R}, \quad n \in \mathbb{N}_0, \quad (3.145)$$

which in particular implies

$$\nu_{\text{out}}^{(n)}(0) = \mathbb{E}[x^n; \mu] \cdot (-D)^n \nu_{\text{in}}, \quad n \in \mathbb{N}_0 \quad (3.146)$$

at the origin $t = 0$. Here, $D^n \nu_{\text{in}}$ denotes the signed measure defined in (3.128), and the signed measure $x^n \odot \mu$ is defined in (3.7). Note that our two conditions above, namely, the compactness of the support of $\mu = x^0 \odot \mu$ and the density of $\nu_{\text{in}} = (-D)^0 \nu_{\text{in}}$ belonging to the Schwartz space, are true not only for $n = 0$, but for all $n \in \mathbb{N}_0$. Note also that compactness of the support of μ guarantees the finiteness of all its higher-order moments $|\mathbb{E}[x^n; \mu]| < \infty$, $n \in \mathbb{N}_0$. Applying (3.146) to our physical situation by letting $\mu = \mu_A^\phi$ and $\nu_{\text{in}} = \mu_Q^\psi$ would prove Proposition 3.5.

Proof of our Main Result. For demonstration, we provide a sketch of the proof by mathematical induction. One may readily confirm by definition that the above statement is trivially true for $n = 0$. Now, assuming that the statement is true for $n \in \mathbb{N}_0$, we rewrite $\tilde{\nu}_{\text{in}} := (-D)^n \nu_{\text{in}}$, $\tilde{\mu} := x^n \odot \mu$ and $\tilde{\nu}_{\text{out}}(t) := \tilde{\nu}_{\text{in}} * \tilde{\mu}_t$ for better readability. Now, recalling that the convolution algebra $L^1(\mathfrak{B})$ is an ideal in the measure algebra $\mathbf{M}_{\mathbb{C}}(\mathfrak{B})$, one finds that $\tilde{\nu}_{\text{out}}(t)$ is absolutely continuous for all $t \in \mathbb{R}$ (in passing, one moreover finds that the density of $\tilde{\nu}_{\text{out}}(t)$ is also a Schwartz function), and that its density $\tilde{\rho}_{\text{out}}(t) := d\tilde{\nu}_{\text{out}}(t)/d\beta$ is given by

$$\tilde{\rho}_{\text{out}}(t)(x) = \int_{\mathbb{R}} \tilde{\rho}_{\text{in}}(x - ty) d\tilde{\mu}(y), \quad t \in \mathbb{R}, \quad (3.147)$$

where $\tilde{\rho}_{\text{in}}$ denotes the density of $\tilde{\nu}_{\text{in}}$ (see (3.24) for this result).

In order to prove the strong differentiability of the map $t \mapsto \tilde{\nu}_{\text{out}}(t)$, we work in the space of density functions. We start by demonstrating the point-wise differentiability of the map $t \mapsto \tilde{\rho}_{\text{out}}(t)$, and to this end, we fix $t_0, x \in \mathbb{R}$ and observe

$$\begin{aligned} (\tilde{\rho}'_{\text{out}}(t_0))(x) &:= \lim_{t \rightarrow t_0} \frac{\tilde{\rho}_{\text{out}}(t)(x) - \tilde{\rho}_{\text{out}}(t_0)(x)}{t - t_0} \\ &= \lim_{t \rightarrow 0} \int_{\mathbb{R}} \frac{\tilde{\rho}_{\text{in}}(x - ty) - \tilde{\rho}_{\text{in}}(x - t_0 y)}{t - t_0} d\tilde{\mu}(y) \\ &= \int_{\mathbb{R}} y(-D\tilde{\rho}_{\text{in}})(x - t_0 y) d\tilde{\mu}(y), \end{aligned} \quad (3.148)$$

where the exchange of the limit and integration in the second equality, while we shall omit any details of its proof, is essentially a consequence of the dominated convergence theorem.

Next, we return to its strong differentiability (*i.e.*, differentiability with respect to the L^1 -norm). To this end, we assume $t_0 < t$ without loss of generality and recall the mean-value theorem, which state that there exists a $t_1 \in]t_0, t[$ such that

$$\frac{\tilde{\rho}_{\text{out}}(t)(x) - \tilde{\rho}_{\text{out}}(t_0)(x)}{t - t_0} = (\tilde{\rho}'_{\text{out}}(t_1))(x) \quad (3.149)$$

holds. Then, one has

$$\begin{aligned} & \left\| \frac{\tilde{\rho}_{\text{out}}(t)(x) - \tilde{\rho}_{\text{out}}(t_0)(x)}{t - t_0} - \tilde{\rho}'_{\text{out}}(t_0) \right\|_1 \\ &= \int_{\mathbb{R}} \left| \int_{\mathbb{R}} y(-D\tilde{\rho}_{\text{in}})(x - t_1y) - y(-D\tilde{\rho}_{\text{in}})(x - t_0y) d\tilde{\mu}(y) \right| d\beta(x) \\ &\leq \int_{\mathbb{R}} |y| \cdot \|\tau_{(-t_1y)}(D\tilde{\rho}_{\text{in}}) - \tau_{(-t_0y)}(D\tilde{\rho}_{\text{in}})\|_1 d\tilde{\mu}(y), \end{aligned} \quad (3.150)$$

where the exchange of the order of integration in the last inequality is guaranteed to hold (Fubini's theorem), and the translation operator τ_a is defined in (3.119). Compactness of the support of $\tilde{\mu}$ together with an analogous argument made in (3.117) implies that the r. h. s. of the above inequality tends to 0 as $t \rightarrow 0$, which completes our proof for strong differentiability. We thus have by (3.148)

$$\begin{aligned} \rho_{\text{out}}^{(n+1)}(t) &= \tilde{\rho}'_{\text{out}}(0) \\ &= (-D\tilde{\rho}_{\text{in}}) * (x \odot \tilde{\mu})_t \\ &= ((-D)^{n+1}\rho_{\text{in}}) * (x^{n+1} \odot \mu)_t, \quad t \in \mathbb{R} \end{aligned} \quad (3.151)$$

and

$$\begin{aligned} \rho_{\text{out}}^{(n+1)}(0) &= ((-D)^{n+1}\rho_{\text{in}}) * (x^{n+1} \odot \mu)_0 \\ &= \mathbb{E}[x^{n+1}; \mu] \cdot (-D)^{n+1}\nu_{\text{in}}, \end{aligned} \quad (3.152)$$

where we have used (3.94) and $(x^{n+1} \odot \mu)(\mathbb{R}) = \mathbb{E}[x^{n+1}; \mu]$ in the last equality. This completes our whole proof. \blacksquare

4. Conditioned Measurement I: In Terms of Conditional Expectations

We shall next embark on our study of the measurement scheme that we call the *conditioned measurement* (CM) scheme. As the name indicates, the CM scheme involves *conditioning*, where one employs the measurement of another observable on the target system on top of the UM scheme studied earlier. The CM scheme can be understood as a natural generalisation of the *post-selected measurement scheme*, which has recently been attracting much attention of several groups among the physics community. While the post-selected measurement scheme itself has been practiced for quite a while, it has caught a renewed interest since Aharonov *et al.* reintroduced it with the term *weak measurement* which in particular applies to the post-selected measurement in the weak limit, along with the complex quantity termed *weak value* purported to be measured by it. Two sections starting from here is devoted to the analysis on the CM scheme, and by following the same line as that of the former unconditioned counterpart, we start by examining the measurement scheme in terms of conditional expectations (Section 4), and subsequently in terms of conditional probabilities (Section 5).

Organisation of this Section. The contents of this section is organised as follows. We first provide a concise summary of some of the necessary mathematical concepts that provides us the tools for conducting the analysis. We then make a brief review on the CM scheme from a relatively general framework, and make some comments on the technique of employing conditioning (or post-selection, as a special case) in precision measurements, whose alleged advantages has recently become the topic of intensive debate. We shall then investigate how one could reclaim the information of the configuration of the target system from the the measured outcomes, and to this end, we concentrate on the behaviour of the conditional expectation of the meter observable around the weak limit $g = 0$ of the interaction parameter. In parallel to the unconditional case, we call this procedure the *weak conditioned measurement scheme* in this paper. We finally close this section by introducing the concept of *conditional quasi-expectations* of a quantum observable given another (not necessarily simultaneously measurable) observable, as a generalisation to that of the standard conditional expectations, and examine some of their notable properties.

4.1. Reference Materials

In this subsection, we shall briefly recall the necessary mathematical definitions and results regarding the formal mathematical description of conditioning.

4.1.1. Conditioning. The essence of the CM scheme lies in the conditioning of the outcomes of a measurement of an observable X of the meter system \mathcal{K} by that of an additional observable B of the target system \mathcal{H} . The quantity of interest is then the *conditional expectation* of X given B , in contrast to the UM scheme described in Section 2, where the quantity of interest was the mere (unconditional) expectation value of X .

Conditional Expectation given a Sub- σ -algebra. Since one may find the general definition of conditional expectations to be rather involved, we start by some preliminary discussion in order to ease the introduction. Let $(\mathbb{R}^n, \mathfrak{B}^n, \mu)$ be a probability space, and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be μ -integrable. Given a Borel set $B \in \mathfrak{B}^n$ with non-vanishing probability $\mu(B) \neq 0$, one

defines the *conditional expectation of f given the measurable set $B \in \mathfrak{B}^n$* by the real number

$$\mathbb{E}[f|B] := \frac{\int_B f(x) d\mu(x)}{\mu(B)}. \quad (4.1)$$

Now, let $\mathbb{R}^n = \cup_{i=1}^N B_i$, $B_i \in \mathfrak{B}^n$ be a decomposition of \mathbb{R}^n into finite numbers of mutually disjoint Borel sets, and let $\mathfrak{E} := \{B_i\}_{i=1, \dots, N}$ denote their collection. We then define

$$\mathfrak{A} := \sigma(\mathfrak{E}) = \left\{ \bigcup_{i \in I} B_i : I \subset \{1, \dots, N\} \right\} \quad (4.2)$$

to be the sub- σ -algebra of \mathfrak{B}^n generated by \mathfrak{E} . Assuming $\mu(B_i) \neq 0$ for all $i = 1, \dots, N$, this gives rise to an \mathfrak{A} - \mathfrak{B} measurable function

$$\mathbb{E}[f|\mathfrak{A}](x) := \sum_{i=1}^N \mathbb{E}[f|B_i] \cdot \chi_{B_i}(x), \quad (4.3)$$

where each χ_{B_i} is the characteristic function of the subset B_i . Observing that each element $A \in \mathfrak{A}$ can be expressed by a union of elements of \mathfrak{E} , one has

$$\begin{aligned} \int_A f(x) d\mu(x) &= \sum_{B_i \subset A} \mathbb{E}[f|B_i] \cdot \mu(B_i) \\ &= \int_A \mathbb{E}[f|\mathfrak{A}](x) d\mu|_{\mathfrak{A}}(x), \quad \forall A \in \mathfrak{A}, \end{aligned} \quad (4.4)$$

where $\mu|_{\mathfrak{A}}$ denotes the restriction of the probability measure μ on the sub- σ -algebra \mathfrak{A} . Guided by this observation, the conditional expectation of an integrable function f given a sub- σ -algebra \mathfrak{A} is defined in the following manner:

Definition (Conditional expectation given a sub- σ -algebra). *Let $(\mathbb{R}^n, \mathfrak{B}^n, \mu)$ be a probability space. For a sub- σ -algebra $\mathfrak{A} \subset \mathfrak{B}^n$ and a μ -integrable function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, the conditional expectation of f given \mathfrak{A} , denoted as $\mathbb{E}[f|\mathfrak{A}]$, is defined as a $\mu|_{\mathfrak{A}}$ -integrable function satisfying*

$$\int_A f(x) d\mu(x) = \int_A \mathbb{E}[f|\mathfrak{A}](x) d\mu|_{\mathfrak{A}}(x), \quad \forall A \in \mathfrak{A}. \quad (4.5)$$

The conditional expectation $\mathbb{E}[f|\mathfrak{A}]$ exists, and is unique $\mu|_{\mathfrak{A}}$ -a.e.

To see the validity of the definition, first observe that the l. h. s. of (4.5) defines a complex measure $A \mapsto (f \odot \mu)(A)$, $A \in \mathfrak{A}$. Since $(f \odot \mu)|_{\mathfrak{A}} \ll \mu|_{\mathfrak{A}}$, the Radon-Nikodým theorem leads to the existence and uniqueness $\mu|_{\mathfrak{A}}$ -a.e. of the conditional expectation

$$\mathbb{E}[f|\mathfrak{A}] := \frac{d(f \odot \mu)|_{\mathfrak{A}}}{d\mu|_{\mathfrak{A}}}, \quad (4.6)$$

which is nothing but the Radon-Nikodým derivative (density) of the restriction $(f \odot \mu)|_{\mathfrak{A}}$ with respect to the restriction $\mu|_{\mathfrak{A}}$. Note that the conditional expectation is defined as a *function* (or more precisely, an equivalent class of functions) rather than a mere number. The elementary definition (4.3) mentioned earlier is in fact a special case of the above general definition, in which the sub- σ -algebra concerned is given by (4.2). The conditional expectation $\mathbb{E}[f|\mathfrak{A}]$ serves as the, so to speak, best approximation of the original function f by measurable functions defined on the coarser¹⁹ σ -algebra $\mathfrak{A} \subset \mathfrak{B}^n$.

¹⁹ Given two σ -algebras $\mathfrak{A} \subset \mathfrak{B}$, \mathfrak{A} is said to be smaller or *coarser* than \mathfrak{B} , and on the other hand, \mathfrak{B} is said to be larger or *finer* than \mathfrak{A} .

Conditional Expectation given another Function. We next recall the definition of the conditional expectation given another real measurable function. As above, we first provide an introductory argument. Let $(\mathbb{R}^n, \mathfrak{B}^n, \mu)$ be a probability space, and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be μ -integrable. Given another measurable function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, suppose that the probability of obtaining the outcome $y \in \mathbb{R}$ of g is non-vanishing $\mu(g^{-1}(y)) \neq 0$. In a similar manner as before, one may define the conditional expectation of f given the outcome y of g as

$$\mathbb{E}[f|g = y] := \mathbb{E}[f|g^{-1}(y)] = \frac{\int_{g^{-1}(y)} f(x) d\mu(x)}{\mu(g^{-1}(y))}, \quad (4.7)$$

where we have just replaced $B = g^{-1}(y)$ in (4.1). It is now tempting to construct a function $y \mapsto \mathbb{E}[f|g = y]$ that maps each of the possible outcomes of g to the corresponding conditional expectation. Assuming that the function g only takes a finite number of distinct outcomes $\{y_i\}_{i=1, \dots, N}$, $y_i \in \mathbb{R}$, one accordingly obtains a decomposition $\mathbb{R}^n = \cup_{i=1}^N g^{-1}(y_i)$ of \mathbb{R}^n into a finite number of mutually disjoint Borel sets. Assuming moreover that $\mu(g^{-1}(y_i)) \neq 0$ for all i , one obtains a well-defined measurable function

$$\mathbb{E}[f|g] : \mathbb{R} \rightarrow \mathbb{R}, \quad y \mapsto \mathbb{E}[f|g = y], \quad (4.8)$$

called the *conditional expectation of f given g* .

To see how this relates to the previous definition of the conditional expectation given a sub- σ -algebra, consider a general situation in which one is given a set X (without a σ -algebra), a measurable space (Y, \mathfrak{A}) and a function $g : X \rightarrow Y$. The collection

$$\mathcal{I}(g) := g^{-1}(\mathfrak{A}) := \{g^{-1}(A) : A \in \mathfrak{A}\} \quad (4.9)$$

makes itself into a σ -algebra, called the *initial σ -algebra* on X with respect to g , and it is the coarsest σ -algebra on \mathbb{R}^n for which the map g is measurable. In the above situation, we take $(Y, \mathfrak{A}) = (\mathbb{R}, \mathfrak{B}^1)$ and define

$$\mathcal{I}(g) := g^{-1}(\mathfrak{B}^1) = \sigma(\mathfrak{E}), \quad (4.10)$$

where we have let $\mathfrak{E} := \{g^{-1}(y_i)\}_{i=1, \dots, N}$. Now, since we have assumed that $\mu(g^{-1}(y_i)) \neq 0$ for all i , the conditional expectation of f given $\mathcal{I}(g)$ can be expressed as

$$\mathbb{E}[f|\mathcal{I}(g)] = \mathbb{E}[f|\sigma(\mathfrak{E})] = \sum_{i=1}^N \mathbb{E}[f|g^{-1}(y_i)] \cdot \chi_{g^{-1}(y_i)}, \quad (4.11)$$

where the last equality is due to (4.3) by replacing $B_i = g^{-1}(y_i)$. It is then fairly straightforward to see that the conditional expectations $\mathbb{E}[f|\mathcal{I}(g)]$, $\mathbb{E}[f|g]$ and the conditioning function g are related to one another through the commutative diagram,

$$\begin{array}{ccc} (\mathbb{R}^n, \mathcal{I}(g)) & \xrightarrow{g} & (\mathbb{R}, \mathfrak{B}^1) \\ & \searrow \mathbb{E}[f|\mathcal{I}(g)] & \downarrow \mathbb{E}[f|g] \\ & & (\mathbb{R}, \mathfrak{B}^1) \end{array} \quad (4.12)$$

where each of the functions is measurable. In this sense, the function $\mathbb{E}[f|g]$ is understood to be nothing but the factorisation of $\mathbb{E}[f|\mathcal{I}(g)]$ by g . The validity of such observation for the general case is guaranteed by the following Factorisation Theorem.

Theorem (Factorisation Theorem). *Let X be a non-empty set, and let $\mathcal{I}(g) := g^{-1}(\mathfrak{A})$ be the initial σ -algebra of a map $g : X \rightarrow (Y, \mathfrak{B})$. A function $h : (X, \mathcal{I}(g)) \rightarrow (\mathbb{R}, \mathfrak{B}^1)$ is measurable if and only if there exists a measurable function $\tilde{h} : (Y, \mathfrak{B}) \rightarrow (\mathbb{R}, \mathfrak{B}^1)$ that makes the diagram*

$$\begin{array}{ccc} (X, \mathcal{I}(g)) & \xrightarrow{g} & (Y, \mathfrak{B}) \\ & \searrow h & \downarrow \tilde{h} \\ & & (\mathbb{R}, \mathfrak{B}^1) \end{array} \quad (4.13)$$

commute.

By letting $(Y, \mathfrak{B}) = (\mathbb{R}, \mathfrak{B}^1)$ and $h = \mathbb{E}[f|\mathcal{I}(g)]$, this guarantees the existence of the function $\mathbb{E}[f|g] := \tilde{h}$ that makes the desired diagram commute, even for the general case.

As for the integrability of the conditional expectation $\mathbb{E}[f|g]$, we first observe that the probability of obtaining the outcome of g in a Borel set $B \in \mathfrak{B}^1$ is dictated by the probability measure

$$g(\mu)(B) := \mu(g^{-1}(B)), \quad B \in \mathfrak{B}^1, \quad (4.14)$$

which is nothing but the image measure of μ with respect to g (see (3.9) for its definition and properties). One thus sees by the formula

$$\begin{aligned} \int_{\mathbb{R}} \mathbb{E}[f|g] dg(\mu) &= \sum_{i=1}^N \mathbb{E}[f|g](y_i) \cdot g(\mu)(\{y_i\}) \\ &= \sum_{i=1}^N \mathbb{E}[f|g = y_i] \cdot \mu(g^{-1}(y_i)) \\ &= \sum_{i=1}^N \int_{g^{-1}(y_i)} f(x) d\mu(x) \\ &= \int_{\mathbb{R}^n} f(x) d\mu(x), \end{aligned} \quad (4.15)$$

that the function $\mathbb{E}[f|g]$ is $g(\mu)$ -integrable, and its expectation value coincides with the expectation value of f under μ , which is what one naturally expects.

Guided by the above observation, the conditional expectation of an integrable function f given another measurable function g is defined in the following manner:

Definition (Conditional expectation given a measurable function). *Let $(\mathbb{R}^n, \mathfrak{B}^n, \mu)$ be a probability space, and let $f : \mathbb{R}^n \rightarrow \mathbb{R}$ be μ -integrable. The conditional expectation of f given a measurable function $g : \mathbb{R}^n \rightarrow \mathbb{R}$, denoted as $\mathbb{E}[f|g]$, is defined as a $g(\mu)$ -integrable function that makes the diagram*

$$\begin{array}{ccc} (\mathbb{R}^n, \mathcal{I}(g), \mu|_{\mathcal{I}(g)}) & \xrightarrow{g} & (\mathbb{R}, \mathfrak{B}^1, g(\mu)) \\ & \searrow \mathbb{E}[f|\mathcal{I}(g)] & \downarrow \mathbb{E}[f|g] \\ & & (\mathbb{R}, \mathfrak{B}^1) \end{array} \quad (4.16)$$

commute. Its existence and uniqueness $g(\mu)$ -a.e. is known to be guaranteed.

Note that integrability of $\mathbb{E}[f|\mathcal{I}(g)]$ is due to the change of variables formula (3.10) for image measures, and its uniqueness $g(\mu)$ -a.e. is immediate by definition. Based on the above definition, let $\mathbb{E}[f|g]$ be (a representative of) the conditional expectation of f given g . We write

$$\mathbb{E}[f|g = y] := \mathbb{E}[f|g](y) \quad (4.17)$$

to denote the *conditional expectation of f given the outcome y of g* . Note that this definition is dependent on the choice of the representative and may admit ambiguity. Indeed, for the choice $y \in \mathbb{R}$ for which the probability of obtaining the outcome of g in $\{y\}$ is vanishing: $g(\mu)(\{y\}) = \mu(g^{-1}(\{y\})) = 0$, one sees that $\mathbb{E}[f|g = y]$ is *indefinite* and may take *any* real number. As exemplified in here, the conditional expectation $\mathbb{E}[f|g]$ of f given g is appropriate to be viewed as an equivalent class of integrable functions, rather than a function alone.

Conditioning by Simultaneously Measurable Observables. As in the previous section, we occasionally denote the Borel sets on \mathbb{R}^n by $\Delta \in \mathfrak{B}^n$ in place of B for better understanding and readability, especially in the context of quantum theory, where the confusion of the notation of B with that of an operator may become a concern. Let A and B be a pair of simultaneously measurable observables on a quantum system \mathcal{H} . We have seen that this yields a probability measure $\mu_{A,B}^\phi$ on $(\mathbb{R}^2, \mathfrak{B}^2)$ (cf. (3.69)), which is interpreted as the joint-probability distribution describing the outcomes of a simultaneous measurement of A and B performed on the quantum system in the state $|\phi\rangle \in \mathcal{H}$. Letting $f(a, b) = \pi_A(a, b) := a$ and $g(a, b) = \pi_B(a, b) := b$ describe the measurement outcomes of each of the observables A and B , we shall briefly see below how the previous discussions on conditioning fits in the context of quantum mechanics. For our purpose, assume $|\phi\rangle \in \text{dom}(A)$ so that the projection $\pi_A(a, b) = a$ may be integrable

$$\begin{aligned} \int_{\mathbb{R}^2} \pi_A(a, b) d\mu_{A,B}^\phi(a, b) &= \int_{\mathbb{R}^2} a d\mu_A^\phi(a) \\ &= \mathbb{E}[A; \phi], \end{aligned} \quad (4.18)$$

with respect to the probability measure $\mu_{A,B}^\phi$. Observing that the image measure of $\mu_{A,B}^\phi$ with respect to the second projection

$$\pi_B \left(\mu_{A,B}^\phi \right) (\Delta_B) := \mu_{A,B}^\phi(\mathbb{R} \times \Delta_B) = \mu_B^\phi(\Delta), \quad \Delta_B \in \mathfrak{B} \quad (4.19)$$

is nothing but the probability measure describing the outcome of B , we define the *conditional expectation* $\mathbb{E}[A|B; \phi]$ of an observable A given B on the state $|\phi\rangle$ as the (equivalence class of) μ_B^ϕ -integrable function(s)

$$\mathbb{E}[A|B; \phi] := \mathbb{E}[\pi_A|\pi_B], \quad (4.20)$$

where the r. h. s. is the conditional expectation of π_A given π_B under the probability measure $\mu_{A,B}^\phi$. Under the same assumption, we analogously define the conditional expectation of an observable A given the outcome b of an observable B on the state $|\phi\rangle \in \text{dom}(A)$ by

$$\mathbb{E}[A|B = b; \phi] := \mathbb{E}[A|B; \phi](b). \quad (4.21)$$

We note again that the last definition incorporates some ambiguity, in which the number $\mathbb{E}[A|B = b; \phi]$ is not well-defined in the case where the probability that the measurement of B yields the outcome b is vanishing.

4.2. Conditioned Measurement

The CM scheme incorporates the measurements of *two* observables, where the experimenter measures one local observable on the meter system and the other on the target system. In this paper, we generally define the CM scheme as the act of measuring the conditional expectation

$$\mathbb{E}[X|B; \Psi^g] := \mathbb{E}[I \otimes X|B \otimes I; \Psi^g] \quad (4.22)$$

of an observable for the choice of either $X = Q$ or $X = P$ of the meter system given another observable B of the target system. Here, for better readability, we have made a little abuse of notation by writing X instead of $I \otimes X$ and B for $B \otimes I$. We emphasise again that the conditional expectation (4.22) is defined as an *equivalence class of functions* that are integrable with respect to the probability measure

$$\mu_B^{\phi^g} := \mu_{B \otimes I}^{\Psi^g}, \quad (4.23)$$

which describes the behaviour of the outcome of the measurement of the local observable B on the target system. Here, we have introduced the density matrix

$$\phi^g := \text{Tr}_{\mathcal{K}}[|\Psi^g\rangle\langle\Psi^g|] \quad (4.24)$$

on the target system defined in a parallel manner as in (2.71). For its well-definedness, we note the following statement for reference.

Proposition 4.1 (Well-definedness of the Conditional Expectation). *In the context of the CM scheme, let*

- (i) *If $X = Y$: $|\phi\rangle \in \mathcal{H}$, $|\psi\rangle \in \text{dom}(X)$*
- (ii) *If $X \neq Y$: $|\phi\rangle \in \text{dom}(A)$, $|\psi\rangle \in \text{dom}(X)$*

be the choice of the initial states of the target and meter systems. Then, the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ is well-defined for all range of the interaction parameter $g \in \mathbb{R}$.

Proof. For demonstration, we shall only refer to Proposition 2.2 that guarantees the integrability of the outcomes of the measurement of X (*i.e.*, $|\mathbb{E}[I \otimes X; \Psi^g]| < \infty$) for all range of $g \in \mathbb{R}$, given the conditions assumed. ■

Post-selected Measurement. As a special subclass of this measurement scheme, we prepare the term *post-selected measurement scheme* to refer to the case where the conditioning observable $B = |\phi'\rangle\langle\phi'|$ happens to be a projection on some one-dimensional subspace of \mathcal{H} spanned by some normalised vector $|\phi'\rangle \in \mathcal{H}$, and in such a case, the act of conditioning will be occasionally referred to as the *post-selection*. It is also a common practice found in various literatures to call the state $|\phi\rangle$ prepared prior to the measurement the *initial* or the *pre-selected* state, and the normalised vector $|\phi'\rangle$ spanning the image of the one-dimensional projection $B = |\phi'\rangle\langle\phi'|$ the *final* or the *post-selected* state.

4.2.1. Topic: ‘Amplification Technique’ by Conditioning. It is widely known that, in general, the range of conditional expectation may exceed the (unconditional) expectation value, *i.e.*, for some clever choice of the conditioning observable B and its outcome $b \in \mathbb{R}$, one has

$$|\mathbb{E}[X; \Psi^g]| \leq |\mathbb{E}[X|B = b; \Psi^g]| \quad (4.25)$$

with non-vanishing probability. Clearly, this property should prove itself useful in some certain situations.

While this property has occasionally been utilised in experiments, it has recently caught wide attention due to the reports on the success of application in precision measurements, including the experimental detection of the spin-Hall effect of light (SHEL) in 2008 [33], and the detection of an ultra-sensitive beam deflection in a Sagnac interferometer in 2009 [34]. The experiments have effectively utilised the technique of conditioning (or post-selection) to yield an enhancement (or ‘amplification’) of an extremely small beam displacement to the extent that it is large enough to overcome various technical imperfections (noise level), and eventually realising significant detection of such tiny effects. In this context, this technique has often been referred to as the ‘weak value amplification’ or as ‘Aharonov-Albert-Vaidman effect’ of amplification [5].

Review of the Recent theoretical Analyses. Extensive theoretical analyses have been conducted in recent years from various viewpoints on the technical advantages of the technique of post-selection over the conventional unconditioned counterpart. Some of them addressed the question of signal amplification and its limit, where one asks the question as to what extent one can amplify the signal [35] and how one could achieve the optimisation [36]; the question of the existence of the limit of amplification will be addressed shortly in a more general framework. As far as the authors are aware of, the first sound analytic result appeared around 2012 [37], in which the limit to the amplification rate, as well as the signal-to-noise ratio has been explicitly presented. The computation was conducted for a special case where the observable A fulfils the condition $A^2 = I$ and the meter wave functions were assumed to be of Gaussian states, which we shall also address in a relatively more general setting later in this section, and also in Appendix A.

Others focused on the statistical loss which occurs due to the post-selection and examine the feasibility of improving the parameter estimation of the coupling constant g by post-selection based on estimation theory (for a concise review on the topic from this point of view, see [38]). The result is that the post-selection statistically deteriorates the quality of estimation, both in the case where ideal noiseless experiments can be performed [39], and also in some case where certain types of fully-known or controllable noise are present [40–42]. In an attempt to address the question of how the post-selection technique, while being statistically inferior to the unconditioned case, could be advantageous in realistic experiments, the authors have conducted a theoretical analysis on post-selected measurement in the presence of some intractable ‘measurement uncertainty’, a relatively modern concept in metrology to express *unknown* or *uncontrollable* source of technical imperfections [43]. It was then found that, while post-selection suffers from statistical deterioration, in certain cases the amplification effect becomes favourable in overcoming the unknown/uncontrollable source of technical imperfections one could not completely eliminate through ‘noise hunting’, which accordingly cannot be reduced from statistical reiteration. This suggests that the post-selection technique should be understood as the *practice of taking advantage of the trade-off relation* between the reduced contribution from intractable source of measurement uncertainty due to its signal amplification effect, and the statistical deterioration caused by the decrease in success probability.

4.2.2. Topic: ‘Limit of Amplification’ in Terms of Essential Suprema. In what follows, we provide a somewhat general result regarding the question of ‘limit of amplification’ by

conditioning, which has been one of the hottest topics among the study of the technical advantages in employing conditioning in experiments. A typical way to address this problem is to ask oneself, to what extent one could enlarge the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ by choosing an appropriate conditioning observable B and its outcome $b \in \sigma(B)$ with non-vanishing probability. By recalling the definition of essential supremum of a function (2.20), one realises that the question is equivalent to asking to what extent one could make the essential supremum of the conditional expectation

$$\|\mathbb{E}[X|B; \Psi^g]\|_\infty \quad (4.26)$$

large by the choice of the conditioning observable B .

Preliminaries. To prepare for our arguments, we first observe some basic facts regarding absolute continuity and essential suprema.

Lemma 4.2. *Let (X, \mathfrak{A}, μ) be a probability space, and let $\nu : \mathfrak{A} \rightarrow \mathbb{C}$ be a complex measure. Then, the following conditions are equivalent:*

- (i) $\nu \ll \mu$.
- (ii) $|\nu| \ll \mu$.
- (iii) *There exists a non-negative number $M \in [0, \infty]$ such that*

$$|\nu(A)| \leq |\nu|(A) \leq M \cdot \mu(A) \quad (4.27)$$

holds for all $A \in \mathfrak{A}$.

In such a cases, the Radon-Nikodým derivative $d\nu/d\mu$ exists by the Radon-Nikodým theorem, and its essential supremum $\|d\nu/d\mu\|_\infty$ gives the smallest of such M that satisfies (4.27).

Proof. For the equivalence of the condition (i) \Leftrightarrow (ii), the reader is referred to any textbooks on measure and integration theory. We already know from the Reference Material in Section 3.1 that $|\nu(A)| \leq |\nu|(A)$, $A \in \mathfrak{A}$. The implication (ii) \Rightarrow (iii) is then trivial by simply taking $M = \infty$. The converse (iii) \Rightarrow (ii) is also immediate by the definition of absolute continuity. Now that we have proved the equivalence of the three conditions, we move on to the demonstration of the final statement. To this end, first observe the evaluation

$$\begin{aligned} |\nu(A)| &= \left| \int_A \frac{d\nu}{d\mu} d\mu \right| \\ &\leq \int_A \left| \frac{d\nu}{d\mu} \right| d\mu \leq \left\| \frac{d\nu}{d\mu} \right\|_\infty \cdot \mu(A). \end{aligned} \quad (4.28)$$

Combining this with the minimality of the variation $|\nu|$, one sees that the choice $M = \|d\nu/d\mu\|_\infty$ of the upper bound satisfies (4.27). Now, suppose that there exists a non-negative number $0 \leq M < \|d\nu/d\mu\|_\infty$ satisfying (4.27). Then, by definition of the essential supremum, there exists a measurable set A satisfying $0 < \mu(A)$ and $M < \|d\nu/d\mu\|_A$ (just take $A := \{x \in X : M < |d\nu/d\mu|(x)\}$), hence

$$|\nu|(A) = \int_A \left| \frac{d\nu}{d\mu} \right| d\mu > M \cdot \mu(A), \quad (4.29)$$

which contradicts the minimality of $|\nu|$. ■

As a corollary to this, the following observation is of special interest.

Corollary 4.3 (Conditional Expectations and Essential Suprema). *Let (X, \mathfrak{A}, μ) be a probability space, $f : X \rightarrow \mathbb{R}$ be μ -integrable, and $\mathfrak{B} \subset \mathfrak{A}$ be a sub- σ -algebra. Then the evaluation*

$$\|\mathbb{E}[f|\mathfrak{B}]\|_\infty \leq \|f\|_\infty \quad (4.30)$$

holds. As a direct consequence, if moreover a measurable function $g : X \rightarrow \mathbb{R}$ is given, the evaluation

$$\|\mathbb{E}[fg]\|_\infty \leq \|f\|_\infty \quad (4.31)$$

naturally holds.

Proof. First recall that the conditional expectation $\mathbb{E}[f|\mathfrak{B}]$ is nothing but the Radon-Nikodým derivative of the complex measure $f \odot \mu$ with respect to the restriction $\mu|_{\mathfrak{B}}$. Letting $\nu := f \odot \mu$ and replacing μ by $\mu|_{\mathfrak{B}}$ in the above Lemma, one finds

$$|\nu(A)| = \left| \int_A f \, d\mu \right| \leq \|f\|_\infty \cdot \mu(A), \quad (4.32)$$

hence

$$\|d\nu/d\mu\|_\infty = \|\mathbb{E}[f|\mathfrak{B}]\|_\infty \leq \|f\|_\infty, \quad (4.33)$$

which was to be demonstrated. ■

In casual language, this is to say that each value of the conditional expectation of f never exceeds the maximum number that f takes under a given probability measure, which is a result that should be intuitively clear. As a direct application of the result in the context of quantum measurement of a pair of simultaneously measurable observables A and B , this reduces to the following.

Corollary 4.4. *Given a pair of strongly commuting self-adjoint operators A and B and a fixed state $|\phi\rangle \in \text{dom}(A)$, the essential supremum of the conditional expectation of A given B is never greater than*

$$\|\mathbb{E}[A|B; \phi]\|_\infty \leq \|A\|_\infty^\phi, \quad (4.34)$$

where $\|A\|_\infty^\phi := \|a\|_\infty$ denotes the essential supremum of the measurable function $a \mapsto a$ under the probability measure μ_A^ϕ describing the behaviour of the outcome of the measurement of A on the state $|\phi\rangle$. If A happens to be bounded, its operator norm²⁰ $\|A\|$ becomes the universal (i.e., state independent) upper bound of $\|A\|_\infty^\phi$, hence

$$\|\mathbb{E}[A|B; \phi]\|_\infty \leq \|A\|_\infty^\phi \leq \|A\| < \infty \quad (4.36)$$

holds for all $|\phi\rangle \in \mathcal{H}$.

²⁰ For a bounded operator X , recall that the *operator norm* of X is defined by

$$\|X\| := \sup\{\|X\phi\| : \|\phi\| = 1\}. \quad (4.35)$$

Proof. The former part of the statement is immediate by Corollary 4.3. For the latter part, we first recall that the *numerical range* of a self-adjoint operator X is defined as

$$W(X) := \{\langle \psi, X\psi \rangle : |\psi\rangle \in \text{dom}(X), \|\psi\|^2 = 1\}, \quad (4.37)$$

which is nothing but the collection of all possible expectation values of X . Now, a direct application of the Cauchy-Schwarz inequality leads to

$$|\mathbb{E}[X; \phi]| \leq \|X\|, \quad \mathbb{E}[X; \phi] \in W(X), \quad (4.38)$$

for bounded X , and by recalling the basic relation $\sigma(X) \subset \overline{W(X)}$, where the overline on $W(X)$ denotes its topological closure, one concludes

$$\begin{aligned} \|X\|_\infty^\phi &\leq \sup\{|x| : x \in \sigma(X)\} \\ &\leq \sup\{|x| : x \in \overline{W(X)}\} \leq \|X\|, \end{aligned} \quad (4.39)$$

which was to be demonstrated. ■

The latter part of the statement is to say that conditional expectations of a bounded observable has a universal upper bound given by its operator norm, which is also a result that should be intuitively clear.

On the ‘Limit of Amplification’ by Conditional Measurement. As a direct application of the above corollary to our problem, we obtain the main result of this passage.

Proposition 4.5 (Amplification by Conditioning). *Under the framework of the CM scheme, the essential supremum of the conditional expectation of X given B is never greater than that of the UM scheme of X*

$$|\mathbb{E}[X|B = b; \Psi^g]| \leq \|\mathbb{E}[X|B; \Psi^g]\|_\infty \leq \|X\|_\infty^{\psi^g}, \quad (4.40)$$

where $\|X\|_\infty^{\psi^g} := \|x\|_\infty$ denotes the essential supremum of x under the probability measure $\mu_X^{\psi^g}$ describing the behaviour of the outcome of the local measurement X on the meter system. In other words, $\|X\|_\infty^{\psi^g}$ gives the (conditioning-observable-independent) upper bound to the extent the conditional expectation can be ‘amplified’ by means of conditioning²¹.

In physical terms, this is to say that the extent one may ‘amplify’ the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ by means of changing the conditioning observable B is predetermined by $\|X\|_\infty^{\psi^g}$. This is one general form to answer the question of the existence of the limit of ‘amplification’ by conditioning.

As the next step, one might eventually be interested in seeking for the condition under which $\|X\|_\infty^{\psi^g}$ is bounded from above, even if we could freely choose the initial state $|\phi\rangle$ of the target system. This would create a universal upper bound of $\|\mathbb{E}[X|B; \Psi^g]\|_\infty$ that is indifferent to both the initial and final configurations of the target system (*i.e.*, the choice of the initial target state $|\phi\rangle$ and the conditioning observable B). As we have learned from the discussions above, this would typically be the case when there exists a subspace $U(g, \psi) \subset$

²¹ Recall the inherent subtlety when we use the expression $\mathbb{E}[X|B = b; \Psi^g]$. The left most inequality in (4.40) should thus be understood to hold $\mu_B^{\phi^g}$ -a.e.

$\mathcal{H} \otimes \mathcal{K}$, for fixed $g \in \mathbb{R}$ and $|\psi\rangle \in \text{dom}(X)$, such that $|\Psi^g\rangle \in U(g, \psi)$ for all $|\phi\rangle \in \text{dom}(A)$, and that the restriction of $I \otimes X$ on $U(g, \psi)$ is bounded.

Proposition 4.6 (Limit of Amplification by Conditioning). *Under the framework of the CM scheme, let both the interaction parameter $g \in \mathbb{R}$ and the initial meter state $|\psi\rangle \in \text{dom}(X)$ be fixed, and suppose that the target observable A has a spectrum $\sigma(A) = \{a_1, \dots, a_N\}$, $N \in \mathbb{N}^\times$ of finite cardinality. Then, the following facts hold:*

- (i) *The density operator ψ^g of the meter system (2.71) can be written as a probabilistic mixture of a finite number of projection operators (pure states) supported on the finite-dimensional (at most N -dimensional) subspace*

$$\mathcal{K}(g, \psi) := \text{span}(\{|e^{-iga_1 Y} \psi\rangle, \dots, |e^{-iga_N Y} \psi\rangle\}), \quad (4.41)$$

which is independent of the initial choice $|\phi\rangle \in \mathcal{H}$ of the target state.

- (ii) *The restriction $X|_{\mathcal{K}(g, \psi)}$ of the meter observable X on the subspace (4.41) is bounded, and thus its operator norm*

$$\|\mathbb{E}[X|B; \Psi^g]\|_\infty \leq \|X|_{\mathcal{K}(g, \psi)}\| < \infty \quad (4.42)$$

provides a finite universal upper bound to the conditional expectation that is independent of the configuration of the target system (i.e., the choice of the initial state $|\phi\rangle \in \mathcal{H}$ and that of the conditioning observable B).

Proof. Under the above condition, first observe that

$$\begin{aligned} |\Psi^g\rangle &= \sum_{n=1}^N (\Pi_{a_n} \otimes e^{-iga_n Y}) |\phi \otimes \psi\rangle \\ &= \sum_{n=1}^N (\Pi_{a_n} |\phi\rangle \otimes |e^{-iga_n Y} \psi\rangle), \end{aligned} \quad (4.43)$$

where we have used (2.77). One readily finds from the above formula that the density operator

$$\psi^g = \text{Tr}_{\mathcal{H}} [|\Psi^g\rangle\langle\Psi^g|], \quad (4.44)$$

defined as in (2.71), can indeed be written as a probabilistic mixture of a finite number of projection operators (pure states) supported on the subspace (4.41). We then recall that any operator X defined on a finite-dimensional Hilbert space are necessarily bounded, and thus observe that the current problem at hand reduces to the situation of Corollary 4.4. \blacksquare

In physical terms, this is to say that there exists a finite limit $\|X|_{\mathcal{K}(g, \psi)}\| < \infty$ to the extent one may ‘amplify’ the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ by means of only changing the configuration of the target system (namely, by changing either or both the conditioning observable B and the initial state $|\phi\rangle \in \mathcal{H}$ of the target system). Specifically, the evaluation

$$|\mathbb{E}[X|B = b; \Psi^g]| \leq \|X|_{\mathcal{K}(g, \psi)}\| < \infty \quad (4.45)$$

holds for all $b \in \sigma(B)$ up to a set of probability zero, and the upper bound $\|X|_{\mathcal{K}(g, \psi)}\|$ does not depend on the choice of B nor $|\phi\rangle$. Naturally, if one could change either the interaction parameter g or the initial state $|\psi\rangle$ of the meter system alongside, the above result is no more valid.

4.3. Recovery of the Target Profile

Parallel to the study of the UM scheme, we are now interested in the information of the target system which is to be extracted from the CM scheme. Following the line of arguments for the UM scheme, we are specifically interested in investigating the local behaviour of the outcome of the CM scheme around $g = 0$, *i.e.*, the *weak conditioned measurement*, in which the target of our analysis is the map

$$g \mapsto \mathbb{E}[X|B; \Psi^g] \quad (4.46)$$

from the interaction parameter g to the conditional expectation of X given B , which was in general defined as a map from the real line to an equivalent class of functions. To this end, we first conduct a preliminary observation.

4.3.1. Preliminary Observation. Since the definition of the conditional expectation is given in a rather abstract way, the conditional expectation (4.22) in general does not admit an explicit expression by vectors and operators (in contrast to the UM case (2.73), which always admits such an explicit expression). In view of this, it would be sometimes helpful if one could find a condition for which the conditional expectation (4.22) of our interest may be explicitly written down. We first point out that this will be indeed the case given that the spectrum of the conditioning observable B has finite cardinality. Now, let

$$B = \sum_{n=1}^N b_n \Pi_{b_n} \quad (4.47)$$

be the spectral decomposition of B , where $\sigma(B) = \{b_1, \dots, b_N\}$ is any enumeration of its eigenvalues, and $\Pi_b := E_B(\{b\})$, $b \in \sigma(B)$ denotes the unique projection on the eigenspace associated to it. It is then fairly straightforward to see by definition that the conditional expectation of X given B is explicitly given by

$$\begin{aligned} & \mathbb{E}[X|B = b; \Psi^g] \\ &= \begin{cases} \mathbb{E}[\Pi_b \otimes X; \Psi^g] / \|(\Pi_b \otimes I)\Psi^g\|^2, & (b \in \sigma(B), \|(\Pi_b \otimes I)\Psi^g\|^2 \neq 0), \\ \text{indefinite}, & (\text{else}). \end{cases} \end{aligned} \quad (4.48)$$

Here, recall that conditional expectations are defined as an equivalence class of functions, and hence its value for the outcome b of the measurement of the observable B such that the probability of observing it is vanishing, is indefinite by definition. The study of the weak CM scheme then reduces to the analysis of the map

$$g \mapsto \mathbb{E}[X|B = b; \Psi^g] \quad (4.49)$$

for each $b \in \sigma(B)$ such that the probability of observing it is non-vanishing. Since this is a map from the real line to itself (*i.e.*, a function), it should be a much more familiar and straightforward object to deal with.

Objective of this Passage. In what follows, we will be discussing the differentiability of the function (4.49) at the point $g = 0$. To this end, first observe that the choice of $b \in \sigma(B)$ for which the probability of observing it is non-vanishing is dependent on g . Hence, for each $b \in \sigma(B)$, we must first guarantee its well-definedness, at least on some neighbourhood of

$g = 0$. Fortunately, this is indeed the case for the choice $b \in \sigma(B)$ such that the probability of finding it on the initial state $|\phi\rangle$ of the target system $\mathbb{E}[\Pi_b \otimes I; \Psi^0] = \|\Pi_b \phi\|^2 \neq 0$ is non-vanishing, due to continuity of the function $g \mapsto \mathbb{E}[\Pi_b \otimes I; \Psi^g]$. The main objective of this passage is to demonstrate the following statement.

Proposition 4.7 (Differentiability of the Conditional Expectation: Preliminary). *Suppose that the conditioning observable B has spectrum of finite cardinality, and moreover let $|\phi\rangle \in \text{dom}(A)$, $|\psi\rangle \in \mathcal{D} \subset \text{dom}(X)$ (the subspace \mathcal{D} is defined as in (2.37)) be assumed, so that the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ is well-defined for all range of $g \in \mathbb{R}$. Then for $b \in \sigma(B)$ such that $\|\Pi_b \phi\|^2 \neq 0$, the conditional expectation $\mathbb{E}[X|B = b; \Psi^g]$ is well-defined on some neighbourhood of $g = 0$. It is moreover differentiable with respect to g at the origin, for which the differential coefficient reads*

$$\begin{aligned} & \left. \frac{d}{dg} \mathbb{E}[X|B = b; \Psi^g] \right|_{g=0} \\ &= 2 \operatorname{Re} \left[\frac{\langle \phi, \Pi_b A \phi \rangle}{\|\Pi_b \phi\|^2} \right] \cdot \mathbb{C}\mathbb{V}_A[X, Y; \psi] + 2 \operatorname{Im} \left[\frac{\langle \phi, \Pi_b A \phi \rangle}{\|\Pi_b \phi\|^2} \right] \cdot \mathbb{C}\mathbb{V}_S[X, Y; \psi]. \end{aligned} \quad (4.50)$$

Here, we have introduced the quantities,

$$\mathbb{C}\mathbb{V}_S[X, Y; \psi] := \mathbb{E}[\{X, Y\}/2; \psi] - \mathbb{E}[X; \psi]\mathbb{E}[Y; \psi], \quad (4.51)$$

$$\mathbb{C}\mathbb{V}_A[X, Y; \psi] := \mathbb{E}[[X, Y]/(2i); \psi], \quad (4.52)$$

occasionally called the symmetric and anti-symmetric (quantum) covariance²² of X and Y on the state $|\psi\rangle \in \mathcal{D}$, respectively, where $\{X, Y\} := XY + YX$ denotes the anti-commutator (not to be confused with the braces denoting sets).

Proof. Throughout the proof, we choose $b \in \sigma(B)$ such that $\|\Pi_b \phi\|^2 \neq 0$. Then, it is fairly straightforward to see that the map

$$\mathbb{E}[X|B = b; \Psi^g] = \frac{\mathbb{E}[\Pi_b \otimes X; \Psi^g]}{\mathbb{E}[\Pi_b \otimes I; \Psi^g]}, \quad g \in U_0, \quad (4.54)$$

is well-defined on some neighbourhood U_0 around the origin $g = 0$. It then follows directly from the expression (4.54) that the differentiability of both the numerator and the denominator of the r. h. s. gives a sufficient condition for the conditional expectation $\mathbb{E}[X|B = b; \Psi^g]$ to be differentiable. In order to simplify our notations, we assume in the following that all the vectors $|\phi\rangle$ and $|\psi\rangle$, respectively representing the initial quantum states of the target and the meter system, are normalised. Since the proof is rather lengthy, we divide it into several parts.

Leibniz Rule. To prepare for our arguments, we first recall some basic facts. Let $F, G : U \rightarrow \mathcal{H}$ be a map from an open subset $U \subset \mathbb{R}$ of the real line to a Hilbert space \mathcal{H} . If both

²² Note that in the case where the two observables coincide $X = Y$, the symmetric quantum covariance reduces to the familiar variance,

$$\mathbb{C}\mathbb{V}_S[X, X; \psi] = \mathbb{V}[X; \psi] := \mathbb{E}[X^2; \psi] - \mathbb{E}[X; \psi]^2, \quad (4.53)$$

which is reminiscent of the familiar result in classical probability theory, whereas the anti-symmetric covariance reduces to null $\mathbb{C}\mathbb{V}_A[X, X; \psi] = 0$.

maps F and G are strongly differentiable at $t_0 \in U$, the inner product $t \mapsto \langle F(t), G(t) \rangle$ is differentiable at $t_0 \in U$, and the derivative satisfies the Leibniz rule,

$$\begin{aligned} \left. \frac{d}{dt} \langle F(t), G(t) \rangle \right|_{t=t_0} &:= \lim_{u \rightarrow 0} \frac{\langle F(u+t_0), G(u+t_0) \rangle - \langle F(t_0), G(t_0) \rangle}{u} \\ &= \lim_{u \rightarrow 0} \frac{\langle F(u+t_0) - F(t_0), G(u+t_0) \rangle + \langle F(t_0), G(u+t_0) - G(t_0) \rangle}{u} \\ &= \left\langle \frac{dF(t_0)}{dt}, G(t_0) \right\rangle + \left\langle F(t_0), \frac{dG(t_0)}{dt} \right\rangle. \end{aligned} \quad (4.55)$$

Differentiability of the Numerator. To prove the differentiability of the numerator of (4.54) and obtain its derivative, we first introduce two auxiliary maps $F(g) := |\Psi^g\rangle$ and $G_X(g) := (\Pi_b \otimes X)F(g)$, by which we rewrite the numerator

$$g \mapsto \mathbb{E}[\Pi_b \otimes X; \Psi^g] = \langle F(g), G_X(g) \rangle \quad (4.56)$$

in terms of their inner products. From the Leibniz rule, one sees that the desired result can be immediately obtained once the differentiability of both the maps $F(g)$ and $G_X(g)$ are proven and their derivatives are given.

As for the strong differentiability of the map $g \mapsto F(g)$, one readily finds by Stone's theorem on one-parameter unitary groups that the condition

$$|\phi\rangle \in \text{dom}(A), \quad |\psi\rangle \in \mathcal{D} \subset \text{dom}(Y) \quad (4.57)$$

would suffice, in which case the derivative is given by

$$\frac{dF(0)}{dg} = -i(A \otimes Y)|\phi \otimes \psi\rangle. \quad (4.58)$$

As for the map $g \mapsto G_X(g)$, we first observe that it is written as

$$G_X(g) = (\Pi_b \otimes I)(I \otimes X)F(g). \quad (4.59)$$

Due to the boundedness (continuity) of the operator $(\Pi_b \otimes I)$, strong differentiability of the vector-valued map $g \mapsto (I \otimes X)F(g)$ would give a sufficient condition for $G_X(g)$ to be strongly differentiable, which one readily proves under the condition

$$|\phi\rangle \in \text{dom}(A), \quad |\psi\rangle \in \mathcal{D} \subset \text{dom}(XY) \cap \text{dom}(Y) \quad (4.60)$$

by imitating the arguments we have made starting from (2.52) with the help of the relation (2.80). Now that the strong differentiability of both the maps $g \mapsto F(g)$, $G_X(g)$ are proven, one finds from the closedness of the self-adjoint operator $(\Pi_b \otimes X)$ that

$$\begin{aligned} \frac{dG_X(0)}{dg} &= (\Pi_b \otimes X) \frac{dF(0)}{dg} \\ &= -i(\Pi_b \otimes X)(A \otimes Y)|\phi \otimes \psi\rangle. \end{aligned} \quad (4.61)$$

Given the results (4.58) and (4.61), the Leibniz rule leads to the desired differentiability of the numerator (4.56), in which one computes its derivative as

$$\begin{aligned}
\left. \frac{d}{dg} \mathbb{E} [\Pi_b \otimes X; \Psi^g] \right|_{g=0} &= \left\langle \frac{dF(0)}{dg}, (\Pi_b \otimes X) F(0) \right\rangle + \left\langle F(0), (\Pi_b \otimes X) \frac{dF(0)}{dg} \right\rangle \\
&= 2 \operatorname{Re} \left[\left\langle F(0), (\Pi_b \otimes X) \frac{dF(0)}{dg} \right\rangle \right] \\
&= 2 \operatorname{Re} [-i \langle F(0), (\Pi_b \otimes X)(A \otimes Y) F(0) \rangle] \\
&= 2 \operatorname{Im} [\langle \phi, \Pi_b A \phi \rangle \langle \psi, XY \psi \rangle] \\
&= 2 \operatorname{Re} [\langle \phi, \Pi_b A \phi \rangle] \cdot \mathbb{E}[\{X, Y\}/(2i); \psi] \\
&\quad + 2 \operatorname{Im} [\langle \phi, \Pi_b A \phi \rangle] \cdot \mathbb{E}[\{X, Y\}/2; \psi], \tag{4.62}
\end{aligned}$$

where we have used the operator equality

$$XY = \frac{\{X, Y\}}{2} + i \frac{[X, Y]}{2i} \tag{4.63}$$

valid on the subspace \mathcal{D} .

Differentiability of the Denominator. The proof for the differentiability of the denominator $\mathbb{E} [\Pi_b \otimes I; \Psi^g]$ goes essentially the same as that for the numerator, where one readily proves its differentiability at $g = 0$ under the condition $|\phi\rangle \in \operatorname{dom}(A)$, $|\psi\rangle \in \operatorname{dom}(Y)$, in which case the derivative reads

$$\left. \frac{d}{dg} \mathbb{E} [\Pi_b \otimes I; \Psi^g] \right|_{g=0} = 2 \operatorname{Im} [\langle \phi, \Pi_b A \phi \rangle] \cdot \mathbb{E}[Y; \psi], \tag{4.64}$$

by formally replacing X with I in (4.62).

Final Result. Combining the above two results (4.62) and (4.64), one concludes that, given the choice $|\phi\rangle \in \operatorname{dom}(A)$ and $b \in \sigma(B)$ with $\|\Pi_b \phi\|^2 \neq 0$ of the target configuration, and $|\psi\rangle \in \mathcal{D}$ for the meter system, the conditional expectation $\mathbb{E}[X|B = b; \Psi^g]$ is indeed differentiable at $g = 0$. Its derivative can then be evaluated based on the classical result of calculus (the quotient rule for derivative) as

$$\begin{aligned}
&\left. \frac{d}{dg} \mathbb{E}[X|B = b; \Psi^g] \right|_{g=0} \\
&= \frac{\left. \frac{d}{dg} \mathbb{E} [\Pi_b \otimes X; \Psi^g] \right|_{g=0} \cdot \mathbb{E} [\Pi_b \otimes I; \Psi^0] - \mathbb{E} [\Pi_b \otimes X; \Psi^0] \cdot \left. \frac{d}{dg} \mathbb{E} [\Pi_b \otimes I; \Psi^g] \right|_{g=0}}{\mathbb{E} [\Pi_b \otimes I; \Psi^0]^2} \\
&= 2 \operatorname{Re} \left[\frac{\langle \phi, \Pi_b A \phi \rangle}{\|\Pi_b \phi\|^2} \right] \cdot \mathbb{E}[\{X, Y\}/(2i); \psi] \\
&\quad + 2 \operatorname{Im} \left[\frac{\langle \phi, \Pi_b A \phi \rangle}{\|\Pi_b \phi\|^2} \right] \cdot (\mathbb{E}[\{X, Y\}/2; \psi] - \mathbb{E}[X; \psi] \mathbb{E}[Y; \psi]) \\
&= 2 \operatorname{Re} \left[\frac{\langle \phi, \Pi_b A \phi \rangle}{\|\Pi_b \phi\|^2} \right] \cdot \mathbb{C}\mathbb{V}_A[X, Y; \psi] + 2 \operatorname{Im} \left[\frac{\langle \phi, \Pi_b A \phi \rangle}{\|\Pi_b \phi\|^2} \right] \cdot \mathbb{C}\mathbb{V}_S[X, Y; \psi]. \tag{4.65}
\end{aligned}$$

We have thus verified our desired statement (4.50). ■

4.3.2. *Conditional Quasi-expectations of Quantum Observables.* Now that we have computed the derivative of the map (4.49) for the special case, we are now interested in the case in which the conditioning observable B is general, and wish to specify the limit of the formal expression

$$\lim_{g \rightarrow 0} \frac{\mathbb{E}[X|B; \Psi^g] - \mathbb{E}[X|B; \Psi^0]}{g} \quad (4.66)$$

and the topology in which the convergence is meant. From the result of Proposition 4.7, one might naturally conjecture that the limit is given by

$$2 \operatorname{Re} f \cdot \mathbb{C}\mathbb{V}_A[X, Y; \psi] + 2 \operatorname{Im} f \cdot \mathbb{C}\mathbb{V}_S[X, Y; \psi] \quad (4.67)$$

with a ‘function’ f defined formally as

$$f(b) := \frac{\langle \phi, \Pi_b A \phi \rangle}{\|\Pi_b \phi\|^2}. \quad (4.68)$$

In order to make this observation a precise mathematical statement, we first introduce a convenient concept.

Conditional Quasi-expectations. Observing that in the case where A and B are simultaneously measurable, the function (4.68) is nothing but the conditional expectation of A given B . In general, however, the target observable and the conditioning observable B need not be simultaneously observable. We thus wish to define a quantum analogue of conditional expectations of an observable A given another observable B , well-defined even for the pair that are not necessarily simultaneously measurable. To this end, we first fix a non-zero vector $|\phi\rangle \in \operatorname{dom}(A)$ and consider a complex measure

$$\nu(\Delta) := \langle \phi, E_B(\Delta) A \phi \rangle / \|\phi\|^2, \quad \Delta \in \mathfrak{B}, \quad (4.69)$$

where E_B is the unique spectral measure accompanying B . Now, a direct application of the Cauchy-Schwarz inequality leads to

$$|\langle \phi, E_B(\Delta) A \phi \rangle| \leq \|E_B(\Delta) \phi\| \cdot \|A \phi\|, \quad (4.70)$$

by which one finds the absolute continuity $\nu \ll \mu_B^\phi$, where $\mu_B^\phi(\Delta) := \|E_B(\Delta) \phi\|^2 / \|\phi\|^2$ as usual. This allows us to define the Radon-Nikodým derivative

$$\mathbb{E}[A|B; \phi] := d\nu / d\mu_B^\phi. \quad (4.71)$$

By definition, it is the unique μ_B^ϕ -integrable (equivalence class of) function(s) that satisfies

$$\langle \phi, E_B(\Delta) A \phi \rangle / \|\phi\|^2 = \int_{\Delta} \mathbb{E}[A|B = b; \phi] d\mu_B^\phi(b), \quad \Delta \in \mathfrak{B}, \quad (4.72)$$

and as such,

$$\mathbb{E}[A; \phi] = \int_{\mathbb{R}} \mathbb{E}[A|B = b; \phi] d\mu_B^\phi(b) \quad (4.73)$$

holds in particular. Incidentally, when the state $|\phi_a\rangle \in \operatorname{dom}(A)$ happens to be an eigenvector of A with the eigenvalue a , the map

$$\mathbb{E}[A|B; \phi_a] = a \quad (4.74)$$

becomes a constant function independent of the choice of the conditioning observable B . The map $\mathbb{E}[A|B; \phi]$ thus shares properties similar to the conditional expectations, and in the

special case in which A and B happens to be simultaneously measurable, it actually reduces to the standard conditional expectation. However, as one finds shortly below, it can be shown by reductio ad absurdum that the map $\mathbb{E}[A|B; \phi]$ may not admit itself to be understood as a standard conditional expectation in the case where the pair of observables concerned does not admit coexistence. These preliminary observations may tempt one to call the map (4.71) a *conditional quasi-expectation* of A given B .

Arbitrariness to Conditional Quasi-expectations. As one may immediately notice, there exists an arbitrariness to the way one may define conditional quasi-expectations. For example, one may just define the complex conjugate of the complex measure (4.69) as

$$\nu^*(\Delta) = \langle \phi, AE_B(\Delta)\phi \rangle / \|\phi\|^2 \quad (4.75)$$

and introduce the Radon-Nikodým derivative as

$$\mathbb{E}^*[A|B; \phi] := d\nu^*/d\mu_B^\phi = \mathbb{E}[A|B; \phi]^*. \quad (4.76)$$

One may conduct analogous reasoning to verify that the function $\mathbb{E}^*[A|B; \phi]$ also satisfies properties similar to the usual conditional expectations, and that both definitions coincide when the pair of A and B happens to be simultaneously measurable. One may even consider a complex linear combination of $\mathbb{E}[A|B; \phi]$ and its complex conjugate to define

$$\begin{aligned} \mathbb{E}^\alpha[A|B; \phi] &:= \frac{1+\alpha}{2} \cdot \mathbb{E}[A|B; \phi] + \frac{1-\alpha}{2} \cdot \mathbb{E}^*[A|B; \phi] \\ &= \operatorname{Re}[\mathbb{E}[A|B; \phi]] + \alpha i \operatorname{Im}[\mathbb{E}[A|B; \phi]], \quad \alpha \in \mathbb{C}, \end{aligned} \quad (4.77)$$

for example, so that $\mathbb{E}^1[A|B; \phi] = \mathbb{E}[A|B; \phi]$ and $\mathbb{E}^{-1}[A|B; \phi] = \mathbb{E}^*[A|B; \phi]$. In fact, it reveals that there exists a multitude of potential candidates for possible definitions of such ‘conditional quasi-expectations’, all sharing desirable properties mentioned earlier. We shall be returning to this problem in a more general framework of quasi-joint-probabilities of quantum observables in Section 6, but for our purpose and the scope of this paper, it suffices to concentrate only on the family (4.77) for definiteness, and we thus introduce:

Definition (Conditional Quasi-expectation of A given B). *Let A and B be self-adjoint operators on a Hilbert space \mathcal{H} , and let E_B be the spectral measure of B . For a given state $|\phi\rangle \in \operatorname{dom}(A)$, we call the family of complex linear combinations of the Radon-Nikodým derivatives (4.77) the complex-parametrised family of conditional quasi-expectations of A given B . They are, by definition, a (family of) complex function(s) defined on the spectrum $\sigma(B)$.*

Note, by definition, that each member $\mathbb{E}^\alpha[A|B; \phi]$, $\alpha \in \mathbb{C}$, of the family of conditional quasi-expectations is integrable with respect to the probability measure μ_B^ϕ , and its total integration coincides with the expectation value $\mathbb{E}[A; \phi]$ of A . If the conditioning observable B happens to possess spectrum with finite cardinality, so that its spectral decomposition reads (4.47), the conditional quasi-expectation admits an expression by operators and vectors as

$$\mathbb{E}[A|B = b; \phi] = \begin{cases} \langle \phi, \Pi_b A \phi \rangle / \|\Pi_b \phi\|^2, & (b \in \sigma(B), \|\Pi_b \phi\|^2 \neq 0), \\ \text{indefinite}, & (\text{else}) \end{cases} \quad (4.78)$$

and

$$\mathbb{E}^\alpha[A|B; \phi] = \frac{1 + \alpha}{2} \cdot \mathbb{E}[A|B; \phi] + \frac{1 - \alpha}{2} \cdot \mathbb{E}[A|B; \phi]^* \quad (4.79)$$

if explicitly written out.

Conditional Quasi-expectations, Two-state Values and the Weak Value. Incidentally, if the conditioning observable happens to be a projection $B = |\phi'\rangle\langle\phi'|$ on a one-dimensional subspace of \mathcal{H} spanned by a unit vector $|\phi'\rangle$ (*i.e.*, a post-selection), the conditional quasi-expectation of A given the outcome $B = 1$ reads

$$\mathbb{E}^\alpha[A|B = 1; \phi] = \frac{1 + \alpha}{2} \cdot \frac{\langle\phi', A\phi\rangle}{\langle\phi', \phi\rangle} + \frac{1 - \alpha}{2} \cdot \frac{\langle\phi, A\phi'\rangle}{\langle\phi, \phi'\rangle}, \quad (4.80)$$

given that the probability of finding the outcome 1 of B is non-vanishing $\mu_B^\phi(\{1\}) = |\langle\phi', \phi\rangle|^2 \neq 0$. Specifically for the choice $\alpha = 1$, this reduces to

$$\mathbb{E}[A|B = 1; \phi] = \frac{\langle\phi', A\phi\rangle}{\langle\phi', \phi\rangle} =: A_w, \quad (4.81)$$

The value A_w is widely referred to as Aharonov's *weak value* [5, 6] of A for the pair of the pre-selected state $|\phi\rangle \in \text{dom}(A)$ and the post-selected state $|\phi'\rangle \in \mathcal{H}$. Historically, the weak value is said to have been originally introduced as a hypothetical value of an observable A assigned to a quantum *process* from the pre-selected to the post-selected state, generalising the common practice of solely assigning values to a single static *state* in the standard framework of quantum mechanics. Following this philosophy, the value (4.80) termed the *two-state value* [44] of A under the respective selections of states was recently introduced in an attempt to generalise the idea of the weak value and to find out the possible form of a quantity of an observable specified by two quantum states. An application of the generalised Gleason's theorem revealed that, under certain desirable conditions, the most general form of the values of an observable A that can be assigned to the two specification of the quantum states $|\phi\rangle \in \text{dom}(A)$, $|\phi'\rangle \in \mathcal{H}$ satisfying $\langle\phi', \phi\rangle \neq 0$ is given by (4.80) with a parameter $\alpha \in \mathbb{C}$ representing the ambiguity inherent to it.

Essential Supremum of Conditional Quasi-expectations. While conditional quasi-expectations and the standard conditional expectations share various properties in common, the non-commutative nature of quantum observables results in some interesting distinctions between the two concepts. In this paper, as an example, we shall focus on the remarkable difference in the behaviour of their essential suprema. Now, as one recalls from Corollary 4.4, for a pair of *simultaneously measurable* observables A and B and a fixed state $|\phi\rangle \in \text{dom}(A)$, the essential supremum of the conditional expectation $\|\mathbb{E}[A|B; \phi]\|_\infty$ is never greater than the essential supremum $\|A\|_\infty^\phi$ of the measurable function $a \mapsto a$ under the probability measure μ_A^ϕ . If A happens to be bounded, the operator norm $\|A\|$ gives the state independent universal upper bound to $\|A\|_\infty^\phi$, which in turn also naturally becomes an upper bound to the conditional expectation $\mathbb{E}[A|B; \phi]$. However, in general, this property is no longer preserved when A and B fail to be simultaneously measurable. There are several possible ways to express this discrepancy, but for brevity, we formulate it in the following manner.

To this end, we first prepare a terminology. In this paper, we say that an observable A on \mathcal{H} is *non-trivial* if A is not a scalar multiple of the identity operator tI , $t \in \mathbb{R}$, or equivalently,

if A has a spectrum $\sigma(A)$ of cardinality not less than 2. Note that the non-triviality of A automatically implies $\dim(\mathcal{H}) \geq 2$, where $\dim(\mathcal{H})$ denotes the dimension of the Hilbert space \mathcal{H} . Since trivial operators strongly commute with any other self-adjoint operators, the function $\mathbb{E}^\alpha[A|B; \phi]$ always become an authentic conditional expectation, revealing itself to be a constant function always taking its unique eigenvalue $\mathbb{E}^\alpha[A|B; \phi] = t$, whose case is not interesting for our purpose. Hence, we shall from now on confine ourselves to the case where A is non-trivial.

Proposition 4.8 (Essential Supremum of Conditional Quasi-expectations). *Let A be a non-trivial observable, $|\phi\rangle \in \text{dom}(A)$ a vector that is not an eigenvector of A , and let $\alpha \in \mathbb{C}$ be any choice of the ambiguity parameter of the conditional quasi-expectation. Then, for any non-negative number $0 \leq M < \infty$, there exists a self-adjoint operator B (not-necessarily simultaneously measurable with A) such that the essential supremum of the conditional quasi-expectation of A given B is not less than*

$$M \leq \|\mathbb{E}^\alpha[A|B; \phi]\|_\infty. \quad (4.82)$$

Specifically, one may always choose such conditioning observable $B = |\phi'\rangle\langle\phi'|$ to be a projection onto a one-dimensional subspace of \mathcal{H} spanned by some unit vector $|\phi'\rangle \in \mathcal{H}$.

Proof. It suffices to prove that, one may always adjust the choice of the conditioning observable $B = |\phi'\rangle\langle\phi'|$ so that the conditional quasi-expectation

$$\mathbb{E}^\alpha[A|B = 1; \phi] = \begin{cases} c, & (c \in \mathbb{C}), \quad (\alpha \neq 0) \\ r, & (r \in \mathbb{R}), \quad (\alpha = 0) \end{cases} \quad (4.83)$$

may take any complex number for the choice $\alpha \neq 0$, and any real number for the choice $\alpha = 0$, while maintaining the probability of observing it to be non-vanishing $\mu_B^\phi(\{1\}) > 0$. The proof is a direct corollary of Proposition 4.9 that follows immediately. \blacksquare

In particular, this result is to say that one may always choose a conditioning observable B such that the essential supremum $\|\mathbb{E}[A|B; \phi]\|_\infty$ of the conditional quasi-expectation exceeds $\|A\|_\infty^\phi$, which is never possible for standard conditional expectations defined for a pair of simultaneously measurable observables. This ‘amplification of conditional quasi-expectations’ is a noteworthy property of quantum mechanics, and the oft-discussed ‘amplification of weak values’ could be understood as its special case.

Proposition 4.9 (Range of the Two-state Value). *Let A be a non-trivial observable on \mathcal{H} , and let $|\phi\rangle \in \text{dom}(A)$ be a pre-selected state that is not an eigenvector of A . Then, the two-state value of A under the pre-selected state $|\phi\rangle$ may take any complex number in the case $\alpha \neq 0$, and in turn any real number in the case $\alpha = 0$, given an appropriate choice of the post-selected state $|\phi'\rangle \in \mathcal{H}$.*

Proof. For simplicity, we only provide the proof of the statement for the specific choice $\alpha = 1$ of the ambiguity parameter without loss of generality.

Now, before we go into the main part of the proof, we first observe that, for a non-trivial self-adjoint operator A and a normalised vector $|\phi\rangle \in \text{dom}(A)$, there exists a normalised

vector $|\chi\rangle \in \mathcal{H}$ orthogonal to $|\phi\rangle$ such that

$$A|\phi\rangle = \mathbb{E}[A; \phi] \cdot |\phi\rangle + \|(A - \mathbb{E}[A; \phi])\phi\| \cdot |\chi\rangle \quad (4.84)$$

holds²³. To see this, we first consider the case

$$A|\phi\rangle = \mathbb{E}[A; \phi] \cdot |\phi\rangle, \quad (4.86)$$

that is, when $|\phi\rangle$ is an eigenvector of A . Then, by choosing any normalised state $|\chi\rangle$ satisfying $\langle\chi, \phi\rangle = 0$ (the existence of such $|\chi\rangle$ is guaranteed by the fact $\dim(\mathcal{H}) \geq 2$), one finds that the above equality is fulfilled. Next, suppose that $A|\phi\rangle \neq \mathbb{E}[A; \phi] \cdot |\phi\rangle$. Then, by defining

$$|\chi\rangle := \frac{(A - \mathbb{E}[A; \phi])|\phi\rangle}{\|(A - \mathbb{E}[A; \phi])\phi\|}, \quad (4.87)$$

one indeed learns that $\|\chi\| = 1$ and $\langle\chi, \phi\rangle = 0$ as stated.

Armed with this fact and by fixing such $|\chi\rangle$, we choose the post-selected state as

$$|\phi'\rangle = \frac{1}{c^*}|\phi\rangle + |\chi\rangle \quad (4.88)$$

with a free parameter $c \in \mathbb{C}^\times$. One then finds

$$\begin{aligned} \mathbb{E}^1[A|B=1; \phi] &= \frac{\langle\phi', A\phi\rangle}{\langle\phi', \phi\rangle} \\ &= \mathbb{E}[A; \phi] \cdot \frac{\langle\phi', \phi\rangle}{\langle\phi', \phi\rangle} + \|(A - \mathbb{E}[A; \phi])\phi\| \cdot \frac{\langle\phi', \chi\rangle}{\langle\phi', \phi\rangle} \\ &= \mathbb{E}[A; \phi] + c \|(A - \mathbb{E}[A; \phi])\phi\|. \end{aligned} \quad (4.89)$$

This shows that, for the choice of an initial state $|\phi\rangle \in \text{dom}(A)$ that is not an eigenvector of A (which is always possible due to the non-triviality of A), the weak value (hence, also the two-state value) may indeed take any complex number by adjusting the free parameter c appropriately. ■

The difference between (standard) conditional expectations and conditional quasi-expectations in the behaviour of their essential suprema makes it clear that, conditional quasi-expectations are not conditional expectations in the classical sense. This provides an indirect proof for the fact that, in general, *the ‘joint behaviour’ of the outcomes of the pair of (generally non-commuting) quantum observables A and B does not allow itself to be described by probability spaces*. This would be accounted for in depth in Section 5 and 6 shortly.

4.3.3. Weak Conditioned Measurement. Armed with our newly introduced concept of conditional quasi-expectations (4.77) of a quantum observable given another (not necessarily simultaneously measurable) quantum observable, we shall summarise our findings regarding the first-order local behaviour of the conditional expectation at the origin. Combining Proposition 4.7 and (4.78), one is naturally tempted to conjecture that:

²³ Note that in the case $|\phi\rangle \in \text{dom}(A^2)$, the equality (4.84) is equivalent to

$$A|\phi\rangle = \mathbb{E}[A; \phi] \cdot |\phi\rangle + \sqrt{\mathbb{V}[A; \phi]} \cdot |\chi\rangle, \quad (4.85)$$

since one has $\|(A - \mathbb{E}[A; \phi])\phi\|^2 = \langle\phi, (A - \mathbb{E}[A; \phi])^2\phi\rangle = \mathbb{V}[A; \phi]$ with the variance defined as in (4.53).

Proposition 4.10 (Weak Conditioned Measurement). *Let A and B be self-adjoint operators defined on the target system \mathcal{H} , and let the respective initial states $|\phi\rangle \in \text{dom}(A)$, $|\psi\rangle \in \mathcal{D}$ be fixed. Then, the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ is well-defined for all range of $g \in \mathbb{R}$, and the limit converges to*

$$\begin{aligned} \left. \frac{d}{dg} \mathbb{E}[X|B; \Psi^g] \right|_{g=0} &:= \lim_{g \rightarrow 0} \frac{\mathbb{E}[X|B; \Psi^g] - \mathbb{E}[X|B; \Psi^0]}{g} \\ &= 2 \text{Re} [\mathbb{E}[A|B; \phi]] \cdot \text{CV}_A[X, Y; \psi] \\ &\quad + 2 \text{Im} [\mathbb{E}[A|B; \phi]] \cdot \text{CV}_S[X, Y; \psi] \end{aligned} \quad (4.90)$$

point-wise μ_B^ϕ -almost everywhere.

While we have explicitly proved the above statement only in the special case where B has spectrum of finite cardinality, the same statement indeed holds for general B , although we do not go into the technical details for its demonstration. One may thus understand the process of the weak CM scheme as the practice of measuring (the real and imaginary parts of) the conditional quasi-expectation $\mathbb{E}[A|B; \phi]$ of the target system. This result is to be compared with the unconditioned counterpart, in which one may extract the standard (unconditional) expectation $\mathbb{E}[A; \phi]$ by means of the weak UM scheme from the first-order differential coefficient of the measurement outcomes.

Topic: Conditional Quasi-expectation as the Merkmal for Amplification. Under the above conditions, Taylor's theorem states that one has the following first-order expansion of the conditional expectation

$$\begin{aligned} \mathbb{E}[X|B; \Psi^g] &= \mathbb{E}[X; \psi] \\ &+ g \cdot (2 \text{Re} [\mathbb{E}[A|B; \phi]] \cdot \text{CV}_A[X, Y; \psi] + 2 \text{Im} [\mathbb{E}[A|B; \phi]] \cdot \text{CV}_S[X, Y; \psi]) \\ &+ o(g), \end{aligned} \quad (4.91)$$

where $o(g)$ (Landau symbol) denotes a member of the class of functions satisfying the asymptotic property

$$\lim_{g \rightarrow 0} \frac{o(g)}{|g|} = 0, \quad (4.92)$$

and the equality (4.91) is understood to hold μ_B^ϕ -almost everywhere. The above fact purports that the conditional quasi-expectation $\mathbb{E}[A|B; \phi]$ gives the (best first-order) indicator on the degree of 'amplification' of the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ one may attain by means of choosing the conditioning observable B on the target system. Colloquially speaking, if one hopes to gain large amplification effect by conditioning, the first place one should look for is its conditional quasi-expectation, and one may hopefully achieve it by adjusting the conditioning observable B so that the conditional quasi-expectation $\mathbb{E}[A|B; \phi]$ becomes large enough. However, note here that while the conditional quasi-expectation (for non-trivial A , and in addition, for the choice of the initial state $|\phi\rangle \in \text{dom}(A)$ that is not an eigenvector of A) admits arbitrary large amplification by a suitable choice of the conditioning observable B (Proposition 4.8), the classical conditional expectation $\mathbb{E}[X|B; \Psi^g]$ may have an upper bound depending on its configuration (Proposition 4.5). This generally suggests that the discrepancies between the full-order behaviour of $\mathbb{E}[X|B; \Psi^g]$ and its first-order

approximation becomes larger (in other words, the higher-order terms $o(g)$ becomes more significant) as one adjusts the choice of the conditioning observable B so that the conditional quasi-expectation may become larger. As for the higher-order terms, although we shall omit details, we note that one may also prove higher-order differentiability of the conditional expectation by placing stricter conditions for the choice of both the initial states of the target system and the meter system, and subsequently compute higher-order derivatives through analogous procedure as demonstrated above.

In order to confirm this observation with a concrete model, we have included in Appendix A an analytic example where we compute the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ for the special case in which the conditioning observable $B = |\phi'\rangle\langle\phi'|$ is a projection onto a one-dimensional subspace spanned by a unit vector $|\phi'\rangle \in \mathcal{H}$ (*i.e.*, the post-selected measurement scheme), and moreover the target observable A is dichotomic. One shall indeed find the existence of the limit of ‘amplification’ of the conditional expectation by the ‘weak value amplification’, and various other general properties alongside that we found in the discussions throughout this section.

5. Conditioned Measurement II: In Terms of Conditional Probabilities

In Section 3, we have elaborated the study of the UM scheme conducted in the preceding Section 2 in terms of probabilities. In this section, we follow the same line and intend to refine our analysis for the conditioned counterpart.

Preliminary Observations. As one may recall, we have seen in Section 2 and Section 3 that, by means of the UM scheme, one could extract the information of the target system in both the form of the expectation value $\mathbb{E}[A; \phi]$ and the probability measure μ_A^ϕ , the former by looking at the expectation value of the meter observable X conjugate to Y , whereas the latter by focusing at the probability measure of it, and they were obtained by either inspecting the strong region $g \rightarrow \pm\infty$ of the interaction or by probing its local behaviour at $g = 0$, both in parallel manners.

Now, as for the conditioned case, while we have not looked into the strong region $g \rightarrow \pm\infty$ of the interaction parameter, our analysis on the local behaviour conducted in Section 4 revealed that the first-order derivative of the expectation value for the choice $X = Q, P$ both contain portions (real and imaginary parts) of the conditional quasi-expectation $\mathbb{E}[A|B; \phi]$. By comparing this result to the unconditioned case, one may come to a naïve conjecture that both the expectation value and the conditional quasi-expectation of an observable A has some quality in common. Namely, since the CM scheme incorporate conditioning, one may speculate that the conditional quasi-expectation $\mathbb{E}[A|B; \phi]$ may be interpreted as some form of a ‘conditional average’ with respect to an underlying ‘probability distribution’ of some kind.

Quasi-joint-probability Distributions in Quantum Mechanics. A quick observation on our previous result (4.50) reveals that, the full description of the CM scheme must incorporate the information of the measurement outcomes of *both* the choice $X = Q, P$ of the meter observables, which is in contrast to the unconditioned case where we may concentrate only on the analysis of the probability distribution describing the outcome of a *single* observable X that is conjugate to Y . In view of this, it would thus be natural to consider some form of a ‘joint-distribution’ describing the measurement outcome of both the observables Q and P . However, as we have seen in Section 3.1.7, and also from an indirect proof by observing the difference of conditional (quasi)-expectations in their behaviour regarding essential suprema that, by definition, only a pair of observables that are simultaneously measurable admits a description by joint-probability distributions in the classical sense and, unfortunately, the pair $\{Q, P\}$ of observables of our present interest does not fall into this category.

On account of this, there have been various attempts to construct some alternative form of ‘joint-distributions’ for pairs of (generally non-commuting) quantum observables that possess convenient or desirable properties in describing the behaviour of both their outcomes. The Wigner-Ville distribution (WD distribution) [1, 2], which purports to describe the ‘joint behaviour’ of the otherwise incompatible pair of observables \hat{x} and \hat{p} on the normalised wave-function $\psi \in L^2(\mathbb{R})$, symbolically defined by

$$W^\psi(x, p) := \frac{1}{\pi} \int_{\mathbb{R}} \psi^*(x+y)\psi(x-y)e^{2ipy} d\beta(y), \quad (5.1)$$

and the Kirkwood-Dirac distribution (KD distribution) [3, 4], which on the other hand allows itself to be defined for arbitrary pair of observables A and B , symbolically defined by

$$K_{A,B}^\phi(a, b) := \frac{\langle \phi, b \rangle \langle b, a \rangle \langle a, \phi \rangle}{\|\phi\|^2} \quad (5.2)$$

with the symbolical decomposition $A = \int_{\mathbb{R}} a d|a\rangle\langle a|$, $B = \int_{\mathbb{R}} b d|b\rangle\langle b|$, are among the most well-known classical proposals. The former allows negative numbers to be assigned, whereas the latter even admits complex numbers. Despite their queerness, they both retain some properties that one finds common in the standard (*i.e.*, real and non-negative) joint-probability distributions, *e.g.*, that they both have total integration of unity, and that the marginals coincide with the probability distribution describing the behaviour of the remaining observable, and in this sense, they are occasionally referred to as *quasi-joint-probability* (QJP) distributions of the specific pairs of observables.

Quasi-joint-probability Distributions and Conditional Quasi-expectations. Now, as some may expect, conditional quasi-expectations are closely related to the notion of quasi-joint-probabilities in quantum mechanics. Indeed, a quick observation reveals that, given a symbolical spectral decomposition $B = \int_{\mathbb{R}} b d|b\rangle\langle b|$ of the conditioning observable, the complex-parametrised conditional quasi-expectation (4.77) for the choice $\alpha = 1$ coincides with the, so to speak, ‘conditional average’ of A given the outcome b of B under the Kirkwood-Dirac distribution, as one finds under the formal computation

$$\mathbb{E}^1[A|B = b; \phi] = \frac{\langle \phi, b \rangle \langle b, A\phi \rangle}{|\langle b, \phi \rangle|^2} = \frac{\int_{\mathbb{R}} a K_{A,B}^\phi(a, b) d\beta(a)}{\int_{\mathbb{R}} K_{A,B}^\phi(a, b) d\beta(a)}. \quad (5.3)$$

As for the Wigner-Ville distribution, pure realness of its values might lead one to think that this is in some form related to the parametrised conditional quasi-expectation for the choice $\alpha = 0$. Indeed, one confirms under the formal computation

$$\begin{aligned} \mathbb{E}^0[\hat{p}|\hat{x} = x; \phi] &:= \frac{1}{2} \left[\frac{\langle x, \hat{p}\psi \rangle}{\langle x, \psi \rangle} + \left(\frac{\langle x, \hat{p}\psi \rangle}{\langle x, \psi \rangle} \right)^* \right] \\ &= |\langle x, \psi \rangle|^{-2} \cdot \frac{[\langle \psi, x \rangle \langle x, \hat{p}\psi \rangle + (\langle \psi, x \rangle \langle x, \hat{p}\psi \rangle)^*]}{2} \\ &= |\langle x, \psi \rangle|^{-2} \cdot (-i)^{-1} \frac{d}{dy} \left(\frac{\langle \psi, e^{-iy\hat{p}}x \rangle \langle x, e^{-iy\hat{p}}\psi \rangle}{2} \right) \Big|_{y=0} \\ &= |\langle x, \psi \rangle|^{-2} \cdot (-i)^{-1} \frac{d}{dy} \left(\frac{\psi^*(x+y)\psi(x-y)}{2} \right) \Big|_{y=0} \\ &= |\langle x, \psi \rangle|^{-2} \cdot \frac{1}{2\pi} \int_{\mathbb{R}} p \psi^*(x+y)\psi(x-y) e^{2ipy} d\beta(y) \\ &= \frac{\int_{\mathbb{R}} p W^\psi(x, p) d\beta(p)}{\int_{\mathbb{R}} W^\psi(x, p) d\beta(p)}, \end{aligned} \quad (5.4)$$

that the conditional quasi-expectation of \hat{p} given $\hat{x} = x$ for the choice $\alpha = 0$ coincides with the, again so to speak, ‘conditional average’ of the momentum \hat{p} given the outcome x of the position \hat{x} under the Wigner-Ville distribution.

Conditioned Measurement. The above observation is instructive in guiding the direction of our analysis. Indeed, it would be natural to expect that the measurement of the meter system in view of QJP distributions of the pair of observables $\{Q, P\}$ would allow us to extract the information of the target system in the form that is ‘akin’ to it, *i.e.*, one might hope to obtain a QJP distributions of the target system, of which ‘conditional average’ coincides with the conditional quasi-expectations $\mathbb{E}^\alpha[A|B = b]$ of our interest. Guided by this formal argument and heuristic observation, in this section, we shall be analysing the CM scheme in terms of quasi-probabilities, or more specifically, in terms of ‘conditional’ quasi-probabilities. Now, as our previous arguments (in particular, those developed in Section 3.3.2) indicate, analysis directly on the level of probabilities is better suited to be performed in the space of generalised functions, rather than density functions or measures, if one is to conduct it with decent mathematical rigour and generality. This becomes especially crucial when introducing ‘quasi-joint-probabilities’ of a pair of (generally not necessarily simultaneously measurable) quantum observables, which is one of the main themes of this paper, and thus examined in depth in the next Section 6. However, since the present authors have judged the theory of generalised functions to be beyond the scope of this paper as a tool for analysis, we shall be working exclusively in the space of complex measures and density functions as usual. While this treatment comes with some unavoidable compromise on generality of the results and loss of transparency of the line of arguments, we hope that we may still convey the essence of the contents.

Conditioned Measurement in View of the WV Distributions. In this section, the target of our interest for our measurement is the QJP distribution of the pair of observables Q and P on the meter, and we shall study how one may extract information of the configuration of the target system from this viewpoint. Now, as one may realise from the two concrete classical proposals given above (namely, WV distribution and KD distribution), there exist an indefiniteness/arbitrariness to the choice of such distributions, and by its very nature, one may equally conduct the analysis in view of any of one’s own selection. In this section, we shall be analysing the CM scheme exclusively in terms of the Wigner-Ville distribution. The primary reason for our choice is merely based on its degree of familiarity in the physics community, and as mentioned above, the choice is essentially arbitrary. One may naturally conduct the same type of analysis in view of another type of quasi-probability distribution (*e.g.*, the Kirkwood-Dirac type) in a similar manner and obtain analogous results, or may treat them collectively from a more general viewpoint (more to this in Section 6).

5.1. Reference Materials

As usual, we first make a brief review on the basic concepts and facts that are used in our later discussion.

5.1.1. Conditional Probabilities. We first introduce some basic definitions and results on the topic of conditioning of probability measures and some intricacies inherent to it. Let (X, \mathfrak{A}, μ) be a probability space, and let $B \in \mathfrak{A}$ such that $\mu(B) \neq 0$. For $A \in \mathfrak{A}$, we define the *conditional probability of A given B* by the number

$$\mu(A|B) := \frac{\mu(A \cap B)}{\mu(B)}, \quad A \in \mathfrak{A}. \quad (5.5)$$

It is immediate that the map $\mu(\cdot|B)$ is itself a probability measure satisfying the relation

$$\mu(A \cap B) = \mu(A|B) \cdot \mu(B), \quad A \in \mathfrak{A}. \quad (5.6)$$

Conditional Probability given a Sub- σ -Algebra. We now intend to generalise the elementary definition above to suit our further needs. In parallel to the manner we have done for conditional expectations in the previous section, let $\mathfrak{B} \subset \mathfrak{A}$ be a sub- σ -algebra, and for each measurable set $A \in \mathfrak{A}$, we define the *conditional probability of A given \mathfrak{B}* by

$$\mu(A|\mathfrak{B}) := \mathbb{E}[\chi_A|\mathfrak{B}], \quad A \in \mathfrak{A}, \quad (5.7)$$

where χ_A is the characteristic function (2.8) of A . For fixed $A \in \mathfrak{A}$, note that by definition, the conditional probability (5.7) is understood as an equivalence class of a *family* of $\mu|_{\mathfrak{B}}$ -integrable functions by identifying those that are indistinguishable under the given probability measure $\mu|_{\mathfrak{B}}$. In the simplest case where $\mathfrak{B} = \sigma(B) = \{\emptyset, B, B^c, X\}$ given some $B \in \mathfrak{A}$, the conditional probability $\mu(A|\sigma(B))$ satisfies

$$\begin{aligned} \mu(A \cap B) &= \int_B \chi_A \, d\mu \\ &= \int_B \mathbb{E}[\chi_A|\sigma(B)] \, d\mu|_{\sigma(B)} = \mu(A|\sigma(B)) \cdot \mu(B), \quad A \in \mathfrak{A}. \end{aligned} \quad (5.8)$$

This clarifies the relation between the general definition (5.7) and the elementary definition (5.6).

Conditional Probability given a Function. Now, under the condition above, instead of being given a sub- σ -algebra, suppose that one is given a measurable function $g : X \rightarrow Y$ for conditioning. We thus define

$$\mu(A|g) := \mu(A|\mathcal{I}(g)), \quad A \in \mathfrak{A}, \quad (5.9)$$

to be the conditional probability of A given g , where $\mathcal{I}(g)$ is the initial σ -algebra of g (see (4.9) for its definition), and also introduce

$$\mu(A|g = y) := \mu(A|g)(y), \quad y \in \mathbb{R}, \quad (5.10)$$

of which notation involves subtlety regarding the choice of the representative, in parallel to the situation of conditional expectations we have seen earlier.

Conditional Probabilities as Equivalent Classes of Functions. Given a probability space (X, \mathfrak{A}, μ) and a sub- σ -algebra $\mathfrak{B} \subset \mathfrak{A}$, the conditional probability $\mu(\cdot|\mathfrak{B})$ satisfies properties analogous to those of probability measures, namely

- (i) $\mu(\emptyset|\mathfrak{B}) = 0$, $\mu(X|\mathfrak{B}) = 1$,
- (ii) $\mu(A|\mathfrak{B}) \geq 0$, $A \in \mathfrak{A}$,
- (iii) for any sequence $(A_n)_{n \geq 1}$ of pairwise disjoint subsets of X , the equality

$$\mu \left(\bigcup_{n=1}^{\infty} A_n \middle| \mathfrak{B} \right) = \sum_{n=1}^{\infty} \mu(A_n|\mathfrak{B}) \quad (5.11)$$

holds.

However, the key distinction to be noted between the usual probability measures is that, the above (in)equalities are guaranteed to hold *almost everywhere*, since by definition, conditional probabilities are equivalent classes of functions. It is thus of natural interest whether we could raise the limitation by dropping ‘validity almost everywhere’, which one may occasionally find troublesome.

Transition Kernels. To this end, we first recall the definition of transition kernels. Let (X, \mathfrak{A}) and (Y, \mathfrak{B}) be measurable spaces. We say that a map $K : X \times \mathfrak{B} \rightarrow [0, \infty]$ that satisfies the conditions

- (i) the map $x \mapsto K(x, B)$ is \mathfrak{A} -measurable for every $B \in \mathfrak{B}$,
- (ii) the map $B \mapsto K(x, B)$ is a measure on (Y, \mathfrak{B}) for every $x \in X$,

a *transition kernel* from (X, \mathfrak{A}) into (Y, \mathfrak{B}) . A transition kernel is said to be (σ) -finite if the map $B \mapsto K(x, B)$ is (σ) -finite for all $x \in X$. If K is normalised to unity $K(x, Y) = 1$ for all $x \in X$, we say that K is a *transition probability kernel*. Given a σ -finite transition kernel $K : X \times \mathfrak{B} \rightarrow [0, \infty]$ from (X, \mathfrak{A}) into (Y, \mathfrak{B}) and a function $f \in \mathcal{M}^+(\mathfrak{B})$, the integral

$$(Kf)(x) := \int_Y f(y)K(x, dy), \quad x \in X \quad (5.12)$$

defines a function $Kf \in \mathcal{M}^+(\mathfrak{A})$. On the other hand, given a measure μ on (X, \mathfrak{A}) , the integral

$$(\mu K)(B) := \int_X K(x, B) d\mu(x), \quad B \in \mathfrak{B} \quad (5.13)$$

defines a measure μK on (Y, \mathfrak{B}) . Associative law is valid, which is to say that

$$\begin{aligned} \mu(Kf) &:= \int_X \left(\int_Y f(y)K(x, dy) \right) d\mu(x) \\ &= \int_Y f(y) \left(\int_X K(x, dy) d\mu(x) \right) =: (\mu K)f \end{aligned} \quad (5.14)$$

holds. The following theorem is of much use.

Theorem (Transition Kernels into Measures on Product Spaces). *Let $K : X \times \mathfrak{B} \rightarrow [0, \infty]$ be a σ -finite transition kernel from (X, \mathfrak{A}) into (Y, \mathfrak{B}) , and let μ be a measure on (X, \mathfrak{A}) . Then, there exists a measure π on the product space $(X \times Y, \mathfrak{A} \otimes \mathfrak{B})$ that satisfies*

$$\int_{X \times Y} f(x, y) d\pi(x, y) := \int_X \int_Y f(x, y)K(x, dy) d\mu(x) \quad (5.15)$$

for all $f \in \mathcal{M}^+(\mathfrak{A} \otimes \mathfrak{B})$. If, moreover, both μ and K happens to be finite, then π is the unique finite measure on the product space satisfying

$$\pi(A \times B) = \int_A K(x, B) d\mu(x), \quad A \in \mathfrak{A}, B \in \mathfrak{B}. \quad (5.16)$$

This provides us a convenient way to construct a measure on the product spaces given a transition kernel and a measure.

Conditional Probability Distributions. We now return to our main line of arguments, and first introduce the definition of conditional probability measures.

Definition (Conditional Probability Measure). *Let (X, \mathfrak{A}, μ) be a probability space, and let $\mathfrak{B} \subset \mathfrak{A}$ be a sub- σ -algebra. We call a transition probability kernel $K : X \times \mathfrak{B} \rightarrow [0, 1]$ a conditional probability measure (or a regular version) of the conditional probability $\mu(\cdot | \mathfrak{B})$ given \mathfrak{B} , if the map $x \mapsto K(x, A)$ happens to be a representative of $\mu(A | \mathfrak{B})$ for all $A \in \mathfrak{A}$, namely*

$$K(\cdot, A) \in [\mu(A | \mathfrak{B})], \quad A \in \mathfrak{A} \quad (5.17)$$

holds, where the brackets around an element denote its equivalence class. If such a transition probability kernel exists, we customarily denote it with the same notation $\mu(\cdot | \mathfrak{B})$, and its images are in turn denoted as

$$K(x, A) = \mu(A | \mathfrak{B})(x) = \mu_x(A), \quad x \in X, A \in \mathfrak{A} \quad (5.18)$$

interchangeably, depending on the aesthetics of the formula in which it should appear.

The presence of conditional probability measures allows us to readily make a connection between conditional expectations (defined previously in (4.5)) and averages with respect to conditional probabilities under consideration.

Proposition (Conditional Expectations as Averages over Conditional Probability Measures). *Let (X, \mathfrak{A}, μ) be a probability space, $\mathfrak{B} \subset \mathfrak{A}$ be a sub- σ -algebra, and suppose that the conditional probability $\mu(\cdot | \mathfrak{B})$ has a conditional probability measure. Then, for every μ -integrable function f , the map*

$$x \mapsto \int_X f(x') d\mu_x(x') \in [\mathbb{E}[f | \mathfrak{B}]] \quad (5.19)$$

is a representative of the conditional expectation of f given \mathfrak{B} .

We note that conditional probability measures do not necessarily exist for general measure spaces. However, fortunately for us, the case $(X, \mathfrak{A}) = (\mathbb{R}^n, \mathfrak{B}^n)$ that we are interested in is known to always admit it.

Conditional Probability Distributions. Given a probability space (X, \mathfrak{A}, μ) and a sub- σ -algebra $\mathfrak{B} \subset \mathfrak{A}$, suppose that a measurable map $f : (X, \mathfrak{A}) \rightarrow (X', \mathfrak{A}')$ is moreover given. In parallel to what we have seen for conditional expectations, this allows us to define an equivalence class of functions

$$\mu(f \in A' | \mathfrak{B}) := \mu(f^{-1}(A') | \mathfrak{B}) \quad (5.20)$$

for all $A' \in \mathfrak{A}'$. Then, a transition probability kernel $K : X \times \mathfrak{A}' \rightarrow [0, 1]$ from (X, \mathfrak{B}) into (X', \mathfrak{A}') satisfying

$$K(\cdot, A') \in [\mu(f \in A' | \mathfrak{B})], \quad A' \in \mathfrak{A}' \quad (5.21)$$

is called a *conditional probability distribution of f given \mathfrak{B}* . Likewise, given another measurable map $g : (X, \mathfrak{A}) \rightarrow (Y', \mathfrak{B}')$, a transition probability kernel $K : X \times \mathfrak{A}' \rightarrow [0, 1]$ from $(X, \mathcal{I}(g))$ into (X', \mathfrak{A}') satisfying

$$K(\cdot, A') \in [\mu(f \in A' | \mathcal{I}(g))], \quad A' \in \mathfrak{A}' \quad (5.22)$$

is called a *conditional probability distribution of f given g* . Such transition probability kernels do not necessarily exist in general, but as above, the case $(X', \mathfrak{A}') = (\mathbb{R}^n, \mathfrak{B}^n)$ that we are interested in is known to always admit it.

Conditioning in Quantum Measurements. Under the context of quantum measurements, let A and B be a pair of simultaneously measurable observables. Given a joint-probability distribution $\mu_{A,B}^\phi$ of A and B on some quantum state $|\phi\rangle \in \mathcal{H}$, we introduce

$$\mu_A^\phi(\Delta_A|B) := \mu_{A,B}^\phi(\pi_A \in \Delta_A|\pi_B), \quad \Delta_A \in \mathfrak{B}^1, \quad (5.23)$$

where $\pi_A(a, b) = a$ and $\pi_B(a, b) = b$ are measurable functions (projections) respectively representing the behaviour of the measurement outcomes of A and B . Accordingly, we define the *conditional probability distribution of A given B* to be a transition probability kernel $K : \mathbb{R} \times \mathfrak{B}^1 \rightarrow [0, 1]$ that satisfies

$$K(\cdot, \Delta_A) \in [\mu_A^\phi(\Delta_A|B)], \quad \Delta_A \in \mathfrak{B}^1, \quad (5.24)$$

which, as guaranteed above, is known to always exist. The values of the conditional probability distribution of A given B are in turn denoted interchangeably by

$$\mu_A^\phi(\Delta_A|B = b) = \mu_{B=b}^\phi(A \in \Delta_A) = \mu_A^\phi(\Delta_A|B)(b) := K(b, \Delta_A), \quad (5.25)$$

depending on the context.

5.1.2. Fourier Transformation. We next recall the basic definitions and properties of the Fourier transformation. For convenience, we first introduce the renormalised n -dimensional Lebesgue-Borel measure on \mathbb{R}^n by

$$dm_n := (2\pi)^{-n/2} d\beta^n. \quad (5.26)$$

Accordingly, in this section we employ the renormalised L^p -norm and the convolution defined by the renormalised Lebesgue-Borel measure,

$$\|f\|_p := \left(\int_{\mathbb{R}^n} |f(x)|^p dm_n(x) \right)^{1/p}, \quad f \in L^p(\mathbb{R}^n), \quad (5.27)$$

$$(f * g)(x) := \int_{\mathbb{R}^n} f(x-y)g(y) dm_n(y), \quad f, g \in L^1(\mathbb{R}^n). \quad (5.28)$$

For brevity, we occasionally write $dm_1 = dm$ whenever there is no risk for confusion.

Now, for a function $f \in L^1(\mathbb{R}^n)$, recall that the functions $\hat{f}, \check{f} : \mathbb{R}^n \rightarrow \mathbb{C}$ defined by

$$\hat{f}(q) := \int_{\mathbb{R}^n} e^{-i\langle q, x \rangle} f(x) dm_n(x), \quad (5.29)$$

$$\check{f}(q) := \int_{\mathbb{R}^n} e^{i\langle q, x \rangle} f(x) dm_n(x), \quad (5.30)$$

with the scalar product $\langle q, x \rangle := \sum_{k=1}^n q_k x_k$ of two real vectors in \mathbb{R}^n , are respectively called the *Fourier transform* and the *inverse Fourier transform* of f . The \mathbb{C} -linear map \mathcal{F} that maps f to its Fourier transform \hat{f} is called the *Fourier transformation*. It is known that the Fourier transformation is injective, *i.e.* $\hat{f} = \hat{g}$ implies $f = g$. For $f, g \in L^1(\mathbb{R}^n)$, the following properties under the convolution (5.28), scaling (3.96), and translation (3.119),

$$\widehat{(f * g)} = \hat{f} \cdot \hat{g}, \quad (5.31)$$

$$\widehat{(f_t)}(q) = \hat{f}(tq), \quad t \neq 0, \quad (5.32)$$

$$\widehat{(\tau_a f)}(t) = e^{i\langle a, x \rangle} \hat{f}(t), \quad a \in \mathbb{R}, \quad (5.33)$$

respectively, are basic. The Fourier transformation \mathcal{F} plays particularly well on the subspace $\mathcal{S}(\mathbb{R}^n) \subset L^1(\mathbb{R}^n)$, where it becomes a linear bijection of $\mathcal{S}(\mathbb{R}^n)$ onto $\mathcal{S}(\mathbb{R}^n)$, whose inverse

is given by the inverse Fourier transformation (recall, on the other hand, that one does not necessarily have $\hat{f} \in L^1(\mathbb{R}^n)$ for $f \in L^1(\mathbb{R}^n)$ in general). One then has

$$\partial^\gamma(\mathcal{F}f) = (-i)^{|\gamma|} \mathcal{F}(x^\gamma f), \quad \gamma \in \mathbb{N}_0^n, \quad (5.34)$$

$$\mathcal{F}(\partial^\gamma f) = i^{|\gamma|} q^\gamma \mathcal{F}f, \quad \gamma \in \mathbb{N}_0^n, \quad (5.35)$$

for $f \in \mathcal{S}(\mathbb{R}^n)$, where we have used the multi-index $\gamma := (\gamma_1, \dots, \gamma_n) \in \mathbb{N}_0^n$ as in (2.24) and introduced the shorthand $|\gamma| := \gamma_1 + \dots + \gamma_n$.

5.1.3. Wigner-Ville Distribution. In order to make our line of arguments self-contained in the framework of density functions $L^1(\mathfrak{B})$, we assume $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ throughout this passage. Given such ψ , we define a complex function $\omega^\psi \in L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$ by

$$\tilde{\omega}^\psi(x, y) := \psi^*(x - y/2)\psi(x + y/2), \quad (5.36)$$

and evaluate its total integration as

$$\begin{aligned} \int_{\mathbb{R}^2} \tilde{\omega}^\psi(x, y) \, dm_2(x, y) &= \int_{\mathbb{R}^2} \psi^*(x - y/2)\psi(x + y/2) \, dm_2(x, y) \\ &= \left(\int_{\mathbb{R}} \psi^*(x) \, dm(x) \right) \left(\int_{\mathbb{R}} \psi(y) \, dm(y) \right) \\ &= \left| \int_{\mathbb{R}} \psi(x) \, dm(x) \right|^2. \end{aligned} \quad (5.37)$$

Whenever the total integration (5.37) is non-vanishing, we introduce

$$\omega^\psi(x, y) := \frac{\tilde{\omega}^\psi(x, y)}{\int_{\mathbb{R}^2} \tilde{\omega}^\psi(x, y) \, dm_2(x, y)}, \quad (5.38)$$

to denote its normalisation.

On the other hand, if we consider the Fourier transform of $\tilde{\omega}^\psi(x, y)$ with respect to its second parameter y ,

$$\begin{aligned} \tilde{W}^\psi(x, p) &:= \int_{\mathbb{R}} e^{-ipy} \tilde{\omega}^\psi(x, y) \, dm(y) \\ &= \int_{\mathbb{R}} \psi^*(x + y/2)\psi(x - y/2) e^{ipy} \, dm(y), \end{aligned} \quad (5.39)$$

we readily find that it is a real function,

$$\left(\tilde{W}^\psi(x, p) \right)^* = \tilde{W}^\psi(x, p), \quad (5.40)$$

whose marginals are given by

$$\begin{aligned} \int_{\mathbb{R}} \tilde{W}^\psi(x, p) \, dm(x) &= |\hat{\psi}(p)|^2, \\ \int_{\mathbb{R}} \tilde{W}^\psi(x, p) \, dm(p) &= |\psi(x)|^2. \end{aligned} \quad (5.41)$$

Applying Plancherel's theorem, one finds that $\tilde{W}^\psi \in L^1(m_2)$, and thus its total integration reads

$$\int_{\mathbb{R}^2} \tilde{W}^\psi(x, p) \, dm_2(x, p) = \|\psi\|_2^2. \quad (5.42)$$

If the total integration (5.42) is non-vanishing, which is equivalent to the condition $\psi \neq 0$, the real quasi-probability density function denoted by

$$W^\psi(x, p) := \frac{\tilde{W}^\psi(x, p)}{\int_{\mathbb{R}^2} \tilde{W}^\psi(x, p) dm_2(x, p)} \quad (5.43)$$

is called the *Wigner-Ville distribution* on ψ . As we have seen in (5.41), the WV distribution possesses useful properties for our analysis, namely, that its marginals yield the probability density function describing the behaviour of the measurement outcomes of the respective observables \hat{x} and \hat{p} on the state ψ , which is to say that

$$\begin{aligned} \int_{\mathbb{R}^2} W^\psi(x, p) dm(x) &= \rho_{\hat{p}}^\psi(p), \\ \int_{\mathbb{R}^2} W^\psi(x, p) dm(p) &= \rho_{\hat{x}}^\psi(x), \end{aligned} \quad (5.44)$$

if explicitly written down. Thus, the choice $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ defines a complex measure

$$\mu_{Q,P}^\psi := W^\psi \odot \beta^2 \quad (5.45)$$

on the measurable space $(\mathbb{R}, \mathfrak{B}^2)$ that satisfies

$$\begin{aligned} \mu_{Q,P}^\psi(\mathbb{R} \times \Delta) &= \mu_P^\psi(\Delta), \\ \mu_{Q,P}^\psi(\Delta \times \mathbb{R}) &= \mu_Q^\psi(\Delta), \end{aligned} \quad \Delta \in \mathfrak{B}. \quad (5.46)$$

For our later argument we note that, since the functions $\omega^\psi(x, y)$ and $W^\psi(x, y)$ are mapped to one another by Fourier transformation, they just represent the same contents seen from different viewpoints, and are thus essentially the same object.

5.2. Conditioned Measurement

We are now interested in simultaneously measuring the probability measure of B on the target system and a QJP distribution of Q and P on the meter system. This should be possible since every local measurements can be simultaneously performed on separate systems, and this leads to an existence of a joint distribution of the probability measure of B on one side, and a QJP distribution of Q and P on the other. Throughout this section, for definiteness, we exclusively treat the special case in which the meter state is described by the one-dimensional Schrödinger representation of the CCR $\{L^2(\mathbb{R}), \mathcal{S}(\mathbb{R}), \{\hat{x}, \hat{p}\}\}$ and choose $Y = \hat{p}$ without loss of generality.

5.2.1. Conditioning over Quasi-probabilities. Since we are now dealing with complex measures, the definitions for conditioning must be suitably expanded accordingly. To this end, we first prepare a terminology:

Definition (Quasi-probabilities). *Let (X, \mathfrak{A}) be a measurable space. We call a complex measure ν on (X, \mathfrak{A}) satisfying the normalisation condition $\nu(X) = 1$ a quasi-probability measure, and accordingly the triplet (X, \mathfrak{A}, ν) , a quasi-probability space.*

If the underlying space is given by $(X, \mathfrak{A}) = (\mathbb{R}^n, \mathfrak{B}^n)$, and the quasi-probability measure ν happens to be absolutely continuous, we call its density $d\nu/d\beta^n \in L^1(\mathbb{R}^n)$, which is in general a complex function that has the total integration of unity, a *quasi-probability density function*.

According to the definition, note that the usual (*i.e.*, real and non-negative) probability measures and density functions are special members of the respective families of quasi-probability measures and density functions. In analogy to the standard probability spaces, given a quasi-probability space (X, \mathfrak{A}, ν) and a ν -integrable function f , we occasionally denote the total integration by

$$\mathbb{E}[f; \nu] := \int_X f \, d\nu, \quad (5.47)$$

and call it the *quasi-expectation value of f under ν* .

Quasi-joint-probabilities. As a special subclass of quasi-probability measures, we say that a quasi-probability measure $\nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ *qualifies as a QJP distribution* of the observables A_1, \dots, A_n on the state $|\phi\rangle \in \mathcal{H}$, if it satisfies

$$\nu(\underbrace{\mathbb{K} \times \dots \times \mathbb{K} \times \Delta}_{n} \times \mathbb{K} \times \dots \times \mathbb{K}) = \mu_{A_k}^{\phi}(B), \quad \Delta \in \mathfrak{B}(\mathbb{K}) \quad (5.48)$$

for all $1 \leq k \leq n$. In parallel to it, we prepare the term *QJP density function* for those ν that are absolutely continuous²⁴. One confirms from (5.46) that, for the choice of the quantum state $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, the quasi-probability measure (5.45) qualifies as a QJP distribution for the pair of observables Q and P . It should be intuitively straightforward to see by the formal arguments made in the introduction that the Kirkwood-Dirac distribution also qualifies as a QJP distribution of the pair of observables under consideration.

Conditional Quasi-expectations. We next intend to introduce analogous definitions regarding conditioning on quasi-probability measure spaces (X, \mathfrak{A}, ν) . To this end, we make some very important remarks on the different properties between standard probability measures and quasi-probability measures. Recall that we have made extensive use of the Radon-Nikodým theory for defining conditional expectations and conditional probabilities. In applying the theory, first note that positiveness of the measure μ is necessary in order for the Radon-Nikodým derivative $d\nu/d\mu$ of some complex measure $\nu \ll \mu$ to be well-defined. Hence, conditioning by a sub- σ -algebra $\mathfrak{B} \subset \mathfrak{A}$ must be such that the restriction $\nu|_{\mathfrak{B}}$ becomes a measure. The second fact to notice is that, for a ν -integrable function f , the complex measure on the sub- σ -algebra defined by

$$B \mapsto (f \odot \nu)(B) := \int_B f \, d\nu, \quad B \in \mathfrak{B} \quad (5.49)$$

is not necessarily absolutely continuous with respect to the restriction $\nu|_{\mathfrak{B}}$, in contrast to that of positive measures. With these in mind, we hereby define:

Definition (Conditional Quasi-expectation). *Let (X, \mathfrak{A}, ν) be a quasi-probability space, and $\mathfrak{B} \subset \mathfrak{A}$ a sub- σ -algebra such that the restriction $\nu|_{\mathfrak{B}}$ becomes a probability measure (*i.e.*,*

²⁴ Here, we occasionally admit complex parameters to describe outcomes of each observable A_k for formal completeness. Accordingly, the r. h. s. of the above formula is understood as the probability measure induced by the two-dimensional spectral measure of A_k seen as a normal operator (*cf.* spectral theorem for normal operators).

real and non-negative). For a ν -integrable function f such that $f \odot \nu \ll \nu|_{\mathfrak{B}}$, we define the conditional quasi-expectation of f given \mathfrak{B}

$$\mathbb{E}[f|\mathfrak{B}] := \frac{d(f \odot \nu)}{d(\nu|_{\mathfrak{B}})} \quad (5.50)$$

by the Radon-Nikodým derivative of the complex measure $f \odot \nu$ with respect to the measure $\nu|_{\mathfrak{B}}$.

Given another measurable function $g : X \rightarrow \mathbb{R}$ such that the above conditions are fulfilled for the initial σ -algebra $\mathfrak{B} = \mathcal{I}(g)$, we define $\mathbb{E}[f|g]$ and any other relevant notations such as $\mathbb{E}[f|g = y]$ etc. in an analogous manner to those defined for standard probability measures.

Conditional Quasi-probabilities. We then intend introduce a complex analogue of conditional probabilities defined for quasi-probability measures.

Definition (Quasi-Conditional Probabilities). *Let (X, \mathfrak{A}, ν) be a quasi-probability space, and $\mathfrak{B} \subset \mathfrak{A}$ a sub- σ -algebra such that the restriction $\nu|_{\mathfrak{B}}$ becomes a probability measure. For a measurable set $A \in \mathfrak{A}$, we define*

$$\nu(A|\mathfrak{B}) := \mathbb{E}[\chi_A|\mathfrak{B}], \quad A \in \mathfrak{A}, \quad (5.51)$$

to be the quasi-conditional probability of A given \mathfrak{B} , whenever $\chi_A \odot \nu \ll \nu|_{\mathfrak{B}}$, where χ_A is the characteristic function of A . Likewise, given a measurable function $f : (X, \mathfrak{A}) \rightarrow (X', \mathfrak{A}')$, we introduce

$$\nu(f \in A'|\mathfrak{B}) := \nu(f^{-1}(A')|\mathfrak{B}), \quad A' \in \mathfrak{A}', \quad (5.52)$$

whenever the r. h. s. is well-defined. If, instead of being given a sub- σ -algebra, one is given a measurable function $g : X \rightarrow Y$ for conditioning, we define

$$\nu(A|g) := \nu(A|\mathcal{I}(g)), \quad A \in \mathfrak{A}, \quad (5.53)$$

where $\mathcal{I}(g)$ is the initial σ -algebra of g , whenever, as usual, the r. h. s. is well-defined.

The conditional quasi-probability $\nu(\cdot|\mathfrak{B})$ satisfies properties analogous to those of quasi-probability measures, namely

- (i) $\nu(\emptyset|\mathfrak{B}) = 0$, $\nu(X|\mathfrak{B}) = 1$,
- (ii) $\nu(A|\mathfrak{B}) \in \mathbb{C}$, $A \in \mathfrak{A}$,
- (iii) for any sequence $(A_n)_{n \geq 1}$ of pairwise disjoint subsets of X , the equality

$$\nu\left(\bigcup_{n=1}^{\infty} A_n \middle| \mathfrak{B}\right) = \sum_{n=1}^{\infty} \nu(A_n|\mathfrak{B}) \quad (5.54)$$

holds,

whenever every component above is well-defined. In parallel to conditional probabilities, the validity of the (in)equalities above are significant only in the sense of $\nu|_{\mathfrak{B}}$ -a.e.

Conditional Quasi-probability Measures. We now expand the definition of transition kernels to fit into the theory of complex measures. Let (X, \mathfrak{A}) and (Y, \mathfrak{B}) be measurable spaces. We say that a map $K : X \times \mathfrak{B} \rightarrow \mathbb{C}$ that satisfies the conditions

- (i) the map $x \mapsto K(x, B)$ is \mathfrak{A} -measurable for every $B \in \mathfrak{B}$,

(ii) the map $B \mapsto K(x, B)$ is a complex measure on (Y, \mathfrak{B}) for every $x \in X$,

a *complex transition kernel* from (X, \mathfrak{A}) into (Y, \mathfrak{B}) . If a complex transition kernel K satisfies $K(x, Y) = 1$ for all $x \in X$, we call such K a *transition quasi-probability kernel*. The following analogous result is of use.

Proposition 5.1 (Complex Transition Kernels into Complex Measures on Product Spaces). *Let $K : X \times \mathfrak{B} \rightarrow \mathbb{C}$ be a complex transition kernel from (X, \mathfrak{A}) into (Y, \mathfrak{B}) , and let μ be a measure on (X, \mathfrak{A}) . Then, there exists a complex measure π on the product space $(X \times Y, \mathfrak{A} \otimes \mathfrak{B})$ that satisfies*

$$\int_{X \times Y} f(x, y) d\pi(x, y) := \int_X \int_Y f(x, y) K(x, dy) d\mu(x) \quad (5.55)$$

for all f , whenever the integration on the r. h. s. is well-defined. In particular, the complex measure π satisfies

$$\pi(A \times B) = \int_A K(x, B) d\mu(x), \quad A \in \mathfrak{A}, B \in \mathfrak{B}. \quad (5.56)$$

Armed with the above concepts, we thus introduce:

Definition (Conditional Quasi-Probability Measure). *Let (X, \mathfrak{A}, ν) be a quasi-probability space, and let $\mathfrak{B} \subset \mathfrak{A}$ be a sub- σ -algebra such that the restriction $\nu|_{\mathfrak{B}}$ becomes a probability measure, and that $\nu(A|\mathfrak{B})$ is well-defined for all $A \in \mathfrak{A}$. We call a transition quasi-probability kernel $K : X \times \mathfrak{B} \rightarrow \mathbb{C}$ a *conditional quasi-probability measure of the conditional quasi-probability $\nu(\cdot|\mathfrak{B})$* , if the map $x \mapsto K(x, A)$ happens to be a representative of $\nu(A|\mathfrak{B})$ for all $A \in \mathfrak{A}$, namely*

$$K(\cdot, A) \in [\nu(A|\mathfrak{B})], \quad A \in \mathfrak{A} \quad (5.57)$$

holds, where the brackets around an element denote its equivalence class. If such a transition quasi-probability kernel exists, we customarily denote it with the same notation $\nu(\cdot|\mathfrak{B})$, and its images are in turn interchangeably denoted by

$$K(x, A) = \nu(A|\mathfrak{B})(x) = \nu_x(A), \quad x \in X, A \in \mathfrak{A}, \quad (5.58)$$

depending on the aesthetics of the formula in which it should appear.

As above, such transition quasi-probability kernels do not exist in general, while the case $(X, \mathfrak{A}) = (\mathbb{R}^n, \mathfrak{B}^n)$ is known to always admit it. We then have:

Proposition (Conditional Quasi-expectations as Averages over Conditional Quasi-probability Measures). *Let (X, \mathfrak{A}, ν) be a quasi-probability space, $\mathfrak{B} \subset \mathfrak{A}$ be a sub- σ -algebra such that the restriction $\nu|_{\mathfrak{B}}$ becomes a probability measure, and suppose that the conditional quasi-probability $\nu(\cdot|\mathfrak{B})$ has a conditional quasi-probability measure. Then, for every ν -integrable function f , the map*

$$x \mapsto \int_X f(x') d\nu_x(x') \in [\mathbb{E}[f|\mathfrak{B}]] \quad (5.59)$$

is a representative of the conditional quasi-expectation of f given \mathfrak{B} .

Conditional Probability Distributions. On a quasi-probability space (X, \mathfrak{A}, ν) , suppose that a measurable map $f : (X, \mathfrak{A}) \rightarrow (X', \mathfrak{A}')$ is moreover given. Choosing a sub- σ -algebra

$\mathfrak{B} \subset \mathfrak{A}$ such that the restriction $\nu|_{\mathfrak{B}}$ is a measure, this allows us to define an equivalence class of functions

$$\nu(f \in A' | \mathfrak{B}) := \nu(f^{-1}(A') | \mathfrak{B}) \quad (5.60)$$

for all $A' \in \mathfrak{A}'$, whenever they are well-defined. Then, a transition quasi-probability kernel $K : X \times \mathfrak{A}' \rightarrow [0, 1]$ from (X, \mathfrak{B}) into (X', \mathfrak{A}') satisfying

$$K(\cdot, A') \in [\mu(f \in A' | \mathfrak{B})], \quad A' \in \mathfrak{A}' \quad (5.61)$$

is called a *conditional quasi-probability distribution of f given \mathfrak{B}* . Likewise, given another measurable map $g : (X, \mathfrak{A}) \rightarrow (Y', \mathfrak{B}')$ such that the restriction of ν over its initial σ -algebra $\mathcal{I}(g)$ is a measure, a transition probability kernel $K : X \times \mathfrak{A}' \rightarrow [0, 1]$ from $(X, \mathcal{I}(g))$ into (X', \mathfrak{A}') satisfying

$$K(\cdot, A') \in [\mu(f \in A' | \mathcal{I}(g))], \quad A' \in \mathfrak{A}' \quad (5.62)$$

is called a *conditional quasi-probability distribution of f given g* .

5.2.2. Conditioned Measurement via the WV Distributions. Now that we have prepared the necessary concepts and results, we may embark on our analysis. By measuring B locally on the target system on one side, and a specific QJP distribution of Q and P locally on the meter system on the other, we obtain a quasi-probability distribution that describes the joint behaviour of the target system and the meter system. If, by haps (*e.g.* by choosing the right initial state $|\psi\rangle \in \mathcal{K}$) the QJP distribution of Q and P on the meter admits representation by a complex measure, the total quasi-probability distribution of both the target and the meter system also admits representation by a complex measure. We thus generally define the CM scheme as an act of measuring the conditional quasi-probability distribution of the ‘joint outcome’ of Q and P of the meter system given the outcome of the conditioning observable B on the target system.

WV Distribution. To demonstrate our point with an example, we shall from now on exclusively concentrate on the *Wigner-Ville distribution* for our choice of the QJP distribution of Q and P for definiteness. Since the choice $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ of the initial meter state allows the WV distribution to be described by quasi-probability measures on $(\mathbb{R}^2, \mathfrak{B}^2)$, we assume such special choice throughout this passage in order to remain contained in the framework of measure and integration theory (so that we may not have to deal with the theory of generalised functions). In this subsection, the CM scheme is studied in view of the WV distribution. We first start by transcribing the CM scheme, which was initially introduced in terms of vectors and operators on Hilbert spaces, into the description by quasi-probability density functions on \mathbb{R}^2 . It is then found that the transcription allows a much simpler expression in view of its Fourier transform (rather than the WV distribution itself), in which the description of the meter system after the interaction is given precisely by the convolution of the configuration of both the meter and the target system, quite analogous to the case of the UM scheme that we have previously seen. This allows us to extract the information of the target system either by means of deconvolution discussed earlier (specifically by constructing an approximate identity on the meter system), or by probing the behaviour of the distribution around the origin $g = 0$. We shall then investigate the properties of the information of the target system we have just obtained, and find that this qualifies as a ‘conditional

quasi-probability distribution of A given B , of which the average has a connection to the conditional quasi-expectation of A given B introduced earlier.

Preliminary Observation. As a preliminary observation, we start by assuming that the target observable A has a spectrum consisting of a finite number of eigenvalues $\sigma(A) = \{a_1, \dots, a_N\}$ so that its spectral decomposition reads (2.76). For the ease of arguments, we further assume that the conditioning observable B also has a spectrum consisting of a finite number of eigenvalues $\sigma(B) = \{b_1, \dots, b_M\}$, that every eigenvalue of B is degenerate, *i.e.*, $\Pi_{b_m} = |b_m\rangle\langle b_m|$ for some normalised vectors $|b_m\rangle \in \mathcal{H}$ for all $1 \leq m \leq M$, and moreover that $\|\Pi_b \phi\|^2 \neq 0$ for all $b \in \sigma(B)$. As for the state preparation, let $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ be a wavefunction of the meter system with normalisation $\|\psi\|_2 = 1$ so that the WV distribution can be represented by a quasi-probability density function, and we also let the initial selection $|\phi\rangle \in \mathcal{H}$ of the target system be normalised $\|\phi\| = 1$.

Computing the WV Distribution. We are now interested in measuring the WV distribution of the meter system given the outcome of B on the target system. Since both the measurements are local measurements performed on the respective systems, this should be statistically equivalent to measuring the WV distribution for the meter state

$$\psi_{B=b}^g := \text{Tr}_{\mathcal{H}} [|\Psi_{B=b}^g\rangle\langle\Psi_{B=b}^g|] \quad (5.63)$$

for all $b \in \sigma(B)$, where

$$|\Psi_{B=b}^g\rangle := \frac{(\Pi_b \otimes I)|\Psi^g\rangle}{\|(\Pi_b \otimes I)\Psi^g\|^2} \quad (5.64)$$

is the, so-to-speak, ‘conditional’ meter state²⁵ given the outcome b of B . Our analysis thus reduces to computing the WV distribution of the density operator $\psi_{B=b}^g$ for each of the outcomes $b \in \sigma(B)$. In our case, in which we assume that the eigenvalues of B are all degenerate, the density operator (5.63) in fact becomes a pure state, of which representation by wave-functions reads

$$\begin{aligned} \psi_{B=b}^g(x) &= \sum_{n=1}^N \frac{\langle b, \Pi_{a_n} \phi \rangle}{\langle b, \phi \rangle} \left(e^{-iga_n \hat{p}} \psi \right) (x) \\ &= \sum_{n=1}^N \frac{\langle b, \Pi_{a_n} \phi \rangle}{\langle b, \phi \rangle} \psi(x - ga_n) \\ &= \sum_{n=1}^N \frac{\langle b, E_A(\{a_n\}) \phi \rangle}{\langle b, \phi \rangle} \int_{\mathbb{R}} \psi(x - ga) d\delta_{a_n}(a) \\ &= \int_{\mathbb{R}} \psi(x - ga) d\nu_b(a), \quad g \in \mathbb{R}, \end{aligned} \quad (5.65)$$

where we have used (3.90) to obtain the first equality. Here, we have introduced an auxiliary quasi-probability measure

$$\nu_b(\Delta) := \frac{\langle b, E_A(\Delta) \phi \rangle}{\langle b, \phi \rangle}, \quad b \in \sigma(B), \quad \Delta \in \mathfrak{B}, \quad (5.66)$$

²⁵ Naturally, (5.64) and (5.63) are nothing but the state one would expect when the ideal measurement of B yielded the outcome $b \in \sigma(B)$, if one adopted the standard von Neumann projection postulate.

defined by means of the spectral measure E_A of A , the initial state $|\phi\rangle \in \mathcal{H}$ of the target system, and the outcome $b \in \sigma(B)$ of the conditioning observable, and have used a result analogous to (3.64) in the last equality. One then finds that the WV distribution of the meter wave-function (5.65) reads

$$\begin{aligned}
W^{\psi_{B=b}^g}(x, p) &:= \int_{\mathbb{R}} (\psi_{B=b}^g(x + y/2))^* \psi_{B=b}^g(x - y/2) e^{ipy} dm(y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \psi^*(x - ga'_1 + y/2) d\nu_b^*(a'_1) \right) \left(\int_{\mathbb{R}} \psi(x - ga'_2 - y/2) d\nu_b(a'_2) \right) e^{ipy} dm(y) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^2} \psi^*(x - ga'_1 + y/2) \psi(x - ga'_2 - y/2) d(\nu_b^* \otimes \nu_b)(a'_1, a'_2) \right) e^{ipy} dm(y), \quad (5.67)
\end{aligned}$$

where we have introduced the product measure $\nu_b^* \otimes \nu_b$ of ν_b and its complex conjugate²⁶ in the last equality. In order to gain a better view of our findings, let us now change variables according to the linear transformation,

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = T \begin{pmatrix} a'_1 \\ a'_2 \end{pmatrix}, \quad T := \begin{pmatrix} 1/2 & 1/2 \\ -1/2 & 1/2 \end{pmatrix}. \quad (5.69)$$

Since $T \in \text{GL}(2, \mathbb{R})$ belongs to the general linear group, for indeed $\det T = 1/2$, note that this transformation is invertible, *i.e.*, it is a linear automorphism. We then introduce the quasi-probability measure

$$\begin{aligned}
\mu_A^\phi(\Delta|B=b) &:= T(\nu_b^* \otimes \nu_b)(\Delta) \\
&:= (\nu_b^* \otimes \nu_b)(T^{-1}\Delta), \quad \Delta \in \mathfrak{B}^2, b \in \sigma(B) \quad (5.70)
\end{aligned}$$

defined on the measurable space $(\mathbb{R}^2, \mathfrak{B}^2)$ as the image measure (*cf.* see (3.9) for the definition of image measures) of the product complex measure with respect to the automorphism T (we shall be shortly returning to the properties of the quasi-probability measure (5.70) and the righteousness of its notation). Then, due to the change of variables formula (3.10), one

²⁶ For a pair of complex measures μ and ν , by observing that $\mu \ll |\mu|$ and $\nu \ll |\nu|$, we define the *product complex measure* of μ and ν by

$$\mu \otimes \nu := \left(\frac{d\mu}{d|\mu|} \cdot \frac{d\nu}{d|\nu|} \right) \odot (|\mu| \otimes |\nu|). \quad (5.68)$$

By definition, product complex measures share properties similar to those of product measures (3.31), and an analogue of Fubini's theorem holds. Product complex measures reduce to the usual product measures when both of the components happen to be finite measures.

may rewrite our previous findings (5.67) by letting $a'_1 = a_1 - a_2$ and $a'_2 = a_1 + a_2$ as

$$\begin{aligned}
& W^{\psi_{B=b}^g}(x, p) \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}^2} \psi^*(x - g(a_1 - a_2) + y/2) \right. \\
&\quad \left. \times \psi(x - g(a_1 + a_2) - y/2) d\mu_A^\phi(a_1, a_2|B=b) \right) e^{ipy} dm(y) \\
&= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}} \psi^*((x - ga_1) + (y + 2ga_2)/2) \right. \\
&\quad \left. \times \psi((x - ga_1) - (y + 2ga_2)/2) e^{ipy} dm(y) \right) d\mu_A^\phi(a_1, a_2|B=b) \\
&= \int_{\mathbb{R}^2} e^{-i2ga_2p} \left(\int_{\mathbb{R}} \psi^*((x - ga_1) + y/2) \right. \\
&\quad \left. \times \psi((x - ga_1) - y/2) e^{ipy} dm(y) \right) d\mu_A^\phi(a_1, a_2|B=b) \\
&= \int_{\mathbb{R}^2} e^{-i2ga_2p} W^\psi(x - ga_1, p) d\mu_A^\phi(a_1, a_2|B=b), \tag{5.71}
\end{aligned}$$

where the change of the order of the integration in the second equality is guaranteed by the Fubini's theorem. For later convenience, we introduce the complex number $a \in \mathbb{C}$ defined by $a := a_1 + ia_2$ by identifying $\mathbb{C} \cong \mathbb{R}^2$ in a usual manner, and write

$$W^{\psi_{B=b}^g}(x, p) = \int_{\mathbb{C}} e^{-i2ga_2p} W^\psi(x - ga_1, p) d\mu_A^\phi(a|B=b), \quad g \in \mathbb{R}. \tag{5.72}$$

To sum up, here we have learned how the CM scheme may be rewritten in terms of quasi-probability measures, in which the WV distribution of the initial meter wave-function ψ is acted upon by the quasi-probability measure (5.70) of the target system to yield the final WV distribution of the meter wave-function $\psi_{B=b}^g$.

Changing the Viewpoint through Fourier Transformation. One finds below that the transcription (5.72) of the CM scheme admits a much simpler expression when described in terms of the inverse Fourier transform (5.38) of the WV distribution, rather than the WV distribution itself. Introducing the (yet to be normalised) function $\tilde{\omega}^{\psi_{B=b}^g}(x, y)$ uniquely specified through the relation

$$W^{\psi_{B=b}^g}(x, p) = \int_{\mathbb{R}} e^{-ipy} \tilde{\omega}^{\psi_{B=b}^g}(x, y) dm(y), \quad g \in \mathbb{R} \tag{5.73}$$

(*cf.*, injectivity of the Fourier transformation), the goal of this small paragraph is to show that our finding (5.72) is equivalent to

$$\tilde{\omega}^{\psi_{B=b}^g}(x, y) = \int_{\mathbb{C}} \tilde{\omega}^\psi(x - ga_1, y - 2ga_2) d\mu_A^\phi(a|B=b), \quad g \in \mathbb{R}, \tag{5.74}$$

which is essentially nothing but the convolution of the initial profile $\tilde{\omega}^\psi$ of the meter state by that of the two-dimensional quasi-probability measure $\Delta \mapsto \mu_A^\phi(\Delta|B=b)$ scaled by g . If, moreover, the total integration of $\tilde{\omega}^\psi$ happens to be non-vanishing, we may renormalise both

sides of the above equality to obtain

$$\omega^{\psi_{B=b}^g}(x, y) = \int_{\mathbb{C}} \omega^{\psi}(x - ga_1, y - 2ga_2) d\mu_A^{\phi}(a|B = b), \quad g \in \mathbb{R}, \quad (5.75)$$

for later use²⁷. Observe here the analogy between the unconditioned case (3.99): in both cases, the profile of the ‘output’ of the meter is given by the convolution of the profile of the ‘input’ of the meter and that of the target system scaled by g .

To verify our statement, one may simply repeat the previous argument to obtain the result directly, but it is actually easier to demonstrate that the Fourier transforms of the two sides of the above equality coincide. Indeed, the Fourier transform of the l. h. s. is nothing but $W^{\psi_{B=b}^g}$, which is just the definition (5.73). As for the r. h. s., one has

$$\begin{aligned} & \int_{\mathbb{R}} e^{-ipy} \left(\int_{\mathbb{C}} \tilde{\omega}^{\psi}(x - ga_1, y - 2ga_2) d\mu_A^{\phi}(a|B = b) \right) dm(y) \\ &= \int_{\mathbb{C}} \left(\int_{\mathbb{R}} e^{-ipy} \tilde{\omega}^{\psi}(x - ga_1, y - 2ga_2) dm(y) \right) d\mu_A^{\phi}(a|B = b) \\ &= \int_{\mathbb{C}} e^{-i2ga_2p} W^{\psi}(x - ga_1, p) d\mu_A^{\phi}(a|B = b), \end{aligned} \quad (5.77)$$

where the exchange of the order of the integration (the first equality) is guaranteed by Fubini’s theorem, and the last equality is due to (5.33). Combining the above two results and by observing (5.72), the injectivity of the Fourier transformation leads to the desired statement. We emphasise again that both (5.72) and (5.75) represent the same contents seen from different viewpoints.

5.3. Recovery of the Target Profile

We are now interested in how one may recover the profile $\Delta \mapsto \mu_A^{\phi}(\Delta|B = b)$ of the target system for each $b \in \sigma(B)$ through CM scheme. As one may expect, the procedure essentially goes analogously to that of the recovery of the probability measure μ_A^{ϕ} in the case of the UM scheme demonstrated in Section 3.2. Recalling the techniques employed there, and by introducing the rescaling

$$v^{\psi}(x, y) := 2^{-1}\omega^{\psi}(x, 2y), \quad (5.78)$$

for the ease of discussion, one may readily rewrite (5.75) into

$$v^{\psi_{B=b}^g} = v^{\psi} * \left(\mu_A^{\phi}(\cdot |B = b) \right)_g, \quad g \in \mathbb{R}, \quad (5.79)$$

or equivalently

$$v_{g^{-1}}^{\psi_{B=b}^g} = v_{g^{-1}}^{\psi} * \mu_A^{\phi}(\cdot |B = b), \quad g \in \mathbb{R}^{\times} \quad (5.80)$$

where the subscript on the respective quasi-probability measures/density functions denotes the scaling (3.94) and (3.96), just as we have done for the case of the UM scheme (see (3.99) and (3.100)). In parallel to the case of the UM case, these two expressions (5.79) and (5.80)

²⁷ Here, note that we have used the general property of convolutions

$$\int_{\mathbb{R}^n} (f * g) d\beta^n = \int_{\mathbb{R}^n} f d\beta^n \cdot \int_{\mathbb{R}^n} g d\beta^n, \quad f, g \in L^1(\mathbb{R}^n) \quad (5.76)$$

regarding integration.

correspond to the manner in which one combines the interaction parameter (3.89), where the former corresponds to the scaling of the target observable $A \rightarrow gA$, whereas the latter corresponds to the scaling of the pair of the meter observables $\{Q, P\} \rightarrow \{g^{-1}Q, gP\}$ (cf. (3.85) and (3.88)).

5.3.1. Strong Conditioned Measurement. We now intend to recover the quasi-probability measure $\Delta \mapsto \mu_A^\phi(\Delta|B=b)$ by making use of the latter expression (5.80). The idea and the procedure are essentially the same as those we have employed in the unconditional case, namely, we manipulate both the interaction parameter $g \in \mathbb{R}^\times$ and the initial meter state $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ so that the scaling $v_{g^{-1}}^\psi$ of the inverse Fourier transform of the WV distribution tends towards the delta measure δ_0 centred at the origin $0 \in \mathbb{R}^2$.

Recovery of the Conditional Quasi-joint-probability. For the same reason discussed in Section 3.3.1, we assume throughout this passage:

- The target observable A admits description by density functions.
- The total integration of $\tilde{\omega}^\psi$ is non-vanishing.

The first condition guarantees that the quasi-probability measure $\Delta \mapsto \mu_A^\phi(\Delta|B=b)$, $\Delta \in \mathfrak{B}^2$ is absolutely continuous for all $b \in \sigma(B)$, of which density we shall write

$$\rho_A^\phi(\cdot|B=b) := \frac{d\mu_A^\phi(\cdot|B=b)}{d\beta^2}. \quad (5.81)$$

The last condition is necessary in order to assure the well-definedness of ω^ψ . Then, one sees from an analogous argument that we have previously made in Section 3.3.1 that, if one adjusts the pair of $g \in \mathbb{R}^\times$ and $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ so that $v_{g^{-1}}^\psi$ makes itself an approximate identity in $L^1(\mathbb{R}^2)$, one may let the product of the convolution (*i.e.*, the ‘outcome’) converge towards the desired target

$$v_{g^{-1}}^{\psi_{B=b}^g} \rightarrow \rho_A^\phi(\cdot|B=b) \quad (5.82)$$

with respect to the L^1 -norm. A typical way to construct such an approximate identity is to start by preparing a compactly supported wave-function ψ , which automatically guarantees $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and to consider a family $\{\psi_{(h)}\}_{h>0}$ of the initial meter state defined as in (3.124). One then finds

$$\begin{aligned} \tilde{\omega}^{\psi_{(h)}}(x, y) &:= \psi_{(h)}^*(x - y/2)\psi_{(h)}(x + y/2) \\ &= |h|^{-1}\psi^*\left(\frac{x - y/2}{h}\right)\psi\left(\frac{x + y/2}{h}\right) \\ &= |h| \cdot \tilde{\omega}_h^\psi(x, y), \end{aligned} \quad (5.83)$$

and hence by observing that the above equality has total integration of $|h| \int_{\mathbb{R}^2} \tilde{\omega}^\psi dm_2$, its normalisation becomes

$$\omega^{\psi_{(h)}}(x, y) = \omega_h^\psi(x, y). \quad (5.84)$$

This should further lead to

$$v_{g^{-1}}^{\psi_{(h)}} = v_{hg^{-1}}^\psi, \quad (5.85)$$

when scaled by g^{-1} . With the initial ω^ψ (or equivalently v^ψ) being compactly supported, one then sees that this indeed makes an example of an approximate identity, and we may

thus achieve our objective by either narrowing the wave-function $h \rightarrow 0$, by intensifying the interaction $g^{-1} \rightarrow 0$ ($g \rightarrow \pm\infty$) or by appropriately balancing both manoeuvres and letting $hg^{-1} \rightarrow 0$ altogether.

5.3.2. Weak Conditioned Measurement. We shall next investigate how the map

$$g \mapsto v^{\psi_{B=b}^g}(x, y) \tag{5.86}$$

behaves locally around $g = 0$, and discuss what information of the target configuration one might reveal through it. In parallel to the case of the UM scheme discussed in Section 3.3.2, one finds below that the information of the configuration of the target system is encoded into the differential coefficients of the above map at $g = 0$, and that by knowing all the higher-order derivatives, one may fully recover the quasi-probability measure $\Delta \mapsto \mu_A^\phi(\Delta|B = b)$ of our interest.

Main Objective. Throughout the following passage, we assume the following.

- The quasi-probability measure $\Delta \mapsto \mu_A^\phi(\Delta|B = b)$ has a compact support.
- The total integration of $\tilde{\omega}^\psi$ is non-vanishing, and its normalisation belongs to the Schwartz space $\omega^\psi \in \mathcal{S}(\mathbb{R}^2)$.

These requirements are imposed primarily for the same reason as we have previously discussed in analysing the weak UM scheme in Section 3.3.2 (which, in short, is to say that we do not wish to get involved in the theory of generalised functions). A sufficient condition for the first and second assumptions would be to respectively require that the spectral measure E_A be compactly supported, and that $\psi \in \mathcal{S}(\mathbb{R})$. Under such conditions, the main objective of this passage is to demonstrate the following Proposition:

Proposition 5.2 (Weak Conditioned Measurement). *Under the above conditions, the map (5.86) is arbitrarily many times strongly differentiable on all the real line \mathbb{R} , and its n th derivatives at the origin $g = 0$ reads*

$$\left. \frac{d^n}{dg^n} v^{\psi_{B=b}^g} \right|_{g=0} = \sum_{|\gamma|=n} \mathbb{E} \left[a^\gamma; \mu_A^\phi(\cdot|B = b) \right] \cdot (-D)^\gamma v^\psi, \quad n \in \mathbb{N}_0. \tag{5.87}$$

Here, $\gamma = (\gamma_1, \gamma_2) \in \mathbb{N}_0^2$ is a multi-index introduced in (2.24), and the ‘quasi-moments’ under the quasi-probability measure $\mu_A^\phi(\cdot|B = b)$ is defined by

$$\mathbb{E} \left[a^\gamma; \mu_A^\phi(\cdot|B = b) \right] := \int_{\mathbb{C}} a_1^{\gamma_1} a_2^{\gamma_2} d\mu_A^\phi(a|B = b), \tag{5.88}$$

in its explicit form, where we understand $a = a_1 + ia_2 \in \mathbb{C}$.

Proof. Since the assumptions and reasonings are essentially the same as those provided for the unconditioned counterpart, we shall avoid reiteration and provide a rough sketch of the proof. In order to avoid clumsiness of notation, we write $v := v^\psi$, $v[g] := v^{\psi_{B=b}^g}$ and $\mu := \mu_A^\phi(\cdot|B = b)$ for simplicity, and denote by $v^{(n)}[g]$ the n th derivative of the map $g \mapsto v[g]$.

We first prove that the n th derivative of $v[g]$ reads

$$v^{(n)}[g](x) = \sum_{|\gamma|=n} \int_{\mathbb{R}^2} (-D)^\gamma v(x - ga) a^\gamma d\mu(a). \quad (5.89)$$

As above, we argue by mathematical induction. The case $n = 0$ is trivial. Suppose that the statement is true for $n \in \mathbb{N}_0$. Then, one may compute its point-wise derivative as

$$\begin{aligned} \frac{d}{dg} v^{(n)}[g](x) &= \sum_{|\gamma|=n} \int_{\mathbb{R}^2} \left(\frac{d}{dg} (-D)^\gamma v(x - ga) \right) a^\gamma d\mu(a) \\ &= \sum_{|\gamma|=n} \int_{\mathbb{R}^2} \left(\sum_{i=1}^2 a_i (-D)_i (-D)^\gamma v(x - ga) \right) a^\gamma d\mu(a) \\ &= \sum_{|\gamma|=n+1} \int_{\mathbb{R}^2} (-D)^\gamma v(x - ga) a^\gamma d\mu(a), \end{aligned} \quad (5.90)$$

and subsequently prove its strong differentiability by employing the same technique as above. This completes our first step of the proof.

Now, by taking $g = 0$ of (5.89), we observe

$$\begin{aligned} v^{(n)}[0](x) &= \sum_{|\gamma|=n} \int_{\mathbb{R}^2} (-D)^\gamma v(x - 0a) a^\gamma d\mu(a) \\ &= \sum_{|\gamma|=n} \int_{\mathbb{R}^2} a^\gamma d\mu(a) \cdot (-D)^\gamma v(x) \\ &= \sum_{|\gamma|=n} \mathbb{E}[a^\gamma; \nu] \cdot (-D)^\gamma v(x), \end{aligned} \quad (5.91)$$

which completes our proof. ■

One then immediately obtains the following corollary by applying the Stone-Weierstraß approximation theorem and the Riesz-Markov-Kakutani representation theorem.

Corollary 5.3 (Recovery of the Target Profile by Weak Conditioned Measurement). *The weak CM scheme (i.e., the knowledge of all the ‘quasi-moments’ (5.88)) allows us to uniquely specify the quasi-probability measure $\mu_A^\phi(\cdot | B = b)$ of our interest.*

Compare these results to those obtained in the case of the weak UM scheme described in Section 3.3.2.

5.4. Profile of the Target System

We have so far investigated how the CM scheme can be transcribed into the language of conditional quasi-probabilities, rather than in terms of mere conditional expectations. As a result, we found that the measurement outcome after the interaction incorporates two components: one being the profile of the meter system in the form of the WV distribution and the other being the that of the target system in the form of the quasi-probability measure $\mu_A^\phi(\cdot | B = b)$ defined in (5.70). Specifically, in view of the (scaled) inverse Fourier transform of the WV distribution, we found that the manner in which the two components interact with each other admits a simple description by convolution (5.75), which is quite analogous to the

unconditioned case. Based on our findings, we have thus analysed how one may recover the profile $\mu_A^\phi(\cdot|B=b)$ by means of both the strong and weak CM schemes, whose procedures are also quite analogous to the unconditioned counterpart. We are now interested in the properties of the quasi-probability measure $\mu_A^\phi(\cdot|B=b)$ we have obtained, which should be expected to convey some information of the target system.

Quasi-joint-probability Distribution of a Pair of Observables. By means of either the strong or weak CM scheme, we have so far obtained the family of quasi-probability measures $\mu_A^\phi(\cdot|B=b)$ for all $b \in \sigma(B)$. Allowing it to extend on the whole real line, one may construct a complex transition kernel by

$$\mu_A^\phi(\Delta_A|B=b) := \begin{cases} \mu_A^\phi(\Delta_A|B=b), & (b \in \sigma(B)) \\ \text{indefinite}, & (b \notin \sigma(B)) \end{cases}, \quad \Delta_A \in \mathfrak{B}(\mathbb{C}) \quad (5.92)$$

from the space $(\mathbb{R}, \mathfrak{B}^1)$ of the measurement outcomes of B into $(\mathbb{C}, \mathfrak{B}(\mathbb{C}))$. For definiteness, we assign to each $b \notin \sigma(B)$ any quasi-probability measure, so that (5.92) defines a *transitional quasi-probability kernel* as a whole. This allows us to construct a quasi-probability measure $\mu_{A,B}^\phi$ on the product space $(\mathbb{C} \times \mathbb{R}, \mathfrak{B}(\mathbb{C}) \otimes \mathfrak{B}^1)$, by combining the transition quasi-probability kernel (5.92) and the probability measure μ_B^ϕ , that satisfies

$$\mu_{A,B}^\phi(\Delta_A \times \Delta_B) = \int_{\Delta_B} \mu_A^\phi(\Delta_A|B=b) d\mu_B^\phi(b), \quad \Delta_A \in \mathfrak{B}(\mathbb{C}), \Delta_B \in \mathfrak{B}^1, \quad (5.93)$$

whose existence is guaranteed by Proposition 5.1. The target of our analysis in this passage is the quasi-probability measure (5.93). As one may expect from the notation employed, we shall shortly see that this qualifies as a QJP distribution of the target observable A and the conditioning observable B .

Proposition 5.4 (Quasi-joint-probability Distribution). *Under the definitions above, the quasi-probability measure $\mu_{A,B}^\phi$ qualifies as a QJP distribution of A and B in the sense of (5.48), namely*

$$\begin{aligned} \mu_{A,B}^\phi(\Delta_A \times \mathbb{R}) &= \mu_A^\phi(\Delta_A), \quad \Delta_A \in \mathfrak{B}(\mathbb{C}), \\ \mu_{A,B}^\phi(\mathbb{C} \times \Delta_B) &= \mu_B^\phi(\Delta_B), \quad \Delta_B \in \mathfrak{B}(\mathbb{R}) \end{aligned} \quad (5.94)$$

holds. Here, μ_A^ϕ denotes the probability measure on $(\mathbb{C}, \mathfrak{B}(\mathbb{C}))$ generated by the two-dimensional spectral measure associated to A understood as a normal operator, whereas μ_B^ϕ denotes the probability measure on $(\mathbb{R}, \mathfrak{B})$ generated by the one-dimensional spectral measure associated to the self-adjoint operator B .

Proof. We start by demonstrating that the marginal of the quasi-probability measure $\mu_{A,B}^\phi$ of the first term coincides with the probability measure μ_A^ϕ on $(\mathbb{C}, \mathfrak{B}(\mathbb{C}))$ generated by the spectral measure of A (seen as a normal operator). To this end, we first observe

$$\begin{aligned} \mu_{A,B}^\phi(\Delta_A \times \mathbb{R}) &= \int_{\mathbb{R}} \mu_A^\phi(\Delta_A|B=b) d\mu_B^\phi(b) \\ &= \sum_{b \in \sigma(B)} (\nu_b^* \otimes \nu_b)(T^{-1}\Delta_A) \cdot |\langle b, \phi \rangle|^2, \end{aligned} \quad (5.95)$$

where we have used (5.70) in the last equality.

Now, in order to proceed further, we then maintain that the measure

$$\Delta \mapsto \mu(\Delta) := \sum_{b \in \sigma(B)} (\nu_b^* \otimes \nu_b)(\Delta) \cdot |\langle b, \phi \rangle|^2, \quad \Delta \in \mathfrak{B}(\mathbb{C}) \quad (5.96)$$

is essentially the same object as the continuous \mathbb{C} -linear map defined by

$$I_{\text{diag}}(f) := \int_{\mathbb{R}} f(a, a) d\mu_A^\phi(a), \quad f \in C_0(\mathbb{C}) \quad (5.97)$$

in the sense of the Riesz-Markov-Kakutani representation theorem. The proof can be carried out in several ways, but for the sake of simplicity, we rather take an elementary approach. Observing that any two measures on a product space $(X \times Y, \mathfrak{A} \otimes \mathfrak{B})$ coincides with each other if they coincide on the subset $\mathfrak{A} * \mathfrak{B} \subset \mathfrak{A} \otimes \mathfrak{B}$ (see (3.29) for the definition), one proceeds as

$$\begin{aligned} \int_{\mathbb{C}} \chi_{\Delta_1}(a_1) \chi_{\Delta_2}(a_2) d\mu(a) &= \mu(\Delta_1 \times \Delta_2) \\ &= \sum_{b \in \sigma(B)} (\nu_b^* \otimes \nu_b)(\Delta_1 \times \Delta_2) \cdot |\langle b, \phi \rangle|^2 \\ &= \sum_{b \in \sigma(B)} \langle \phi, E_A(\Delta_1)b \rangle \langle b, E_A(\Delta_2)\phi \rangle \\ &= \langle \phi, E_A(\Delta_1)E_A(\Delta_2)\phi \rangle \\ &= \int_{\mathbb{R}} \chi_{\Delta_1}(a) \chi_{\Delta_2}(a) d\mu_A^\phi(a) \\ &= I_{\text{diag}}(\chi_{\Delta_1} \chi_{\Delta_2}), \end{aligned} \quad (5.98)$$

which proves

$$\mu \cong I_{\text{diag}}. \quad (5.99)$$

Armed with the findings, we return to our original problem (5.95) and finally obtain

$$\begin{aligned} \sum_{b \in \sigma(B)} (\nu_b^* \otimes \nu_b)(T^{-1}\Delta_A) \cdot |\langle b, \phi \rangle|^2 &= I_{\text{diag}}(\chi_{(T^{-1}\Delta_A)}) \\ &= \int_{\mathbb{R}} \chi_{(T^{-1}\Delta_A)}(a, a) d\mu_A^\phi(a) \\ &= \int_{\mathbb{R}} \chi_{\Delta_A}(a, 0) d\mu_A^\phi(a) \\ &= \langle \phi, E_{A,0}(\Delta_A)\phi \rangle \\ &= \mu_A^\phi(\Delta_A), \quad \Delta_A \in \mathfrak{B}(\mathbb{C}), \end{aligned} \quad (5.100)$$

where $E_{A,0}$ denotes the product spectral measure of the one-dimensional spectral measure E_A of A (as a self-adjoint operator) and that of the 0 operator E_0 (*i.e.*, the ‘delta spectral measure’ (3.102) centred at the origin), and the last equality is due to the observation that the two-dimensional spectral measure \tilde{E}_A of A as a normal operator coincides with the product spectral measure $\tilde{E}_A = E_{A,0}$ introduced above. This completes our proof for the marginal of the first term.

It now remains to compute the marginal of $\mu_{A,B}^\phi$ of the second term, which one carries out as

$$\begin{aligned}
\mu_{A,B}^\phi(\mathbb{C} \times \Delta_B) &= \int_{\mathbb{C} \times \mathbb{R}} \chi_{\Delta_B}(b) d\mu_{A,B}^\phi(a, b) \\
&= \int_{\mathbb{C} \times \mathbb{R}} \chi_{\Delta_B}(b) \mu_A^\phi(\mathbb{C} | B = b) d\mu_B^\phi(b) \\
&= \int_{\mathbb{C} \times \mathbb{R}} \chi_{\Delta_B}(b) d\mu_B^\phi(b) \\
&= \mu_B^\phi(\Delta_B), \quad \Delta_B \in \mathfrak{B},
\end{aligned} \tag{5.101}$$

where the second equality is due to the definition of $\mu_{A,B}^\phi$, and the third equality is due to the fact that $\mu_A^\phi(\mathbb{C} | B = b) = 1$ is normalised to unity (*i.e.*, a quasi-probability measure) for all $b \in \sigma(B)$. \blacksquare

As for the relation between the QJP distribution $\mu_{A,B}^\phi$ and the transition quasi-probability kernel $(b, \Delta_A) \mapsto \mu_A(\Delta_A | B = b)$, one immediately has the following corollary by construction.

Corollary 5.5. *The transition quasi-probability kernel (5.92) is a conditional quasi-probability distribution of A given B under the QJP distribution $\mu_{A,B}^\phi$.*

Conditional Quasi-expectation of A given B. It is now tempting to investigate how the ‘conditional average’ of the QJP distribution $\mu_{A,B}^\phi$ relates to the conditional quasi-expectation $\mathbb{E}^\alpha[A|B; \phi]$ we have introduced earlier in (4.77).

Proposition 5.6 (Conditional Average of the Quasi-joint-probability Distribution). *Under the definitions above, the conditional average of A given B under the QJP distribution $\mu_{A,B}^\phi$ reads*

$$\int_{\mathbb{C}} a d\mu_A^\phi(a|B = b) = \mathbb{E}^i[A|B = b; \phi], \tag{5.102}$$

where the *r. h. s.* is the member of the complex-parametrised sub-family of conditional quasi-expectations of A given B introduced in (4.77) for the purely imaginary choice $\alpha = i$ of the parameter.

Proof. For the demonstration, let $b \in \sigma(B)$. One then has

$$\begin{aligned}
\int_{\mathbb{C}} a d\mu_A^\phi(a|B = b) &= \int_{\mathbb{R}^2} (a_1 + ia_2) dT(\nu_b^* \otimes \nu_b)(a_1, a_2) \\
&= \int_{\mathbb{C}} \frac{a + a'}{2} + i \frac{a - a'}{2} d(\nu_b^* \otimes \nu_b)(a', a) \\
&= \int_{\mathbb{C}} \frac{1+i}{2} a + \frac{1-i}{2} a' d(\nu_b^* \otimes \nu_b)(a, a') \\
&= \frac{1+i}{2} \cdot \frac{\langle b, A\phi \rangle}{\langle b, \phi \rangle} + \frac{1-i}{2} \cdot \left(\frac{\langle b, A\phi \rangle}{\langle b, \phi \rangle} \right)^* \\
&= \mathbb{E}^i[A|B = b; \phi],
\end{aligned} \tag{5.103}$$

where the second equality is due to the change of variables formula (3.10) for image measures. \blacksquare

Obtaining Conditional Quasi-probability Distribution by Conditioned Measurement. We now realise that the CM scheme, in view of conditional quasi-probabilities, can be regarded as a *method of obtaining conditional quasi-probability distributions of the target observable A given the conditioning observable B* , and that it implies the existence of QJP distributions of a pair of (generally not necessarily simultaneously measurable) quantum observables lying underneath. Moreover, we have seen a connection between the concept of conditional quasi-expectations and the ‘conditional average’ of the QJP distributions, which is reminiscent of the familiar relation between classical conditional expectations and conditional average of probability measures. While we have conducted an analysis for the special case in which both A and B happen to possess spectra of finite cardinalities (and that B is degenerate), we note that one may suitably generalise the results obtained here by introducing appropriate mathematical tools and some little more advanced mathematical languages.

6. Quasi-probabilities of Quantum Observables

By studying the both the UM and CM schemes in depth throughout the preceding four sections, we have so far naturally arrived, by a purely bottom-up construction, at the concept of *quasi-joint-probability* (QJP) of an arbitrary pair of quantum observables. While such an operational way of demonstration has its own merit of being solid and down to earth, it has an apparent downside in that the line of argument lacks transparency and that the whole structure may become obscure on occasions. In this section, we will be conducting a top-down study on the topic as a complement to the analyses made in the preceding sections.

Organisation of this Section. In this section, we first devote several pages to introducing some mathematical tools for our analysis as usual. We then propose a general prescription for the construction of QJP distributions of a given pair of quantum observables, and observe their basic properties. Since it is difficult to perform a general analysis on the whole class of all possible candidates of QJP distributions with full mathematical rigour due to the limited framework and tools available, for our demonstration we shall mostly concentrate on a special sub-family of such distributions parametrised by a single complex number, hopefully without loss of too much essence. We finally close this section by observing where the bottom-up line of discussion performed in Section 5 fits in this more general framework.

6.1. Reference Materials

As usual, we first prepare some necessary mathematical tools for reference. As a generalisation to those defined on integrable functions, we now introduce Fourier transforms of complex measures.

6.1.1. Fourier Transform of Complex Measures. Analogous to the manner in which we have defined Fourier transforms of elements of $L^1(\mathbb{R}^n)$ (namely, the density functions), one may define Fourier transforms of complex measures. Given a complex measure $\mu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{A})$ on a measurable space (X, \mathfrak{A}) , we define the Fourier transform and the inverse Fourier transform of μ , respectively by the functions

$$\hat{\mu}(q) := \int_X e^{-i\langle q, x \rangle} d\mu(x), \quad (6.1)$$

$$\check{\mu}(q) := \int_X e^{i\langle q, x \rangle} d\mu(x), \quad (6.2)$$

where $\langle q, x \rangle := \sum_{k=1}^n q_k x_k$ denotes the scalar product on \mathbb{R}^n as usual. Note that the functions $\hat{\mu}$, $\check{\mu}$ are well-defined, for indeed $|\hat{\mu}(q)| \leq \int_X |e^{-i\langle q, x \rangle}| d|\mu|(x) = \|\mu\| < \infty$ for all $q \in \mathbb{R}^n$, where $|\mu|$ and $\|\mu\|$ are respectively the variation and the total variation of μ (a similar evaluation holds for $\check{\mu}$).

Basic Properties. To see how this newly introduced definition of Fourier transforms relates to that of integrable functions introduced earlier, let $L^1(\mathfrak{B}^n) \subset \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ be the sub-algebra of absolutely continuous complex measures with respect to m_n , where m_n denotes the renormalised n -dimensional Lebesgue-Borel measure on $(\mathbb{R}^n, \mathfrak{B}^n)$ defined in (5.26). Choosing $\mu \in L^1(\mathfrak{B}^n)$ and letting $\rho := d\mu/dm_n$ be the Radon-Nikodým derivative of μ , one finds by a

direct application of (3.20) that

$$\begin{aligned}\hat{\mu}(q) &:= \int_{\mathbb{R}^n} e^{-i\langle q, x \rangle} d\mu(x) \\ &= \int_{\mathbb{R}^n} e^{-i\langle q, x \rangle} \rho(x) dm_n(x) =: \hat{\rho}(q),\end{aligned}\tag{6.3}$$

holds as expected. An analogous relation holds for the inverse Fourier transform as well. The \mathbb{C} -linear map \mathcal{F} that maps a complex measure into its Fourier transform is called the *Fourier transformation*. In parallel to that defined for integrable functions, the Fourier transformation on the measure algebra is injective, *i.e.*, $\hat{\mu} = \hat{\nu}$ implies $\mu = \nu$. For $\mu, \nu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$, the properties

$$\widehat{(\mu * \nu)} = \hat{\mu} \cdot \hat{\nu},\tag{6.4}$$

$$\widehat{(\mu_t)}(q) = \hat{\mu}(tq), \quad t \neq 0,\tag{6.5}$$

$$\widehat{(\tau_a \mu)}(t) = e^{i\langle a, x \rangle} \hat{\mu}(t), \quad a \in \mathbb{R},\tag{6.6}$$

are basic, in which one sees how the Fourier transformation behaves under the convolution (3.11), scaling (3.94), and translation

$$(\tau_a \mu)(B) := \mu(B + a), \quad a \in \mathbb{R}^n,\tag{6.7}$$

respectively.

Linear Transformation. Let T be a linear operator on \mathbb{R}^n (*i.e.*, an $n \times n$ real matrix), and let $\mu \in \mathbf{M}_{\mathbb{C}}(\mathfrak{B}^n)$ be a complex measure. We then define the *linear transform*

$$\mu_T(B) := \mu(T^{-1}B), \quad B \in \mathfrak{B}^n,\tag{6.8}$$

of μ with respect to T by its image measure. By definition, one readily finds the validity of the product rule

$$(\mu_T)_S = \mu_{(ST)}\tag{6.9}$$

for a pair of linear operators S and T on \mathbb{R}^n , and that

$$\int_X f(x) d\mu_T(x) = \int_X f(Tx) d\mu(x)\tag{6.10}$$

by the change of variables formula (3.10), whenever the integration exists. Note that the familiar scaling μ_t , $t \neq \mathbb{R}$ defined in (3.94), and the translation $\tau_a \mu$, $a \in \mathbb{R}^n$ defined in (6.7) are respectively special cases of the linear transform of μ with respect to $T = tI$ and $T = I - a$, where I denotes the identity operator. In such a cases, note also that the linear operators involved are automorphisms, hence members of the general linear group $\text{GL}(n; \mathbb{R})$. In relation to the Fourier transformation, one finds

$$\begin{aligned}(\mathcal{F} \mu_T)(q) &:= \int_{\mathbb{R}^n} e^{-i\langle q, x \rangle} d\mu_T(x) \\ &= \int_{\mathbb{R}^n} e^{-i\langle q, Tx \rangle} d\mu(x) \\ &= \int_{\mathbb{R}^n} e^{-i\langle T^* q, x \rangle} d\mu(x) \\ &= (\mathcal{F} \mu)(T^* q),\end{aligned}\tag{6.11}$$

where T^* denotes the adjoint (in this case, the transpose $T^* = T^t$) of the Matrix T .

Complex Conjugate. We finally review how the Fourier transform behaves regarding the operation of taking the complex conjugate of a complex measure. To this end, let μ be a complex measure on $(\mathbb{R}^n, \mathfrak{B}^n)$, and define the complex conjugate of μ by $\mu^*(\Delta) := \mu(\Delta)^*$, $\Delta \in \mathfrak{B}^n$ in a natural manner. One then readily finds

$$\begin{aligned} (\mathcal{F}\mu^*)(q) &:= \int_{\mathbb{R}^n} e^{-i\langle q, x \rangle} d\mu^*(x) \\ &= \left(\int_{\mathbb{R}^n} e^{-i\langle -q, x \rangle} d\mu(x) \right)^* \\ &= (\mathcal{F}\mu)^*(-q) \\ &= (\mathcal{F}\mu)^\dagger(q), \end{aligned} \tag{6.12}$$

where $f^\dagger(x) := f^*(-x)$ denotes the *involution* of a function f .

Differentiation. We finally make a brief note on the basic results regarding differentiability and derivatives of a Fourier transform of a complex measure at the origin.

Lemma 6.1. *Let μ be a complex measure on $(\mathbb{R}^n, \mathfrak{B}^n)$, and let $\gamma \in \mathbb{N}_0^n$ be a multi-index. If the integration*

$$\int_{\mathbb{R}^n} x^{\gamma'} d\mu(x), \tag{6.13}$$

exists for all $0 \leq \gamma' \leq \gamma$, then the derivative $D^\gamma \hat{\mu}$ of the Fourier transform of μ exists at the origin, in which case the derivative reads

$$(D^\gamma \hat{\mu})(0) = (-i)^{|\gamma|} \int_{\mathbb{R}^n} x^\gamma d\mu(x). \tag{6.14}$$

Proof. One readily computes

$$\begin{aligned} (D^\gamma \hat{\mu})(0) &:= D^\gamma \left(\int_{\mathbb{R}^n} e^{-i\langle q, x \rangle} d\mu(x) \right) \Big|_{q=0} \\ &= \int_{\mathbb{R}^n} \left(D^\gamma e^{-i\langle q, x \rangle} \Big|_{q=0} \right) d\mu(x) \\ &= (-i)^{|\gamma|} \int_{\mathbb{R}^n} x^\gamma d\mu(x), \end{aligned} \tag{6.15}$$

where the second equality (exchange of the differentiation and integration) is a consequence of the dominated convergence theorem. ■

Compare this result to that for Schwartz functions (5.34).

6.2. Quasi-joint-probabilities of a Combination of Quantum Observables

We now intend to provide a general prescription for defining a QJP distribution of a combination of generally not necessarily simultaneously measurable quantum observables.

6.2.1. Preliminary Observations. In this passage, we conduct some formal discussions on the topic of QJP distributions of a combination of quantum observables. Since rigorous treatment requires advanced mathematical tools that is beyond the scope of this paper, we first conduct a formal and intuitive argument to obtain the essence of the idea. Now,

before we embark on our main objective, we first recall a basic theorem regarding strong commutativity of A and B and that of their unitary operators.

Theorem. *Let A and B be self-adjoint. Then, the following conditions are equivalent.*

- (i) *The operators A and B strongly commute with each other.*
- (ii) *The operators e^{isA} and e^{itB} commute with each other for all $s, t \in \mathbb{R}$, namely*

$$e^{itA} e^{isB} = e^{isB} e^{itA}, \quad s, t \in \mathbb{R} \quad (6.16)$$

holds.

This familiar theorem builds the starting point of our discussion that follows.

Fourier Transform of Product Spectral Measures. Recall that the joint behaviour of the outcomes of an ideal measurement of a pair of simultaneously measurable observables A and B is governed by the product spectral measure $E_{A,B}$ of their respective spectral measures E_A, E_B introduced earlier in (3.67). An important observation here is to see that the ‘Fourier transform’ of the product spectral measure $E_{A,B}$ is nothing but the product (6.16) of the parametrised unitary operators

$$\begin{aligned} (\mathcal{F} E_{A,B})(s, t) &:= \int_{\mathbb{R}^2} e^{-i((s,a)+(t,b))} dE_{A,B}(a, b) \\ &= e^{-i\overline{(sA+tB)}} \\ &= \lim_{N \rightarrow \infty} (e^{-isA/N} e^{-itB/N})^N \\ &= e^{-isA} e^{-itB}, \end{aligned} \quad (6.17)$$

where the overline on the essentially self-adjoint operator $sA + tB$ denotes its unique self-adjoint extension as usual, and the second equality is due to the familiar Trotter formula.

Hashed Operators. We now consider a pair of *arbitrary* (not necessarily strongly-commuting) self-adjoint operators A and B . Guided by the above observation, we formally introduce

$$\#(s, t) := \text{a ‘decent’ mixture of the disintegrated components of } e^{-isA} \text{ and } e^{-itB} \quad (6.18)$$

for the pair of A and B . Example of such mixtures of the disintegrated components of the unitary operators are given by:

$$\#(s, t) = \begin{cases} e^{-isA} e^{-itB}, \\ e^{-itB} e^{-isA}, \\ \prod_{k=1}^N e^{-isA/L_k} e^{-itB/M_k}, & \left(\sum_{k=1}^N L_k^{-1} = 1, \sum_{k=1}^N M_k^{-1} = 1 \right), \\ (e^{-isA/N} e^{-itB/N})^N, \\ e^{-i\overline{(sA+tB)}} = \lim_{N \rightarrow \infty} (e^{-isA/N} e^{-itB/N})^N, \\ \text{etc.}, \end{cases} \quad (6.19)$$

or even any linear combinations of them. The term ‘decent’ is intended to express a mathematical condition as to what qualifies as a reasonable ‘mixture’ to meet our purpose. However, we do not intend to discuss its precise mathematical definition here, for it is beyond

the scope of this paper. In this paper, the ‘parametrised family of operators’ $\#(s, t)$ shall occasionally be referred to as *hashed operators* of the unitary operators, in a rather casual manner. Due to the commutativity of the unitary operators for a simultaneously measurable pair, the hashed operator $\#(s, t) = e^{-isA}e^{-itB}$ is always unique, while on the other hand, hashed operators admit variety for non-commutative pairs. Now, given a hashed operator $\#$ of A and B , we then introduce the collection of all parametrised operators of the form

$$\hat{\mathfrak{M}}_{A,B} := \{ \# : \# \text{ is a hashed operator of } A \text{ and } B \text{ defined as in (6.18)} \}. \quad (6.20)$$

As we have seen above, in the case in which A and B are simultaneously measurable, the above collection in fact consists of only one trivial element

$$\hat{\mathfrak{M}}_{A,B} = \{ e^{-isA}e^{-itB} \}, \quad (6.21)$$

due to the strong commutativity of the two operators. On the other hand, one readily observes that the cardinality of $\hat{\mathfrak{M}}_{A,B}$ is always greater than unity if the pair of observables A and B fails to strongly commute.

Lemma 6.2. *The cardinality of the collection $\hat{\mathfrak{M}}_{A,B}$ is equal to unity if and only if the observables A and B strongly commute with each other. Otherwise, the cardinality is always greater than unity.*

Distributions generated by Hashed Operators. We now consider the inverse Fourier transform of all the elements of the hashed operators $\hat{\mathfrak{M}}_{A,B}$, and thus formally introduce

$$\mathfrak{M}_{A,B} := \{ \mathcal{F}^{-1}\# : \# \in \hat{\mathfrak{M}}_{A,B} \}, \quad (6.22)$$

without any consideration of the mathematical intricacies involved in its well-definedness. By the injectivity of the Fourier transformation, one intuitively expects that the collection reduces to the single element

$$\mathfrak{M}_{A,B} = \{ E_{A,B} \} \quad (6.23)$$

when the operators A and B strongly commute with each other, which should be nothing but the original product spectral measure of A and B . On the other hand, Lemma 6.2 implies that the cardinality of the collection $\mathfrak{M}_{A,B}$ is always greater than unity in the case where A and B are not simultaneously measurable. Although being possibly highly non-unique, each element of the collection $\mathfrak{M}_{A,B}$ defined for non-commuting pairs retains similar properties to those of the standard product spectral measures. Incidentally, choosing any element $\Pi \in \mathfrak{M}_{A,B}$ of the collection, the ‘total integration’ reduces to the unit I , as one finds under the formal computation

$$\begin{aligned} \int_{\mathbb{K}^2} \Pi(a, b) dm_2(a, b) &= \int_{\mathbb{K}^2} e^{-i((0,a)+(0,b))} \Pi(a, b) dm_2(a, b) \\ &= (\mathcal{F}\Pi)(0, 0) \\ &= \#(0, 0) = I, \end{aligned} \quad (6.24)$$

where $\#$ is the hashed operator whose inverse Fourier transform is the element $\Pi = \mathcal{F}^{-1}\#$ of our choice. As for the marginals, by formally introducing

$$\Pi_B(b) := \int_{\mathbb{K}} \Pi(a, b) dm(a), \quad (6.25)$$

one observes under a formal computation that

$$\begin{aligned}
(\mathcal{F}\Pi_B)(t) &= \int_{\mathbb{K}} e^{-i\langle t, b \rangle} \left(\int_{\mathbb{K}} \Pi(a, b) dm(a) \right) dm(b) \\
&= \int_{\mathbb{K}^2} e^{-i(\langle 0, a \rangle + \langle t, b \rangle)} \Pi(a, b) dm_2(a, b) \\
&= \#(0, t) \\
&= e^{-itB} \\
&= (\mathcal{F}E_B)(t).
\end{aligned} \tag{6.26}$$

The injectivity of the Fourier transformation \mathcal{F} leads us to conclude that the marginal $\Pi_B = E_B$ is essentially the same object as the original spectral measure governing the probabilistic behaviour of the outcomes of B . By a parallel argument, one also finds that the marginal

$$\Pi_A(a) := \int_{\mathbb{K}} \Pi(a, b) dm(b) \tag{6.27}$$

is nothing but $\Pi_A = E_A$. These properties are naturally found common in product spectral measures defined for strongly commuting pairs of self-adjoint operators, although each $\Pi(a, b)$ is not necessarily a projection, or may not be even positive. This tempts us to introduce the term *quasi-joint-spectral distributions* of a pair of observables, which can be understood as a generalisation of the concept of spectral measures or POVMs.

Definition (Quasi-joint-spectral Distribution of a Pair of Quantum Observables). *Let A and B be self-adjoint operators on \mathcal{H} . We call an element of $\mathfrak{M}_{A,B}$ a quasi-joint-spectral distribution of the pair of observables A and B . The cardinality of the collection $\mathfrak{M}_{A,B}$ is equal to unity if and only if A and B strongly commute with each other. Otherwise, it is always greater than unity.*

In the case where the observables A and B happen to strongly commute with each other, we specifically call the unique element of $\mathfrak{M}_{A,B}$ the *joint-spectral distribution* of A and B , which is nothing but the product spectral measure $E_{A,B}$ of the pair in standard terminology. We note that the terminologies introduced above are non-standard, and are to be used only in this paper.

6.2.2. Quasi-joint-probability Distributions. Although the study on the precise definitions and properties of the family of quasi-joint-spectral distributions would be of mathematical interest in its own right, we shall refrain from going further due to the limited mathematical tools available. Instead, we turn to a more elementary object to ease our discussion.

Now, given a quasi-joint-spectral distribution $\Pi \in \mathfrak{M}_{A,B}$ of A and B , we fix a specific quantum state $|\phi\rangle \in \mathcal{H}$, and consider a distribution of the form formally defined by

$$\begin{aligned}
p(a, b) &:= \frac{\langle \phi, \Pi(a, b)\phi \rangle}{\|\phi\|^2} \\
&= \frac{\langle \phi, (\mathcal{F}^{-1}\#)(a, b)\phi \rangle}{\|\phi\|^2} \\
&= \left(\mathcal{F}^{-1} \frac{\langle \phi, \#(\cdot, \cdot)\phi \rangle}{\|\phi\|^2} \right) (a, b),
\end{aligned} \tag{6.28}$$

where $\#$ is the hashed operator of which inverse Fourier transform $\Pi = \mathcal{F}^{-1}\#$ is the quasi-joint-spectral distribution under consideration. Since the distribution p is ‘scalar valued’, it

should be a much more feasible object to deal with than the ‘operator valued’ distribution Π introduced earlier. We thus introduce the collection

$$\hat{\mathfrak{M}}_{A,B}^\phi := \left\{ \frac{\langle \phi, \#(s,t)\phi \rangle}{\|\phi\|^2} : \# \in \hat{\mathfrak{M}}_{A,B}, |\phi\rangle \in \mathcal{H} \right\} \quad (6.29)$$

of all distributions generated by the hashed operators of the parametrised unitary operators give a fixed state, and in turn formally define

$$\mathfrak{M}_{A,B}^\phi := \left\{ \mathcal{F}^{-1}u : u \in \hat{\mathfrak{M}}_{A,B}^\phi \right\}, \quad (6.30)$$

by their inverse Fourier transforms. We thus summarise as:

Definition (Quasi-joint-probability Distribution of a Pair of Quantum Observables). *Let A and B be self-adjoint operators on \mathcal{H} , and let $|\phi\rangle \in \mathcal{H}$. We call an element of $\mathfrak{M}_{A,B}^\phi$ a quasi-joint-probability (QJP) distribution of the pair of observables A and B on $|\phi\rangle$. The cardinality of the collection $\mathfrak{M}_{A,B}^\phi$ is equal to unity for every choice of the vector $|\phi\rangle$ if and only if the observables A and B strongly commute with each other. Otherwise, there exists a vector $|\phi\rangle$ such that the cardinality is greater than unity.*

In the case where the observables A and B happen to strongly commute with each other, we specifically call the unique element of $\mathfrak{M}_{A,B}^\phi$ the joint-probability distribution of A and B on $|\phi\rangle$, which is nothing but the probability measure $\mu_{A,B}^\phi$ of the pair introduced in (3.69). Given a hashed operator $\#$ of the parametrised unitary groups and a quantum state $|\phi\rangle \in \mathcal{H}$, we call an element $p \in \mathfrak{M}_{A,B}^\phi$ specified by

$$(\mathcal{F}p)(s,t) = \frac{\langle \phi, \#(s,t)\phi \rangle}{\|\phi\|^2}, \quad (6.31)$$

the QJP distribution generated by $\#$ and $|\phi\rangle$. Our choice of the denomination of the elements of $p \in \mathfrak{M}_{A,B}^\phi$ is due to the fact that they retain similar properties to those of classical joint-probability distributions. Indeed, the ‘total integration’ reduces to

$$\begin{aligned} \int_{\mathbb{K}^2} p(a,b) dm_2(a,b) &= \int_{\mathbb{K}^2} e^{-i(\langle 0,a \rangle + \langle 0,b \rangle)} p(a,b) dm_2(a,b) \\ &= (\mathcal{F}p)(0,0) \\ &= \frac{\langle \phi, \#(0,0)\phi \rangle}{\|\phi\|^2} = 1, \end{aligned} \quad (6.32)$$

where $\#$ is the hashed operator that, together with $|\phi\rangle$, generates p . As for the marginals, by introducing the marginal distribution formally defined by

$$p_B(b) := \int_{\mathbb{K}} p(a,b) dm(a), \quad (6.33)$$

one observes through a formal computation that

$$\begin{aligned}
(\mathcal{F}p_B)(t) &= \int_{\mathbb{K}} e^{-i\langle t, b \rangle} \left(\int_{\mathbb{K}} p(a, b) dm(a) \right) dm(b) \\
&= \int_{\mathbb{K}^2} e^{-i(\langle 0, a \rangle + \langle t, b \rangle)} p(a, b) dm_2(a, b) \\
&= \frac{\langle \phi, \#(0, t)\phi \rangle}{\|\phi\|^2} \\
&= \frac{\langle \phi, e^{-itB}\phi \rangle}{\|\phi\|^2} \\
&= \left(\mathcal{F}\mu_B^\phi \right)(t).
\end{aligned} \tag{6.34}$$

The injectivity of the Fourier transformation \mathcal{F} leads us to conclude that the distribution $p_B(b)$ is essentially the same object as the probability measure μ_B^ϕ describing the probabilistic behaviour of the outcomes of B . By a parallel argument, one also finds that the marginal

$$p_A(b) := \int_{\mathbb{K}} p(a, b) dm(b) \tag{6.35}$$

is nothing but $p_A = \mu_A^\phi$. Before we proceed further, we make notes on some mathematical intricacies involved in their definitions for the interested.

Mathematical Remarks. One may notice some subtleties inherent to the definition of $\mathfrak{M}_{A,B}^\phi$. The first problem might be the domain of the definition of the inverse Fourier transformation: while the Fourier transform of a complex measure μ is a function, in regard that it does not necessarily lie in $\hat{\mu} \notin L^1(\mathbb{R}^n)$, its inverse Fourier transform may not be well-defined, even in the case where A and B strongly commute with each other. This can be temporarily remedied by understanding the inverse Fourier transform of an element $u \in \hat{\mathfrak{M}}_{A,B}^\phi$ to be the unique complex measure μ such that $u = \hat{\mu}$ holds, which should be a reasonable treatment due to the injectivity of the Fourier transformation. This provides a sufficient cure in the case where the pair of observables strongly commutes.

On the other hand, another problem arises in the case in which the pair of self-adjoint operators fails to strongly commute: it might happen that, for some element $u \in \hat{\mathfrak{M}}_{A,B}^\phi$, there is no complex measure μ such that its Fourier transform coincides with $u = \hat{\mu}$. A straightforward and more fundamental cure for this would be to expand our framework into that of generalised functions, specifically, by embedding the space of complex measures into that of tempered distributions. Indeed, since the Fourier transformation is a bijection on the space of tempered distributions, by understanding that each of the elements of $\hat{\mathfrak{M}}_{A,B}^\phi$ to be a tempered distribution, its inverse Fourier transform itself always exists as a tempered distribution. In consideration of this, since we do not wish to get involved with the theory of generalised functions, we shall be exclusively dealing with those elements $u \in \hat{\mathfrak{M}}_{A,B}^\phi$ for which there exists a complex measure μ satisfying $u = \hat{\mu}$, and understand the element $\mu := \mathcal{F}^{-1}u \in \mathfrak{M}_{A,B}^\phi$ to be the complex measure. To this end, we introduce:

Definition (Representation by Quasi-probability Measures). *Under the above situation, let $p \in \mathfrak{M}_{A,B}^\phi$ be a QJP distribution of A and B , and let $u \in \hat{\mathfrak{M}}_{A,B}^\phi$ be an element such that $p = \mathcal{F}^{-1}u$. We say that the QJP distribution p admits representation by a quasi-probability*

measure, if there exists a quasi-probability measure μ on \mathbb{K}^2 such that $u = \hat{\mu}$ holds, and understand the QJP distribution $p = \mu$ to be the quasi-probability measure.

A similar concern arises for the definition of quasi-joint-spectral distributions $\Pi = \mathcal{F}^{-1}\#$ defined as inverse Fourier transforms of hashed operators $\#$ of the unitary operators e^{-isA} and e^{-itB} . Parallel to the ‘scalar valued’ case seen above, quasi-joint-spectral distributions Π are better understood as an object generalising the concept of spectral measures (or POVMs), in the sense that, while spectral measures E (or POVMs) yield probability measures $\langle \phi, E(\cdot)\phi \rangle / \|\phi\|^2$ when combined with a vector $|\phi\rangle$, quasi-joint-spectral distributions Π yield generalised functions, symbolically denoted by $\langle \phi, \Pi(a, b)\phi \rangle / \|\phi\|^2$. In this respect, quasi-joint-spectral distributions are to be understood as elements of the space of *operator valued (tempered) distributions* (OVDs), which should serve as a generalisation to that of POVMs.

We also note that the methods introduced above in defining QJSDs admit a straightforward generalisation in defining them, not only for a *pair* ($N = 2$) of quantum observables as presented above, but also for *arbitrary combinations* ($N \geq 2$) of quantum observables, or even for *arbitrary combinations of POVMs*. Also, one may readily generalise the discussion for defining QJP distributions, not just for pure states as presented above by sandwiching the QJSPs by kets and bras, but also for mixed states by taking the trace of the product of QJSPs and density operators.

6.3. Complex-parametrised Sub-families

Since we have decided to confine ourselves in the framework of complex measures rather than that of generalised functions due to our restricted mathematical tools available, we would mostly refrain from treating the general cases, and shall concentrate on a special sub-families of QJP distributions of a pair of quantum observables A and B .

6.3.1. Additive Sub-family. As a simple example of QJP distributions admitting representation by quasi-probability measures, we observe:

Lemma 6.3. *Let A and B be self-adjoint operators on \mathcal{H} , and consider the hashed operator of either of the form*

$$\#(s, t) = \begin{cases} e^{-itA}e^{-isB} \\ e^{-isB}e^{-itA} \end{cases}, \quad s, t \in \mathbb{R}. \quad (6.36)$$

Then, the QJP distributions generated by $\#$ and any choice of the vector $|\phi\rangle \in \mathcal{H}$ admit representation by quasi-probability measures.

Proof. We provide the proof for the first case without loss of generality. Observe that the complex measure

$$\Delta \mapsto \nu(\Delta, \Delta_A) := \frac{\langle \phi, E_B(\Delta)E_A(\Delta_A)\phi \rangle}{\|\phi\|^2}, \quad \Delta \in \mathfrak{B} \quad (6.37)$$

is absolutely continuous with respect to μ_B^ϕ for all fixed $\Delta_A \in \mathfrak{B}$. This allows us to construct a transition quasi-probability kernel by taking the Radon-Nikodým derivative of the above complex measure with respect to μ_B^ϕ . A direct application of Proposition 5.1 with $f(a, b) := e^{-i(as+bt)}$ then leads to the desired statement. \blacksquare

This inspires us to introduce the complex linear combinations of the above two distributions. We hereby consider the hashed operators of the form

$$\#_{\text{add}}^\alpha(s, t) := \frac{1 + \alpha}{2} e^{-itB} e^{-isA} + \frac{1 - \alpha}{2} e^{-isA} e^{-itB}, \quad s, t \in \mathbb{R}, \alpha \in \mathbb{C}, \quad (6.38)$$

and observe that the QJP distributions induced by them naturally admit representation by quasi-probability measures.

Corollary 6.4. *The QJP distributions generated by the hashed operators of the form (6.38) and $|\phi\rangle \in \mathcal{H}$ admits representation by quasi-probability measures.*

In this paper, we call the above sub-family of QJP distributions the *additive complex-parametrised sub-family* of QJP distributions of A and B on $|\phi\rangle$ (or simply, the additive sub-family, for short).

6.3.2. Convolutional Sub-family. One realises below that another class of QJP distributions parametrised by a complex number can be introduced. We hereby consider the hashed operators of the form

$$\#_{\text{cnv}}^\alpha(s, t) := e^{-i\langle s, (1-\alpha)/2 \rangle A} e^{-itB} e^{-i\langle s, (1+\alpha)/2 \rangle A}, \quad s \in \mathbb{C}, t \in \mathbb{R}, \alpha \in \mathbb{C} \quad (6.39)$$

where $\langle s, \alpha \rangle := s_1 \alpha_1 + s_2 \alpha_2$ denotes the inner product of

$$s = s_1 + is_2, \quad \alpha = \alpha_1 + i\alpha_2, \quad (6.40)$$

each of them understood as real vectors of $\mathbb{R}^2 \cong \mathbb{C}$, and introduce the *convolutional complex-parametrised sub-family* of QJP distributions of A and B on $|\phi\rangle$ (or simply, the convolutional sub-family, for short) by those elements of $\mathfrak{M}_{A,B}^\phi$ that are generated by the hashed operators of the form (6.39) and $|\phi\rangle$.

Linear Transformation. It is of natural interest to find out the condition as to when an element of the convolutional sub-family admits representation by quasi-probability measures. Obviously, the choice $\alpha = \pm 1$ admits it, since they are also members of the additive sub-family introduced earlier. As for the other choices of the complex parameter $\alpha \in \mathbb{C}$, we first introduce an auxiliary distribution defined by

$$\tilde{u}(s, t) := \frac{\langle \phi, e^{-is_1 A} e^{-itB} e^{-is_2 A} \phi \rangle}{\|\phi\|^2}, \quad s \in \mathbb{C}, t \in \mathbb{R}, \quad (6.41)$$

where $s = s_1 + is_2 \in \mathbb{C}$, $s_1, s_2 \in \mathbb{R}$ is defined as (6.40). Once there exists a quasi-probability measure $\tilde{\mu}$ such that its Fourier transform coincides with $\mathcal{F}\tilde{\mu} = \tilde{u}$, one finds below that every member of the convolutional sub-family is a linear transform of the quasi-probability measure $\tilde{\mu}$, hence themselves admit representation by quasi-probability measures. To see this, we first introduce the matrix

$$T_\alpha := \begin{pmatrix} (1 - \alpha_1)/2 & (1 + \alpha_1)/2 \\ -\alpha_2/2 & \alpha_2/2 \end{pmatrix}, \quad (6.42)$$

defined for each complex number $\alpha = \alpha_1 + i\alpha_2 \in \mathbb{C}$, $\alpha_1, \alpha_2 \in \mathbb{R}$. The Fourier transform $\mathcal{F}\tilde{\mu}_{(T_\alpha \times I)}$ of the linear transform of the quasi-probability measure $\tilde{\mu}$ with respect to the

operator

$$(T_\alpha \times I)(a, b) := (T_\alpha a, b), \quad a \in \mathbb{C}, b \in \mathbb{R}, \quad (6.43)$$

reads

$$\begin{aligned} (\mathcal{F}\tilde{\mu}_{(T_\alpha \times I)})(s, t) &= \tilde{u}(T_\alpha^* s, t) \\ &= \frac{\langle \phi, e^{-i\langle s, (1-\alpha)/2 \rangle A} e^{-itB} e^{-i\langle s, (1+\alpha)/2 \rangle A} \phi \rangle}{\|\phi\|^2}, \end{aligned} \quad (6.44)$$

where we have used (6.11) in the first equality. We thus have:

Lemma 6.5 (Transformation between Parameters). *Let A and B be self-adjoint operators on \mathcal{H} , and let $|\phi\rangle \in \mathcal{H}$.*

- (i) *If the inverse Fourier transform of the auxiliary distribution (6.41) admits representation by a quasi-probability measure, then all the members of the convolutive sub-family admit representation by quasi-probability measures.*
- (ii) *Under the above situation, let T_α be the linear transformation defined for every choice of the complex parameter $\alpha \in \mathbb{C}$ as in (6.42), and let \tilde{u} be the auxiliary distribution (6.41) and $\tilde{\mu}$ be the quasi-probability measure such that $\mathcal{F}\tilde{\mu} = \tilde{u}$. Then, every member of the convolutive sub-family can be described as the linear transform of $\tilde{\mu}$ as*

$$\mu_{\text{cnv}}^{\phi, \alpha} = \tilde{\mu}_{(T_\alpha \times I)}, \quad (6.45)$$

where $\mu_{\text{cnv}}^{\phi, \alpha}$ denotes the quasi-probability measure generated by the hashed operator of the form (6.39), and $T_\alpha \times I$ is the operator defined as in (6.43).

Representation by Quasi-probability Measures. Now, observing that the determinant of the linear transform T_α reads

$$\det T_\alpha = \text{Im } \alpha / 2, \quad (6.46)$$

one finds that the transformation $\mu \mapsto \mu_{(T_\alpha \times I)}$ is invertible if and only if $\alpha \in \mathbb{C} \setminus \mathbb{R}$, for indeed $T_\alpha \in \text{GL}(2; \mathbb{R}) \Leftrightarrow \alpha \in \mathbb{C} \setminus \mathbb{R}$. The product rule (6.9) then reveals that, one may move from one member of the convolutive sub-family to another by a sequential application of the transformations as

$$\mu^\alpha \xrightarrow{T_\alpha^{-1} \times I} \tilde{\mu} \xrightarrow{T_{\alpha'} \times I} \mu^{\alpha'} \quad (6.47)$$

for the choice $\alpha \in \mathbb{C} \setminus \mathbb{R}$ and $\alpha' \in \mathbb{C}$. Combining Lemma 6.5 with the above observation, one concludes:

Corollary 6.6 (Representation by Quasi-probability Measures). *The following conditions are equivalent.*

- (i) *The inverse Fourier transform of the distribution \tilde{u} defined in (6.41) admits representation by quasi-probability measures.*
- (ii) *A member of the convolutive sub-family for the choice of the parameter $\alpha \in \mathbb{C} \setminus \mathbb{R}$ admits representation by quasi-probability measures.*
- (iii) *Every member of the convolutive sub-family admits representation by quasi-probability measures.*

Explicit Computation of the Members of the convolutive Sub-family. We shall provide an explicit example of the case in which every member of the convolutive complex-parametrised sub-family admits representation by quasi-probability measures.

Proposition 6.7. *Let A and B self-adjoint, and suppose that B has spectrum $\sigma(B)$ of finite cardinality and that it is non-degenerate*

$$B = \sum_{b \in \sigma(B)} b \cdot |b\rangle\langle b|. \quad (6.48)$$

For a quantum state $|\phi\rangle \in \mathcal{H}$ such that the probability of finding the outcomes of B is non-vanishing $|\langle b, \phi\rangle|^2 \neq 0$ for all its eigenvalues $b \in \sigma(B)$, every member of the convolutive sub-family of QJP distributions admits representation by quasi-probability measures.

Proof. Corollary 6.6 purports that it suffices to construct the quasi-probability measure $\tilde{\mu}$ that satisfies $\mathcal{F}\tilde{\mu} = \tilde{u}$, where \tilde{u} is the auxiliary distribution (6.41). Now, under the above conditions, let $b \in \sigma(B)$ and $\Delta_A \in \mathfrak{B}^1$ be fixed, and introduce the Radon-Nikodým derivative

$$\begin{aligned} \nu_b(\Delta_A) &:= (d\nu(\cdot, \Delta_A)/d\mu_B^\phi)(b) \\ &= \frac{\langle b, E_A(\Delta_A)\phi \rangle}{\langle b, \phi \rangle}, \end{aligned} \quad (6.49)$$

where the complex measure $\nu(\cdot, \Delta_A)$ was defined in (6.37). For every fixed $b \in \sigma(B)$, this defines a quasi-probability measure $\Delta_A \mapsto \nu_b(\Delta_A)$, which is in fact nothing but a slight generalisation of the quasi-probability measure (5.66) previously introduced in Section 5. Defining the product complex measure

$$K(b, \cdot) := \nu(b, \cdot)^* \otimes \nu(b, \cdot), \quad (6.50)$$

on the product space $\mathbb{R}^2 \cong \mathbb{C}$ for each $b \in \sigma(B)$, we intend to extend the domain of the variable b to the whole real line to make a transition quasi-probability kernel from $(\mathbb{R}, \mathfrak{B})$ into $(\mathbb{C}, \mathfrak{B}(\mathbb{C}))$ by defining, for example,

$$\tilde{K}(b, \Delta_A) := \begin{cases} K(b, \Delta_A), & (b \in \sigma(B)) \\ \delta_0(\Delta_A), & (b \notin \sigma(B)), \end{cases} \quad (6.51)$$

where δ_0 is the delta measure centred at the origin (for the extension into $\mathbb{R} \setminus \sigma(B)$, we could have assigned any quasi-probability measure so that the extension makes a transition quasi-probability kernel as a whole). Letting $\tilde{\mu}$ denote the quasi-probability measure on the product space $\mathbb{C} \times \mathbb{R}$ defined by \tilde{K} and μ_B^ϕ by means of

$$\int_{\mathbb{C} \times \mathbb{R}} f(a, b) d\tilde{\mu}(a, b) = \int_{\mathbb{R}} \int_{\mathbb{C}} f(a, b) \tilde{K}(b, da) d\mu_B^\phi(b), \quad (6.52)$$

(see (5.1)), we maintain that $(\mathcal{F}\tilde{\mu})(s, t) = \langle \phi, e^{-is_1 A} e^{-itB} e^{-is_2 A} \phi \rangle / \|\phi\|^2$. To see this, just let $f(a, b) := e^{-i\langle s, a \rangle} e^{-itb}$ above and compute

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{C}} e^{-i\langle s, a \rangle} e^{-itb} \tilde{K}(b, da) d\mu_B^\phi(b) &= \sum_{b \in \sigma(B)} \int_{\mathbb{C}} e^{-i\langle s, a \rangle} e^{-itb} K(b, da) \cdot \frac{|\langle b, \phi \rangle|^2}{\|\phi\|^2} \\ &= \sum_{b \in \sigma(B)} e^{-itb} \frac{\langle \phi, e^{-is_1 A} b \rangle \langle b, e^{-is_2 A} \phi \rangle}{\|\phi\|^2} \\ &= \frac{\langle \phi, e^{-is_1 A} e^{-itB} e^{-is_2 A} \phi \rangle}{\|\phi\|^2}, \end{aligned} \quad (6.53)$$

which was to be demonstrated. We have thus achieved a concrete construction of the quasi-probability measure, whose Fourier transform is the distribution \tilde{u} defined in (6.41). \blacksquare

In passing, we note that, by comparing the transformation matrices (6.42) and (5.69), one finds that the quasi-probability measure $\mu_{A,B}^\phi$ obtained in the preceding Section 5 defined as in (5.93) is nothing but the member of the convolutive complex-parametrised sub-family

$$\mu_{A,B}^\phi = \mu_{\text{cnv}}^{\phi,i} \quad (6.54)$$

for the purely imaginary choice $\alpha = i$ of the complex parameter.

6.3.3. Qualification as Quasi-joint-probability Distributions. Although we have provided a formal discussion to the problem, it yet remains to be confirmed by a rigorous treatment that every member of either the additive or the convolutive complex-parametrised sub-family of QJP distributions of a pair of quantum observables indeed qualifies as what its name indicates itself to be. Without loss of generality, we only provide the demonstration for the convolutive sub-family, since the proof for the additive subfamily is essentially the same.

Proposition 6.8 (Qualification as Quasi-joint-probability Distributions). *Let A and B be self-adjoint, $|\phi\rangle \in \mathcal{H}$, and suppose that there exists a quasi-probability measure μ on the product space $\mathbb{C} \times \mathbb{R}$ such that*

$$(\mathcal{F}\mu)(s, t) = \frac{\langle \phi, \#_{\text{cnv}}^\alpha(s, t)\phi \rangle}{\|\phi\|^2} \quad (6.55)$$

holds for some $\alpha \in \mathbb{C}$. Then, μ qualifies as a QJP distribution of A and B on $|\phi\rangle$, in the sense that (5.48) holds.

Proof. We first observe a general result regarding marginals of complex measures and Fourier transformations. Let μ be a complex measure on the product space $\mathbb{R}^m \times \mathbb{R}^n$, and define the marginal of μ by

$$\mu_2 : \Delta \mapsto \mu_2(\Delta) := \mu(\mathbb{R}^m \times \Delta), \quad \Delta \in \mathfrak{B}^n, \quad (6.56)$$

which is itself a complex measure on $(\mathbb{R}^n, \mathfrak{B}^n)$. One then observes

$$\begin{aligned} \hat{\mu}_2(p) &:= \int_{\mathbb{R}^n} e^{-i\langle p, y \rangle} d\mu_2(y) \\ &= \int_{\mathbb{R}^{(m+n)}} e^{-i\langle 0, x \rangle} e^{-i\langle p, y \rangle} d\mu(x, y) \\ &= \hat{\mu}(0, p), \end{aligned} \quad (6.57)$$

where the second equality is due to the change of variables formula (3.10) for the image measure $\mu_2 = \pi_2(\mu)$, where $\pi_2(x, y) = y$, $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$ is the projection on the second variable. Applying this fact to our situation as

$$(\mathcal{F}\mu)(0, t) := \frac{\langle \phi, e^{-itB}\phi \rangle}{\|\phi\|^2} = (\mathcal{F}\mu_B^\phi)(t), \quad (6.58)$$

one readily finds

$$\mu(\mathbb{C} \times \Delta) = \mu_B^\phi(\Delta), \quad \Delta \in \mathfrak{B}^1, \quad (6.59)$$

by the injectivity of the Fourier transformation. One may also demonstrate $\mu(\Delta \times \mathbb{R}) = \mu_A^\phi(\Delta)$, $\Delta \in \mathfrak{B}(\mathbb{K})$ by an analogous reasoning, which completes our proof. \blacksquare

6.3.4. *Relation to other known Proposals.* We demonstrate below, in passing, that the complex-parametrised sub-families of the QJP distributions of a pair of quantum observables serve as generalisations to the other well known proposals of quasi-probability distributions.

Kirkwood-Dirac Distribution. We first note that the Kirkwood-Dirac distribution, introduced in (5.2) in a formal manner, can be given a mathematically rigorous definition within our framework, and that it belongs to both the additive and convolutive sub-families of the QJP distributions for the choice $\alpha = 1$.

Definition (Kirkwood-Dirac Quasi-joint-probability Distribution). *Let A and B be self-adjoint on \mathcal{H} , and let $|\phi\rangle \in \mathcal{H}$. We call the member of the additive/convolutive sub-family of the QJP distributions of the pair of observables A and B for the choice $\alpha = 1$, the Kirkwood-Dirac QJP distribution of the pair.*

To see how this definition can be justified, observe the following formal chain of expressions

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-i(as+bt)} K_{A,B}^\phi(a,b) dm_2(a,b) &= \int_{\mathbb{R}^2} e^{-i(sa+tb)} \frac{\langle \phi, b \rangle \langle b, a \rangle \langle a, \phi \rangle}{\|\phi\|^2} dm_2(a,b) \\ &= \frac{\langle \phi, e^{-itB} e^{-isA} \phi \rangle}{\|\phi\|^2}, \end{aligned} \quad (6.60)$$

where $K_{A,B}^\phi$ is the formal definition of the Kirkwood-Dirac distribution introduced in (5.2). The injectivity of the Fourier transformation leads to the desired statement.

Wigner-Ville Distribution. We next note that the Wigner-Ville distribution, introduced in (5.43), is also a special member of the convolutive sub-family of QJP distributions.

Proposition 6.9 (Wigner-Ville Distribution). *Let $\{L^2(\mathbb{R}), \mathcal{S}(\mathbb{R}), \{\hat{x}, \hat{p}\}\}$ denote the one-dimensional Schrödinger representation of the CCR. Then, for the choice $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ of the wave-function, the member of the convolutive sub-family of the QJP distributions of the canonically conjugate pair \hat{p} and \hat{x} admits representation by quasi-probability measures for the choice $\alpha = 0$, which we denote by $\mu_{\text{cinv}}^{\psi,0}$. The quasi-probability measure $\mu_{\text{cinv}}^{\psi,0}$ is absolutely continuous, and its Radon-Nikodým derivative with respect to the renormalised two-dimensional Lebesgue-Borel measure reads*

$$\left(d\mu_{\text{cinv}}^{\psi,0} / dm_2 \right) (p, x) := W^\psi(x, p), \quad (6.61)$$

where the r. h. s. is the Wigner-Ville distribution introduced in (5.43).

Proof. Observe that the condition $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ guarantees the integrability $W^\psi \in L^1(\mathbb{R}^2)$ of the WV distribution, based on which we compute

$$\begin{aligned}
& \int_{\mathbb{R}^2} e^{-i(sp+tx)} W^\psi(x, p) \, dm_2(x, p) \\
& := \int_{\mathbb{R}^2} e^{-i(sp+tx)} \left(\int_{\mathbb{R}} \psi^*(x + y/2) \psi(x - y/2) e^{ipy} \, dm(y) \right) dm_2(x, p) \\
& = \int_{\mathbb{R}} e^{-itx} \psi^*(x + s/2) \psi(x - s/2) \, dm(x) \\
& = \int_{\mathbb{R}} e^{-itx} (e^{is\hat{p}/2} \psi)^*(x) (e^{-is\hat{p}/2} \psi)(x) \, dm(x) \\
& = \langle e^{is\hat{p}/2} \psi, e^{-it\hat{x}} e^{-is\hat{p}/2} \psi \rangle \\
& = \mathcal{F} \left(\mu_{\text{cnv}}^{\psi, 0} \right). \tag{6.62}
\end{aligned}$$

Combining (6.3) and the injectivity of the Fourier transformation, one arrives at the desired statement. \blacksquare

6.4. Some General Properties

We next observe some general properties of QJP distributions. We first provide some discussion regarding the operation of taking the complex conjugate, and subsequently seek for the condition for their realness.

6.4.1. Complex Conjugate. We are interested in the complex conjugate of QJP distributions of a pair of observables A and B on $|\phi\rangle$. To this, let $\#$ be a hashed operator of the unitary operators e^{-isA} , e^{-itB} , and let $|\phi\rangle \in \mathcal{H}$ be such that the QJP distribution generated by them admits representation by a quasi-probability measure μ . By applying (6.12), one readily finds that the Fourier transform of the complex conjugate μ^* reads

$$\begin{aligned}
(\mathcal{F}\mu^*)(s, t) &= (\mathcal{F}\mu)^*(-s, -t) \\
&= \frac{\langle \#(-s, -t)\phi, \phi \rangle}{\|\phi\|^2} \\
&= \frac{\langle \phi, \#(-s, -t)^*\phi \rangle}{\|\phi\|^2}, \quad s, t \in \mathbb{K} \tag{6.63}
\end{aligned}$$

where $\#(s, t)^*$ denotes the ‘adjoint’ of the hashed operator. Since the ‘involution’ $\#(-s, -t)^*$ is itself a hashed operator of e^{-isA} and e^{-itB} , one concludes that the complex conjugate μ^* is again a QJP distribution of the pair of observables A and B , and that it is precisely the distribution generated by the ‘involution’ of the original hashed operator. One also specifically finds that the sub-family of the QJP distributions $\mathfrak{M}_{A, B}^\phi$ that admit representations by quasi-probability measures is closed under the operation of taking the complex conjugate.

Parallel to this, by observing that the left most hand side of (6.63) can be written as

$$(\mathcal{F}\mu^*)(s, t) = \frac{\langle \phi, (\mathcal{F}\Pi^*)(s, t)\phi \rangle}{\|\phi\|^2}, \quad s, t \in \mathbb{K} \tag{6.64}$$

where $\Pi = \mathcal{F}^{-1}\#$ is the quasi-joint-spectral distribution, one concludes the validity of the equality

$$\begin{aligned} (\mathcal{F}\Pi^*)(s, t) &= \#(-s, -t)^* \\ &= (\mathcal{F}\Pi)(-s, -t)^* \\ &=: (\mathcal{F}\Pi)^\dagger(s, t), \quad s, t \in \mathbb{K}, \end{aligned} \tag{6.65}$$

where $(\mathcal{F}\Pi)^\dagger$ denotes the ‘involution’ (observe the analogy between (6.12)). This shows that the ‘adjoint’ Π^* of the quasi-joint-spectral distribution of A and B is again a quasi-joint-spectral distribution of the pair, and that it is precisely the inverse Fourier transform of the ‘involution’ of the original hashed operator.

Complex-parametrised Sub-families. Armed with our findings, one may explicit compute the complex conjugate of the elements of both the additive and the convolutive sub-families, and see that the sub-families are also closed under the operation of taking the complex conjugate. Indeed, if we respectively introduce

$$\Pi_{\text{add}}^\alpha := \mathcal{F}^{-1}\#_{\text{add}}^\alpha, \quad \Pi_{\text{cnv}}^\alpha := \mathcal{F}^{-1}\#_{\text{cnv}}^\alpha, \tag{6.66}$$

for the members of the additive and convolutive sub-families, by observing that the ‘involution’ of the hashed operators read

$$\#_{\text{add}}^\alpha(-s, -t)^* = \#_{\text{add}}^{-\alpha^*}(s, t), \tag{6.67}$$

$$\#_{\text{cnv}}^\alpha(-s, -t)^* = \#_{\text{cnv}}^{-\alpha}(s, t), \tag{6.68}$$

one finds

$$(\Pi_{\text{add}}^\alpha)^* = \Pi_{\text{add}}^{-\alpha^*}, \tag{6.69}$$

$$(\Pi_{\text{cnv}}^\alpha)^* = \Pi_{\text{cnv}}^{-\alpha}. \tag{6.70}$$

This provide explicit formulae for the computation of the complex conjugate of the members of the sub-families, and one specifically finds from it that both the sub-families are closed under the operation of taking the complex conjugate as promised. Now, fixing $|\phi\rangle \in \mathcal{H}$ of one’s choice, one observes:

Lemma 6.10 (Complex Conjugate: Additive Sub-family). *Let μ_{add}^α denote the QJP distribution of A and B generated by $\#_{\text{add}}^\alpha$ and $|\phi\rangle$. Then, its complex conjugate reads*

$$\left(\mu_{\text{add}}^{\phi, \alpha}\right)^* = \mu_{\text{add}}^{\phi, -\alpha^*}. \tag{6.71}$$

Lemma 6.11 (Complex Conjugate: Convolutive Sub-family). *Suppose that the member of the convolutive sub-family admits representation by the quasi-probability measure $\mu_{\text{cnv}}^{\phi, \alpha}$ for the choice $\alpha \in \mathbb{C}$. Then, the member for the choice $-\alpha \in \mathbb{C}$ also admits representation by quasi-probability measures, and the equality*

$$\left(\mu_{\text{cnv}}^{\phi, \alpha}\right)^* = \mu_{\text{cnv}}^{\phi, -\alpha}. \tag{6.72}$$

holds.

6.4.2. *Realness of the QJP Distributions.* One may naturally be interested in the condition as to when the quasi-joint-spectral distribution $\Pi = \mathcal{F}^{-1}\#$ becomes ‘self-adjoint’ so that the resulting QJP distribution, symbolically denoted by $p(a, b) = \langle \phi, \Pi(a, b)\phi \rangle / \|\phi\|^2$, is also ‘real’ for any choice of the vector $|\phi\rangle \in \mathcal{H}$. While the task of finding the explicit condition for which $\Pi(a, b)$ becomes ‘self-adjoint’ seems at first non-trivial, the problem becomes significantly tractable if one considers its Fourier transform. Indeed, combining (6.65) with the injectivity of the Fourier transform, one concludes that $\Pi = \Pi^*$ is ‘self-adjoint’ if and only if its Fourier transform (namely, the hashed operator) $\# = \#^\dagger$ is a ‘self-involution’. Examples of such ‘self-involutive’ hashed operators are provided by

$$\#(s, t) = \begin{cases} e^{-itB/2} e^{-isA} e^{-itB/2}, \\ e^{-isA/2} e^{-itB} e^{-isA/2}, \\ \frac{1}{2} (e^{-itB} e^{-isA} + e^{-isA} e^{-itB}), \\ e^{-itB/L_N} e^{-isA/M_N} \dots e^{-itB/L_1} e^{-isA/M_1} e^{-itB/L_1} \dots e^{-isA/M_N} e^{-itB/L_N}, \\ e^{-i\overline{(sA+tB)}} = \lim_{N \rightarrow \infty} (e^{-isA/N} e^{-itB/N})^N, \\ \text{etc.}, \end{cases} \quad (6.73)$$

where $\sum_{k=1}^N M_k^{-1} = 1$, $\sum_{k=1}^N L_k^{-1} = 1$ in the third example. Colloquially speaking, hashed operators in which the disintegrated components of the unitary operators appear ‘symmetrically’ provide straightforward examples. As for our concrete examples, one finds:

Corollary 6.12 (Condition for Realness). *A member of either the additive or convolutive sub-families of QJP distributions of A and B for the choice $\alpha = 0$ is always real.*

6.5. Conditioned Measurement Revisited

We finally investigate how the CM scheme described in Section 5 fits into our general framework of quasi-joint-probabilities of quantum observables. What we see below is that the CM scheme is essentially *a measurement scheme for measuring QJP distributions of an arbitrary pair of quantum observables*. As before, since the tools for the analysis of the most general cases are beyond the scope of this paper, we shall exclusively concentrate on the subfamily of quasi-joint-probabilities parametrised by a single complex number. Without loss of generality, we only provide below a demonstration for the convolutive sub-family for simplicity.

6.5.1. *Short Introduction.* We now intend to construct a measurement scheme for obtaining the member of the convolutive sub-family of the QJP distributions for arbitrary choices of the parameter $\alpha \in \mathbb{C}$. As for the problem, let us first recall that the quasi-probability measure (5.93) obtained in Section 5 was nothing but the member for the choice of the parameter $\alpha = i$ (see (6.54) for the discussion). In fact, as we have seen before, once we know the member of the subfamily for the parameter $\alpha \in \mathbb{C} \setminus \mathbb{R}$, we may compute all other members of the complex parameters by sequentially applying linear transformations as depicted in (6.47). Hence, the knowledge of the distribution for the choice $\alpha = i$, obtained by means of the CM scheme in view of the WV distribution, actually suffices for our purpose. Even so, one might be interested in how one could measure the QJP distribution for some specific parameter in a more direct manner. This should also provide a much more transparent view of the

measurement scheme described in Section 5 from a more general viewpoint, which may be beneficial in its own right.

Model and Assumption. Throughout this subsection, we let A denote an observable on the target system \mathcal{H} , and assume that the meter system \mathcal{K} is described by the one-dimensional Schrödinger representation of the CCR $\{L^2(\mathbb{R}), \mathcal{S}(\mathbb{R}), \{\hat{x}, \hat{p}\}\}$ for simplicity. As usual, we prepare the two systems into their respective initial states $|\phi\rangle \in \mathcal{H}$, $|\psi\rangle \in \mathcal{K}$, and let them interact under the unitary operator $e^{-igA \otimes Y}$, $g \in \mathbb{R}$, for which we choose $Y = \hat{p}$ for definiteness, and let $|\Psi^g\rangle$ denote the state of the composite system after the interaction. Since we intend to confine ourselves within the framework of complex measures, we place several conditions throughout this passage, so that, given a conditioning observable B on the target system \mathcal{H} , all the members of the convolutive sub-family of the QJP distributions of A and B on $|\phi\rangle$ admits representation by quasi-probability measures.

6.5.2. Conditioned Measurement Revisited. In the previous section, the choice of the QJP distribution we intend to measure on the meter system was the WV distribution, which we found to be nothing but the member of our convolutive sub-family of the QJP distributions of the canonically conjugate pair of observables $A = \hat{p}$, $B = \hat{x}$ for the choice of the parameter $\alpha = 0$. The result was that, one could obtain the member of the convolutive sub-family of the QJP distributions for arbitrary pairs of quantum observables for the choice $\alpha = i$. Motivated by this finding, it is then natural to conjecture that a different choice of the meter QJP distribution results in different choice of the target QJP distribution.

Meter QJP. The starting point would be to find the equivalent object to ω^ψ for the other choices of the parameter $\alpha \in \mathbb{C}$. To this end, we first assume $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, and introduce the function

$$\tilde{\omega}^{\psi, \alpha_1}(x, y) := \psi^*(x - y(1 - \alpha_1)/2) \psi(x + y(1 + \alpha_1)/2) \quad (6.74)$$

and also its Fourier transform

$$\tilde{W}^{\psi, \alpha_1}(x, p) := \int_{\mathbb{R}} e^{-ipy} \tilde{\omega}^{\psi, \alpha_1}(x, y) dm(y), \quad (6.75)$$

where we let $\alpha = \alpha_1 + i\alpha_2$, $\alpha_1, \alpha_2 \in \mathbb{R}$. Needless to say, the function $\tilde{\omega}^{\psi, 0} = \tilde{\omega}^\psi$ for the choice $\alpha_1 = 0$ reduces to the original function introduced in (5.36), and thus $\tilde{W}^{\psi, 0} = \tilde{W}^\psi$ is nothing but the (yet-to-be-normalised) WV distribution. By computing the Fourier transform

$$\begin{aligned} \left(\mathcal{F}\tilde{W}^{\psi, \alpha_1}\right)(q, y) &= \int_{\mathbb{R}^2} e^{-i(qx+yp)} \left(\int_{\mathbb{R}} e^{-ipy} \tilde{\omega}^{\psi, \alpha_1}(x, y) dm(y) \right) dm_2(x, p) \\ &= \int_{\mathbb{R}} e^{-iqx} \tilde{\omega}^{\psi, \alpha_1}(x, -y) dm(x) \\ &= \int_{\mathbb{R}} e^{-iqx} \psi^*(x + y(1 - \alpha_1)/2) \psi(x - y(1 + \alpha_1)/2) dm(x) \\ &= \left\langle \psi, e^{-i\langle y, (1-\alpha_1)/2 \rangle \hat{p}} e^{-iq\hat{x}} e^{-i\langle y, (1+\alpha_1)/2 \rangle \hat{p}} \psi \right\rangle, \end{aligned} \quad (6.76)$$

one concludes from the injectivity of the Fourier transformation that the normalisation

$$\begin{aligned} W^{\psi, \alpha_1} &:= \tilde{W}^{\psi, \alpha_1} / \|\psi\|^2 \\ &= \left(d\mu_{\text{cnv}}^{\psi, \alpha_1} / dm_2 \right) \end{aligned} \quad (6.77)$$

is nothing but the Radon-Nikodým derivative of the member of the convolutive sub-family of the QJP distributions of the canonically conjugate pair $A = \hat{p}$, $B = \hat{x}$ for the choice of the real parameter $\alpha_1 \in \mathbb{R}$.

Rescaling. For simplicity of the argument, we only treat the case for the choice $\alpha \in \mathbb{C} \setminus \mathbb{R}$, and for later convenience, we introduce the function

$$\tilde{v}^{\psi, \alpha}(x, y) := |2/\alpha_2|^{-1} \tilde{\omega}^{\psi, -\alpha_1}(x, (2/\alpha_2)y), \quad (\alpha_2 \neq 0) \quad (6.78)$$

for a given choice of the parameter $\alpha \in \mathbb{C} \setminus \mathbb{R}$ (note the minus sign for the real part $\alpha_1 := \text{Re } \alpha$ in the definition). Its Fourier transform then reads

$$\begin{aligned} \int_{\mathbb{R}} e^{-iqx} \tilde{v}^{\psi, \alpha}(x, -y) dm(x) &= \left\langle \psi, e^{-i\langle (2/\alpha_2)y, (1+\alpha_1)/2 \rangle \hat{p}} e^{-iq\hat{x}} e^{-i\langle (2/\alpha_2)y, (1-\alpha_1)/2 \rangle \hat{p}} \psi \right\rangle \\ &= \left\langle \psi, e^{-i\langle y, (1+\alpha_1)/\alpha_2 \rangle \hat{p}} e^{-iq\hat{x}} e^{-i\langle y, (1-\alpha_1)/\alpha_2 \rangle \hat{p}} \psi \right\rangle, \end{aligned} \quad (6.79)$$

where we have combined the second and the last equality of (6.76), and applied the result (5.32).

QJP of the ‘conditional’ Meter State. The next step is to compute the function $\tilde{v}^{\psi_b^g, \alpha}$ for the ‘conditional’ meter state $\psi_b^g := \psi_{B=b}^g$ introduced in (5.65). What we find below is that, parallel to the findings in Section 5, the resulting function $\tilde{v}^{\psi_b^g, \alpha}$ is provided by the convolution of the initial profiles of both the meter and the target configurations. As above, we assume, for the ease of demonstration, that both the target and the conditioning observables A and B have spectra of finite cardinality, that B is degenerate, and the probability of finding the outcomes of B is non-vanishing $|\langle b, \phi \rangle|^2 \neq 0$ for all its eigenvalues $b \in \sigma(B)$. Since the essence of the demonstration is substantially the same as those provided in Section 5, we proceed by sketching the proofs.

In computing the function of our interest, we first compute its Fourier transform to observe

$$\begin{aligned} &\int_{\mathbb{R}} e^{-iqx} \tilde{v}^{\psi_b^g, \alpha}(x, -y) dm(x) \\ &= \left\langle \psi_b^g, e^{-i\langle y, (1+\alpha_1)/\alpha_2 \rangle \hat{p}} e^{-iq\hat{x}} e^{-i\langle y, (1-\alpha_1)/\alpha_2 \rangle \hat{p}} \psi_b^g \right\rangle \\ &= \int_{\mathbb{R}^2} \left\langle e^{-igs_1 \hat{p}} \psi, e^{-i\langle y, (1+\alpha_1)/\alpha_2 \rangle \hat{p}} e^{-iq\hat{x}} e^{-i\langle y, (1-\alpha_1)/\alpha_2 \rangle \hat{p}} e^{-igs_2 \hat{p}} \psi \right\rangle d(\nu_b^* \otimes \nu_b)(s_1, s_2), \end{aligned} \quad (6.80)$$

where we have used (6.79) in the first equality, and where ν_b is the quasi-probability measure introduced in (5.66). We next change variables of the above equality according to the linear

transformation

$$\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = T_\alpha \begin{pmatrix} s_1 \\ s_2 \end{pmatrix}, \quad T_\alpha := \begin{pmatrix} (1 - \alpha_1)/2 & (1 + \alpha_1)/2 \\ -\alpha_2/2 & \alpha_2/2 \end{pmatrix} \quad (6.81)$$

by substituting

$$s_1 = a_1 - \frac{1 + \alpha_1}{\alpha_2} a_2, \quad s_2 = a_1 + \frac{1 - \alpha_1}{\alpha_2} a_2, \quad (6.82)$$

to find

$$\begin{aligned} & \int_{\mathbb{R}} e^{-iqx} \tilde{\nu}^{\psi_b^g, \alpha}(x, -y) dm(x) \\ &= \int_{\mathbb{R}^2} \left\langle \psi, e^{-i(y - ga_2, (1 + \alpha_1)/\alpha_2)\hat{p}} e^{-i(q\hat{x} - ga_1 I)} e^{-i(y - ga_2, (1 - \alpha_1)/\alpha_2)\hat{p}} e^{-igs\hat{p}} \psi \right\rangle d\mu_A^{\phi, \alpha}(a|B = b) \\ &= \int_{\mathbb{R}^2} \left(\int_{\mathbb{R}} e^{-iqx} \tilde{\nu}^{\psi, \alpha}(x - ga_1, -(y - ga_2)) dm(x) \right) d\mu_A^{\phi, \alpha}(a|B = b) \\ &= \int_{\mathbb{R}} e^{-iqx} \left(\int_{\mathbb{R}^2} \tilde{\nu}^{\psi, \alpha}(x - ga_1, -(y - ga_2)) d\mu_A^{\phi, \alpha}(a|B = b) \right) dm(x), \end{aligned} \quad (6.83)$$

where $\mu_A^{\phi, \alpha}(\Delta|B = b) := (\nu_b^* \otimes \nu_b)(T_\alpha^{-1}\Delta)$, $\Delta \in \mathfrak{B}^2$ is the image measure, and we have combined (6.79) with (5.33) to obtain the second equality. One thus concludes from the injectivity of the Fourier transformation that

$$\tilde{\nu}^{\psi_b^g, \alpha}(x, y) = \int_{\mathbb{R}^2} \tilde{\nu}^{\psi, \alpha}(x - ga_1, y - ga_2) d\mu_A^{\phi, \alpha}(a|B = b), \quad (6.84)$$

as promised.

Recovery of the Target QJP. As for the recovery of the target information $\Delta \mapsto \mu_A^{\phi, \alpha}(\Delta|B = b)$, $\Delta \in \mathfrak{B}(\mathbb{C})$, one may resort to the familiar techniques we have discussed so far in depth, namely, one may recover the full profile by either probing the strong or the weak region of the interaction parameter. Once we obtained $\mu_A^{\phi, \alpha}(\cdot|B = b)$ for all $b \in \sigma(B)$, one may extend the domain of $b \in \sigma(B)$ to the whole real line \mathbb{R} in a consistent manner, making it a transition quasi-probability kernel. This allows us to construct the QJP $\mu_{A, B}^{\phi, \alpha}$ of the pair of A and B in a manner described in Proposition 5.1 that satisfies

$$\mu_{A, B}^{\phi, \alpha}(\Delta_A \times \Delta_B) = \int_{\Delta_B} \mu_A^{\phi, \alpha}(\Delta_A|B = b) d\mu_B^\phi(b), \quad \Delta_A \in \mathfrak{B}(\mathbb{C}), \Delta_B \in \mathfrak{B}^1. \quad (6.85)$$

A close look on the proof of Proposition 6.7 leads one to conclude that the QJP obtained here

$$\mu_{A, B}^{\phi, \alpha} = \mu_{\text{cnv}}^{\phi, \alpha}, \quad (6.86)$$

is in fact nothing but the member of the convolutive sub-family for the choice $\alpha \in \mathbb{C} \setminus \mathbb{R}$, and that $\mu_A^{\phi, \alpha}(\cdot|B = b)$ is the conditional quasi-probability distribution of A given $B = b$.

7. Application: Interpretation of Aharonov's Weak Value

As an application of the findings on the QJP distributions of quantum observables, we now focus on the geometric structure that the QJP distributions induce in the space of quantum observables. Specifically, by drawing an analogy between the result of classical probability theory, we provide a geometric and statistical interpretation of Aharonov's weak value as 'orthogonal projection' and 'conditional average', respectively.

7.1. Reference Materials

As usual, we start by preparing some necessary materials that become useful for our analysis. The main objective of this subsection is to obtain a geometric understanding of conditional expectations in classical probability theory.

7.1.1. L^2 -Theory of Conditional Expectations.

L^p -spaces for finite Measures. Let μ be a finite measure on a measurable space (X, \mathfrak{A}) , i.e., $\mu(X) < \infty$, and let $1 \leq p < q \leq \infty$. By defining $r > 0$ satisfying $\frac{1}{r} = \frac{1}{p} - \frac{1}{q}$, a direct application of Hölder's inequality yields

$$\|f\|_p \leq \|f\|_q \cdot \|1\|_r = \|f\|_q \cdot |\mu(X)|^{1/r} < \infty \quad (7.1)$$

for $f \in L^q(\mu)$. The following Lemma is worth of special notice.

Lemma 7.1. *Let μ be a finite measure on a measurable space (X, \mathfrak{A}) . Then, for any $1 \leq p \leq q \leq \infty$, the relation*

$$L^q(\mu) \subset L^p(\mu) \quad (7.2)$$

holds.

Specifically, for probability spaces, note that one has the evaluation $\|f\|_p \leq \|f\|_q$ for the choice of the parameters $1 \leq p \leq q \leq \infty$.

Conditional Expectations for square-integrable Functions. Now, consider a probability space $(\mathbb{R}^n, \mathfrak{B}^n, \mu)$, and let $\mathfrak{A} \subset \mathfrak{B}^n$ be a sub- σ -algebra. Since every square-integrable function $f \in L^2(\mu) \subset L^1(\mu)$ is integrable due to the above Lemma, its conditional expectation $\mathbb{E}[f|\mathfrak{A}] := d(f \odot \mu)|_{\mathfrak{A}}/d\mu|_{\mathfrak{A}} \in L^1(\mu|_{\mathfrak{A}})$ is well-defined, where $(f \odot \mu)|_{\mathfrak{A}}$ and $\mu|_{\mathfrak{A}}$ denotes the restriction of the respective (complex) measures on the sub- σ -algebra. Now, observe that, for any square-integrable function $g \in L^2(\mu|_{\mathfrak{A}})$, the equality

$$\begin{aligned} \langle g, f \rangle &:= \int_{\mathbb{R}^n} g^*(x) f(x) d\mu(x) \\ &= \int_{\mathbb{R}^n} g^*(x) d(f \odot \mu)(x) \\ &= \int_{\mathbb{R}^n} g^*(x) \frac{d(f \odot \mu)|_{\mathfrak{A}}}{d\mu|_{\mathfrak{A}}}(x) d\mu|_{\mathfrak{A}}(x) \\ &=: \langle g, \mathbb{E}[f|\mathfrak{A}] \rangle \end{aligned} \quad (7.3)$$

holds by the definition of the Radon-Nikodým derivative. Specifically, note that this leads to the fact that the conditional expectation $\mathbb{E}[f|\mathfrak{A}] \in L^2(\mu|_{\mathfrak{A}})$ of a square-integrable function $f \in L^2(\mu)$ is again square-integrable.

Conditioning as Projection. Another important observation to make from the above equality is that, the act of conditioning

$$\mathbb{E}[\cdot|\mathfrak{A}] : L^2(\mu) \rightarrow L^2(\mu|\mathfrak{A}), \quad f \mapsto \mathbb{E}[f|\mathfrak{A}] \quad (7.4)$$

that takes a μ -square-integrable function to its conditional expectation, is an *orthogonal projection*. To see this, first observe that linearity $\mathbb{E}[af + bg|\mathfrak{A}] = a\mathbb{E}[f|\mathfrak{A}] + b\mathbb{E}[g|\mathfrak{A}]$, $f, g \in L^2(\mu)$, $a, b \in \mathbb{C}$ follows immediately by definition (naturally, equality is only valid $\mu|_{\mathfrak{A}}$ -almost everywhere). Now, since $L^2(\mu|\mathfrak{A})$ is itself a complex Hilbert space, it is a topologically closed subspace of the larger complex Hilbert space $L^2(\mu)$. By recalling that there is a one-to-one correspondence between closed subspaces and orthogonal projections in Hilbert spaces, let $P(\cdot|\mathfrak{A}) : L^2(\mu) \rightarrow L^2(\mu|\mathfrak{A})$ denote the unique orthogonal projection associated with it. By observing that

$$\langle g, f \rangle = \langle g, P(f|\mathfrak{A}) \rangle \quad (7.5)$$

holds for all $g \in L^2(\mu|\mathfrak{A})$ and $f \in L^2(\mu)$ by definition of orthogonal projections, one realises that the equality (7.3) combined with the non-degenerateness of inner products leads to

$$P(f|\mathfrak{A}) = \mathbb{E}[f|\mathfrak{A}], \quad f \in L^2(\mu). \quad (7.6)$$

We summarise the results as follows.

Proposition 7.2. (*Orthogonal Projection and Conditional Expectation*) *Let μ be a probability measure on $(\mathbb{R}^n, \mathfrak{B}^n)$, and let $\mathfrak{A} \subset \mathfrak{B}^n$ be a sub- σ -algebra. Then, the unique orthogonal projection $P(\cdot|\mathfrak{A}) : L^2(\mu) \rightarrow L^2(\mu|\mathfrak{A})$ associated with the subspace $L^2(\mu|\mathfrak{A})$ is provided by the conditional expectation*

$$P(f|\mathfrak{A}) = \mathbb{E}[f|\mathfrak{A}], \quad (7.7)$$

where $f \in L^2(\mu)$.

We here see how the geometric concept of orthogonality relates to the statistical concept of conditioning in L^2 -spaces.

7.1.2. Conditioning as Optimal Approximation. The geometric property mentioned above leads to several important interpretation of conditional expectations. One of the prominent characteristics of orthogonal projections is the validity of the *Pythagorean identity*

$$\|f - g\|_2^2 = \|f - \mathbb{E}[f|\mathfrak{A}]\|_2^2 + \|\mathbb{E}[f|\mathfrak{A}] - g\|_2^2, \quad f \in L^2(\mu), \quad g \in L^2(\mu|\mathfrak{A}), \quad (7.8)$$

where $\|\cdot\|_2$ denotes the standard L^2 -norm introduced earlier in (2.19). An immediate consequence of the above Pythagorean identity is the following equality

$$\|f - \mathbb{E}[f|\mathfrak{A}]\|_2 = \min_{g \in L^2(\mu|\mathfrak{A})} \|f - g\|_2, \quad f \in L^2(\mu), \quad (7.9)$$

which states that the optimal $\mu|_{\mathfrak{A}}$ -square-integrable function one can find in approximating a function f is explicitly provided by the conditional expectation of f given \mathfrak{A} , and the positive-definiteness of the L^2 -norm shows that the optimum is unique $\mu|_{\mathfrak{A}}$ -a.e.

7.2. Statistical Interpretation of Geometric Structures

Now that we have reviewed the geometric interpretation of conditional expectations in classical probability theory, we shall begin our main analysis.

7.2.1. *Preliminary Observation.* In classical probability theory, probability measures equip the space of square-integrable functions with a geometry, *i.e.*, an inner product defined by (2.21), which we reiterate for the readers' convenience as

$$\langle g, f \rangle_\mu := \int g^* f \, d\mu \quad (7.10)$$

(here, we have also explicitly written the probability measure μ under consideration for clarity). In the context of classical physics in which observables are represented by functions, a probability measure defines quantities on a given pair of square-integrable classical observables interpreted as *correlations* or *covariances* between them.

'Correlations' in Quantum Theory. In quantum mechanics, observables are represented by self-adjoint operators on Hilbert spaces, and the statistics of the system are in turn represented by vectors of Hilbert spaces. In order to see how the two distinct frameworks of classical and quantum theory on correlations play together, first let A and B be a pair of simultaneously measurable bounded quantum observables on a Hilbert space \mathcal{H} , $|\psi\rangle \in \mathcal{H}$ a vector, and introduce

$$\langle B, A \rangle_\psi := \frac{\langle B\psi, A\psi \rangle}{\|\psi\|^2}. \quad (7.11)$$

One readily sees that, for the present case, the above geometry induced by the vector $|\psi\rangle$ is in accordance with the classical theory. Indeed, the unique product spectral measure $E_{A,B}$ of A and B admits a unique representation of the observables given by

$$A = \int_{\mathbb{R}^2} A(a, b) \, dE_{A,B}(a, b), \quad B = \int_{\mathbb{R}^2} B(a, b) \, dE_{A,B}(a, b) \quad (7.12)$$

with $A(a, b) = a$, $B(a, b) = b$, and moreover defines a joint-probability measure $\mu_{A,B}^\psi$ of the pair (see (3.69)) on the state $|\psi\rangle \in \mathcal{H}$. It is then straightforward to see the validity of the equality

$$\begin{aligned} \langle B, A \rangle_\psi &= \int_{\mathbb{R}^2} B^*(a, b) A(a, b) \, d\mu_{A,B}^\psi(a, b) \\ &= \langle B, A \rangle_{\mu_{A,B}^\psi}, \end{aligned} \quad (7.13)$$

based on which one obtains a statistical interpretation of the geometry (7.11) as the correlation or covariance between the observables in the classical sense.

Non-commutative Case. On the other hand, the problem is not so straightforward for the case where the pair A and B does not admit simultaneous measurability. This is essentially to do with the lack of the unique product spectral measure of the pair. As we have seen in the previous Section 6, the non-commutative analogues of product spectral measures are the quasi-joint-spectral distributions (QJSDs) defined as the inverse Fourier transforms of the hashed unitary groups (6.22). The non-uniqueness of the QJSDs for the non-commuting case generally leads to the non-uniqueness of the representation of operators and vectors by functions and quasi-probability distributions. Specifically, given a choice of a QJSD $\Pi_{A,B}$ for a pair of generally non-commuting observables A and B , one obtains a functional

representation of operators as

$$A = \int_{\mathbb{R}^2} A(a, b) d\Pi_{A,B}(a, b), \quad B = \int_{\mathbb{R}^2} B(a, b) d\Pi_{A,B}(a, b) \quad (7.14)$$

with $A(a, b) = a$, $B(a, b) = b$ (fortunately, as for this specific case, all representations coincide irrespective of the choice of the QJSD), and also a representation of quantum states $|\psi\rangle \in \mathcal{H}$ by QJP distributions $p_{A,B}^\psi(a, b)$ defined as in (6.28). Guided by a straightforward analogy, one realises that the quantity formally defined by

$$\langle\langle B, A \rangle\rangle_\psi := \int_{\mathbb{R}^2} B^*(a, b) A(a, b) p_{A,B}^\psi(a, b) dm_2(a, b) \quad (7.15)$$

defines various different geometries between quantum observables dependent on the choice of the QJSDs. This implies that, parallel to the classical case, QJSDs serves as a bridge that offers a ‘statistical’ interpretation of the (non-unique) geometric structures that can be introduced in the space of quantum observables.

7.2.2. Geometry induced by a specific Sub-family of QJSD. The general treatment involving the entire class of QJSDs makes extensive use of the theory of generalised functions and its operator valued analogue (operator valued distributions), which may be far beyond the scope of this paper. In this paper, mainly in order to confine our argument in the theory of complex measures and its operator valued analogue, we concentrate on the specific choice of the QJSD of a pair of quantum observables, namely, to the additive sub-family (introduced in Section 6.3.1) for the choice $-1 \leq \alpha \leq 1$ of the complex parameter, hopefully without essential loss of generality. We shall moreover confine ourselves to bounded operators for simplicity, but the general treatment including unbounded operators is also possible without any essential alteration.

Sesquilinear Forms. We are interested in introducing geometries in the space $L(\mathcal{H})$ of all bounded linear operators on a Hilbert space \mathcal{H} given a fixed state $|\psi\rangle \in \mathcal{H}$. To this end, let $\alpha \in \mathbb{C}$ be a complex number, and define

$$\langle\langle Y, X \rangle\rangle_{\psi, \alpha} := \frac{1 + \alpha}{2} \cdot \frac{\langle Y\psi, X\psi \rangle}{\|\psi\|^2} + \frac{1 - \alpha}{2} \cdot \frac{\langle X^*\psi, Y^*\psi \rangle}{\|\psi\|^2}, \quad X, Y \in L(\mathcal{H}). \quad (7.16)$$

As described earlier, this is just one possible straightforward way to extend the geometry (7.11) so that it can be defined even for non-commuting observables. One readily sees that this satisfies

- (i) $\langle\langle Y + Y', X + X' \rangle\rangle_{\psi, \alpha} = \langle\langle Y, X \rangle\rangle_{\psi, \alpha} + \langle\langle Y, X' \rangle\rangle_{\psi, \alpha} + \langle\langle Y', X \rangle\rangle_{\psi, \alpha} + \langle\langle Y', X' \rangle\rangle_{\psi, \alpha}$,
- (ii) $\langle\langle bY, aX \rangle\rangle_{\psi, \alpha} = \bar{b}a \langle\langle Y, X \rangle\rangle_{\psi, \alpha}$

for any $X, X', Y, Y' \in L(\mathcal{H})$ and $a, b \in \mathbb{C}$, hence it defines a *sesquilinear form* on $L(\mathcal{H})$. By definition, one has

$$\langle\langle Y^*, X^* \rangle\rangle_{\psi, \alpha} = \langle\langle X, Y \rangle\rangle_{\psi, -\alpha}. \quad (7.17)$$

By observing moreover that

$$\langle\langle Y, X \rangle\rangle_{\psi, \alpha}^* = \langle\langle X, Y \rangle\rangle_{\psi, \alpha^*}, \quad X, Y \in L(\mathcal{H}) \quad (7.18)$$

holds, the sesquilinear form (7.16) is *symmetric* (Hermitian) if the given parameter $\alpha \in \mathbb{R}$ is real. If one takes $X = Y$, this reads

$$\langle\langle X, X \rangle\rangle_{\psi, \alpha} = \frac{1 + \alpha}{2} \cdot \frac{\|X\psi\|^2}{\|\psi\|^2} + \frac{1 - \alpha}{2} \cdot \frac{\|X^*\psi\|^2}{\|\psi\|^2}, \quad X \in L(\mathcal{H}). \quad (7.19)$$

Specifically for the choice $-1 \leq \alpha \leq 1$ of the parameter, note that the above quantity happens to be always *positive*, hence becomes a *semi-norm*, for which we introduce the notation

$$\|X\|_{\psi, \alpha} := \langle\langle X, X \rangle\rangle_{\psi, \alpha}^{\frac{1}{2}}, \quad X \in L(\mathcal{H}), \quad -1 \leq \alpha \leq 1. \quad (7.20)$$

Note that

$$\|X\|_{\psi, \alpha}^2 \leq \frac{1 + \alpha}{2} \cdot \|X\|^2 + \frac{1 - \alpha}{2} \cdot \|X^*\|^2 = \|X\|^2, \quad (7.21)$$

where $\|X\|$ denotes the usual operator norm of $X \in L(\mathcal{H})$, which shows that the semi-norm $\|\cdot\|_{\psi, \alpha}$ induces a topology on $L(\mathcal{H})$ coarser than that induced by the usual operator norm $\|\cdot\|$.

Statistical Interpretation. In classical theory, the natural geometry (7.10) introduced on the space of observables admits statistical interpretation as correlations by means of probability measures. Parallel to this, we next intend to provide a statistical representation of the sesquilinear form (7.16) for the quantum case, and this is achieved by means of QJP distributions. As mentioned earlier, we let $\mu_{\text{add}}^{\psi, \alpha}$ denote a quasi-probability measure satisfying

$$(\mathcal{F}\mu_{\text{add}}^{\psi, \alpha})(s, t) = \frac{1 + \alpha}{2} \cdot \frac{\langle \psi, e^{-itB} e^{-isA} \psi \rangle}{\|\psi\|^2} + \frac{1 - \alpha}{2} \cdot \frac{\langle \psi, e^{-isA} e^{-itB} \psi \rangle}{\|\psi\|^2}, \quad \alpha \in \mathbb{C}, \quad (7.22)$$

which are namely members of the additive complex-parametrised sub-family of QJP distributions of A and B introduced in Section 6.3.1. One readily sees from a direct application of Lemma 6.1 that the integration of polynomial functions reduces to

$$\begin{aligned} \int_{\mathbb{R}^2} b^m a^n d\mu_{\text{add}}^{\psi, \alpha}(a, b) &= (i\partial_t)^m (i\partial_s)^n (\mathcal{F}\mu_{\text{add}}^{\psi, \alpha})(s, t) \Big|_{(s, t) = (0, 0)} \\ &= \langle\langle B^m, A^n \rangle\rangle_{\psi, \alpha}. \end{aligned} \quad (7.23)$$

One may also readily obtain a generalisation of this observation to continuous functions. Indeed, according to the Stone-Weierstraß approximation theorem, since continuous functions f, g defined on a compact space admit uniform approximations by polynomial functions as

$$f(a) = \sum_{n=0}^{\infty} f_n a^n, \quad g(b) = \sum_{n=0}^{\infty} g_n b^n, \quad (7.24)$$

one has

$$\begin{aligned} \int_{\mathbb{R}^2} g^*(b) f(a) d\mu_{\text{add}}^{\psi, \alpha}(a, b) &= \sum_{n, m=0}^{\infty} g_m^* f_n \int_{\mathbb{R}^2} b^m a^n d\mu_{\text{add}}^{\psi, \alpha}(a, b) \\ &= \sum_{n, m=0}^{\infty} g_m^* f_n \langle\langle B^m, A^n \rangle\rangle_{\psi, \alpha} \\ &= \left\langle\left\langle \sum_m g_m B^m, \sum_n f_n A^n \right\rangle\right\rangle_{\psi, \alpha} \\ &= \langle\langle g(B), f(A) \rangle\rangle_{\psi, \alpha}. \end{aligned} \quad (7.25)$$

This observation can be summarised as:

Lemma 7.3 (Statistical Representation of Sesquilinear Forms). *Let $\mu_{\text{add}}^{\psi,\alpha}$ be a member of the additive complex-parametrised sub-family of QJP distributions of the pair of observables $A, B \in L(\mathcal{H})$ defined as in (7.22). Then, for any continuous functions f and g defined on the real line \mathbb{R} , the equality*

$$\langle\langle g(B), f(A) \rangle\rangle_{\psi,\alpha} = \int_{\mathbb{R}^2} g^*(b) f(a) d\mu_{\text{add}}^{\psi,\alpha}(a, b) \quad (7.26)$$

holds.

We have thus obtained a convenient representation of the sesquilinear form by integration with respect to QJP distributions, which offers a ‘statistical’ interpretation to the geometry as ‘correlations’ of a pair of generally non-commuting quantum observables.

Topic: Quasi-covariances (Quantum Covariances). As a natural extension to the classical notion of covariances, we may introduce the term ‘quasi-covariance’ (or ‘quantum-covariance’) of a pair of quantum observables under a given QJP distribution $p_{A,B}^{\psi}(a, b)$ for the quantity formally defined by

$$\int_{\mathbb{R}^2} (a - \mathbb{E}[A; \psi])(b - \mathbb{E}[B; \psi]) p_{A,B}^{\psi}(a, b) dm_2(a, b), \quad (7.27)$$

whenever the integration exists. Specifically, from an immediate application of the above Lemma, one may readily compute the quantum-covariances with respect to the additive subgroup $\mu_{\text{add}}^{\psi,\alpha}(a, b)$ of the QJP distributions as

$$\begin{aligned} \mathbb{C}\mathbb{V}^{\alpha}[A, B; \psi] &:= \int_{\mathbb{R}^2} (a - \mathbb{E}[A; \psi])(b - \mathbb{E}[B; \psi]) d\mu_{\text{add}}^{\psi,\alpha}(a, b) \\ &= \langle\langle A - \mathbb{E}[A; \psi], B - \mathbb{E}[B; \psi] \rangle\rangle_{\psi,\alpha} \\ &= \mathbb{C}\mathbb{V}_{\text{S}}[A, B; \psi] + \alpha i \mathbb{C}\mathbb{V}_{\text{A}}[A, B; \psi], \end{aligned} \quad (7.28)$$

where $\mathbb{C}\mathbb{V}_{\text{S}}[A, B; \psi]$ and $\mathbb{C}\mathbb{V}_{\text{A}}[A, B; \psi]$ are respectively the symmetric and anti-symmetric quantum covariances introduced in (4.51).

7.2.3. The Hilbert Space of Bounded Operators given a fixed State. In classical theory, observables were described by functions, whereas in quantum theory, observables become self-adjoint operators. In order to conduct an analogue of the L^2 -theory for quantum observables, it reveals for our purpose that it is convenient not just to deal with self-adjoint operators, but rather to consider the collection $L(\mathcal{H})$ of all bounded linear operators defined on the Hilbert space \mathcal{H} .

Identification. In classical theory, recall that we made identification of observables that cannot be distinguished in view of the probability measure μ by introducing the equivalence relation $f \sim g \Leftrightarrow f = g$ μ -a.e., and slimmed down the space of functions into quotient spaces (see Section 2.1.1). We intend to follow the same path for the quantum case, and to this, we

introduce the subspace

$$Z_\psi(H) := \{X \in L(H) : X|\psi\rangle = X^*|\psi\rangle = 0\} \quad (7.29)$$

and define the \mathbb{C} -linear quotient space

$$L_\psi(\mathcal{H}) := L(H)/Z_\psi(H) \quad (7.30)$$

by identifying those operators for which the action of both themselves and their adjoints are indistinguishable on the state $|\psi\rangle$. In other words, this is to say that we identify two operators $X, Y \in L(\mathcal{H})$ by the equivalence relation

$$X \sim Y \Leftrightarrow X|\psi\rangle = Y|\psi\rangle \text{ and } X^*|\psi\rangle = Y^*|\psi\rangle. \quad (7.31)$$

One readily sees that the sesquilinear form (7.16) passes to the quotient, and we thus obtain a sesquilinear form

$$\langle\langle [Y]_\psi, [X]_\psi \rangle\rangle_{\psi, \alpha} := \langle\langle Y, X \rangle\rangle_{\psi, \alpha}, \quad [X]_\psi, [Y]_\psi \in L_\psi(\mathcal{H}) \quad (7.32)$$

on the quotient space $L_\psi(\mathcal{H})$. Whenever there is no risk of confusion, we shall mostly denote equivalence classes by their representatives for simplicity of notation. Note also that the *involution* $*$: $L(\mathcal{H}) \rightarrow L(\mathcal{H})$, $X \mapsto X^*$ that takes a bounded linear operator to its adjoint is also well-defined on the quotient space.

Hilbert Space of Operators. We have already seen that the original sesquilinear form (7.16) becomes positive and symmetric for the choice $-1 \leq \alpha \leq 1$ of the parameter. Based on the identification above, the sesquilinear form (7.32) on the quotient space $L_\psi(\mathcal{H})$ becomes *positive definite*, which is to say that

- (i) $\langle\langle X, X \rangle\rangle_{\psi, \alpha} \geq 0, \quad X \in L_\psi(\mathcal{H}),$
- (ii) $\langle\langle X, X \rangle\rangle_{\psi, \alpha} = 0 \Leftrightarrow X = 0$

for the choice $-1 \leq \alpha \leq 1$. This makes (7.32) an *inner product* on $L_\psi(\mathcal{H})$ for $-1 \leq \alpha \leq 1$, allowing us to define the norm

$$\|X\|_{\psi, \alpha} := \langle\langle X, X \rangle\rangle_{\psi, \alpha}^{\frac{1}{2}}, \quad X \in L_\psi(\mathcal{H}), \quad -1 \leq \alpha \leq 1. \quad (7.33)$$

One moreover proves by rudimentary technique that the space is in fact complete with respect to the norm. We thus have the following result.

Proposition 7.4 (Hilbert Space of Operators). *For a fixed $|\psi\rangle \in \mathcal{H}$ and the choice $-1 \leq \alpha \leq 1$ of the parameter, the ordered pair $(L_\psi(\mathcal{H}), \langle\langle \cdot, \cdot \rangle\rangle_{\psi, \alpha})$ defines a complex Hilbert Space.*

This convenient property greatly facilitates our further argument. Hence, in what follows, we will be treating only those geometries associated with the specific choice $-1 \leq \alpha \leq 1$ of the complex parameter.

Sub-algebra generated by an Observable. We next introduce an important subspace of $L(\mathcal{H})$. Given a bounded self-adjoint operator $A \in L(\mathcal{H})$, we prepare a special symbol

$$\mathfrak{E}(A) := \{f(A) : f \text{ is a continuous function on } \sigma(A)\} \quad (7.34)$$

for the \mathbb{C} -linear subspace of $L(\mathcal{H})$ consisting of all operators defined by means of the functional calculus (3.56). By definition, one proves that

$$\|f(A)\| = \|f\|_\infty, \quad (7.35)$$

where the l. h. s. is the operator norm of $f(A)$, and the r. h. s. is the supremum norm of the continuous function f (note that the spectrum $\sigma(A)$ of a bounded self-adjoint operator A is compact, hence any continuous function defined on the spectrum is necessarily bounded). Moreover, it is easy to see that $f(A)^* = f^*(A)$ holds, where the l. h. s. denotes the adjoint of the linear operator $f(A)$, whereas the r. h. s. denotes the operator induced by the complex conjugate f^* of the original function f . This implies that all the operators of the form $f(A)$ are normal²⁸, and that the space $\mathfrak{E}(A)$ is closed under the operation of taking the adjoint. Moreover, one sees that any two operators $f(A), g(A) \in \mathfrak{E}(A)$ commute with each other $f(A)g(A) = (fg)(A) = g(A)f(A)$, and that the space $\mathfrak{E}(A)$ thus makes itself into a commutative C^* -algebra. We call the space $\mathfrak{E}(A)$ the sub-algebra *generated by* A .

Identification. An immediate observation one makes is that the sesquilinear form (7.16) is independent of the choice of the parameter $\alpha \in \mathbb{C}$ on the space $\mathfrak{E}(A)$. Indeed, for any choice of a pair of continuous functions f, g defined on the spectrum $\sigma(A)$, the equality

$$\begin{aligned} \langle\langle g(A), f(A) \rangle\rangle_{\psi, \alpha} &= \frac{\langle g(A)\psi, f(A)\psi \rangle}{\|\psi\|^2} \\ &= \int_{\mathbb{R}} g^*(a)f(a) d\mu_A^\psi(a) = \langle g, f \rangle_{\mu_A^\psi} \end{aligned} \quad (7.36)$$

holds, where the right-most hand side denotes the standard inner product introduced on the space of square-integrable complex functions. Following the same line of arguments we have made in the previous discussion of this section, we next intend to identify those operators that are not distinguishable in view of a given state $|\psi\rangle \in \mathcal{H}$. To this, we introduce the subspace

$$Z_\psi(A) := \{N \in \mathfrak{E}(A) : N|\psi\rangle = N^*|\psi\rangle = 0\}, \quad (7.37)$$

and define the space

$$\mathfrak{E}_\psi(A) := \overline{\mathfrak{E}(A)/Z_\psi(A)} \quad (7.38)$$

by identifying those normal operators for which the action of both themselves and their adjoints on the state $|\psi\rangle$ are indistinguishable. Here, the overline on the quotient space $\mathfrak{E}(A)/Z_\psi(A) \subset L_\psi(\mathcal{H})$ denotes its topological closure with respect to the topology on the superset $L_\psi(\mathcal{H})$ induced by the norm $\|\cdot\|_{\psi, \alpha}$ (7.33). Note here that, as a set, the closure $\mathfrak{E}_\psi(A)$ is independent of the choice of the parameter $-1 \leq \alpha \leq 1$, since all the norms $\|\cdot\|_{\psi, \alpha}$ coincide on the subspace $\mathfrak{E}(A)/Z_\psi(A)$. Moreover, one may readily check that all the inner

²⁸ Recall that a bounded operator $N \in L(\mathcal{H})$ is *normal* if and only if $\|N\psi\| = \|N^*\psi\|$ holds for all $|\psi\rangle \in \mathcal{H}$.

products $\langle\langle \cdot, \cdot \rangle\rangle_{\psi, \alpha}$ also coincide for the pair of elements of the closure $\mathfrak{E}_\psi(A)$ for any choice of the parameter $-1 \leq \alpha \leq 1$.

Now, since by definition the space $\mathfrak{E}_\psi(A)$ is a closed subspace of the complex Hilbert space, it is itself a complex Hilbert space. By denoting the restriction of the inner product as $\langle\langle \cdot, \cdot \rangle\rangle_\psi := \langle\langle \cdot, \cdot \rangle\rangle_{\psi, \alpha}|_{\mathfrak{E}_\psi(A)}$, which does not depend on the choice of the parameter as we have mentioned above, we have:

Lemma 7.5. *For a fixed $|\psi\rangle \in \mathcal{H}$, the ordered pair $(\mathfrak{E}_\psi(A), \langle\langle \cdot, \cdot \rangle\rangle_\psi)$ defines a complex Hilbert space.*

The next Lemma is of special interest for our purpose.

Lemma 7.6. *Let $A \in L(\mathcal{H})$ be self-adjoint, $|\psi\rangle \in \mathcal{H}$, and let $\mathfrak{E}_\psi(A)$ be defined as in (7.38). Then, there exists a unique unitary operator $\Phi : L^2(\mu_A^\psi) \mapsto \mathfrak{E}_\psi(A)$ such that*

$$\Phi(f) = [f(A)]_\psi \quad (7.39)$$

holds for every continuous function f on $\sigma(A)$, where the r. h. s. denotes the equivalent class of $f(A)$, which in turn is a bounded linear operator defined by means of the functional calculus (3.56).

Proof. We will construct the map Φ by continuous linear extension. To this, first recall that, since $\sigma(A)$ is compact, the space $C(\sigma(A))$ of all continuous functions on $\sigma(A)$ is dense in $L^2(\mu_A^\psi)$. Since the map

$$\tilde{\Phi} : L^2(\mu_A^\psi) \supset C(\sigma(A)) \rightarrow \mathfrak{E}_\psi(A), \quad f \mapsto f(A), \quad (7.40)$$

is an isometry $\|\tilde{\Phi}(f)\|_\psi := \|f(A)\psi\| = \|f\|_2$ from a dense subspace of a normed space to a Banach space, there exists a unique isometric extension $\Phi : L^2(\mu_A^\psi) \rightarrow \mathfrak{E}_\psi(A)$. By construction, one may also prove the surjectivity of Φ , hence Φ is unitary. ■

This is to say that the Hilbert spaces $\mathfrak{E}_\psi(A) \cong L^2(\mu_A^\psi)$ are unitarily isomorphic, and that Φ gives an embedding of the space of square-integrable functions $L^2(\mu_A^\psi)$ into the space of bounded operators on a Hilbert space. For simplicity of notation, we occasionally denote the image of a square-integrable function $f \in L^2(\mu_A^\psi)$ by $f(A) := \Phi(f)$. A word of caution is to be made here for the possible confusion for the notation used. Here, the notation $f(A) := \Phi(f)$ is meant to denote (a representative of) the equivalence class of bounded linear operators, whereas the notation $f(A)$ is usually used to represent (generally unbounded) linear operator defined by means of the functional calculus (3.56). The relation between the two different notations can be understood in the following way. For $f \in L^2(\mu_A^\psi)$, let $T \in \Phi(f)$ be a representative of the equivalence class of bounded operators, and let $f(A)$ be a (generally unbounded) operator defined by means of the functional calculus (3.56), and note that $|\psi\rangle \in \text{dom}(f(A))$ by definition. Then, we have $T|\psi\rangle = f(A)|\psi\rangle$.

7.3. Interpretation of Conditional Quasi-expectations

Now that we have sufficiently prepared our tools, the most important among which is the embedding

$$\Phi : L^2(\mu_A^\psi) \cong \mathfrak{E}_\psi(A) \subset L_\psi(\mathcal{H}) \quad (7.41)$$

of the L^2 -space of functions into that of bounded linear operators on a Hilbert space, we next focus on orthogonal projections and ‘conditioning’ with respect to QJP distributions.

7.3.1. *Geometric Interpretation of Conditional Quasi-expectations.* Recall that, with each closed subspace of a Hilbert space, a unique *orthogonal projection* is associated. In what follows, we are interested in the orthogonal projection of an observable A onto the subspace $\mathfrak{E}_\psi(B)$ generated by another observable B , and see that this provides a geometric interpretation of the conditional quasi-expectation $\mathbb{E}^\alpha[A|B; \psi]$ introduced earlier in (4.77).

To this, let $B \in L(\mathcal{H})$ be self-adjoint, $|\psi\rangle \in \mathcal{H}$, $-1 \leq \alpha \leq 1$, and let $\mathfrak{E}_\psi(B)$ be the space generated by B , which is a closed subspace of the Hilbert spaces $(L_\psi(\mathcal{H}), \langle \cdot, \cdot \rangle_{\psi, \alpha})$. As a closed subspace of a Hilbert space, there exists a unique orthogonal projection

$$P_\alpha(\cdot | B; \psi) : L_\psi(\mathcal{H}) \rightarrow \mathfrak{E}_\psi(B), \quad X \mapsto P_\alpha(X|B; \psi) \quad (7.42)$$

associated to $\mathfrak{E}_\psi(B)$. Recalling the relation between orthogonal projections and conditional expectations in classical probability theory (see Section 7.1), it is natural to conjecture that an analogous relation holds for the quantum case. To this, let $A \in L(\mathcal{H})$ be self-adjoint, and consider the projection $P_\alpha(A|B; \psi)$ of A onto $\mathfrak{E}_\psi(B)$. We have seen in Section 4 that the conditional quasi-expectations $\mathbb{E}^\alpha[A|B; \psi] \in L^1(\mu_B^\psi)$ introduced in (4.77) serve as possible candidates of quantum analogues of conditional expectations that can even be defined for non-commuting pair of quantum observables. Since, the observable A we consider here is bounded, one may prove that the conditional quasi-expectation $\mathbb{E}^\alpha[A|B; \psi] \in L^2(\mu_B^\psi)$ is in fact square-integrable. By letting $\mathbb{E}^\alpha[A|B; \psi]$ denote both the square-integrable function and its image by the unitary map $\Phi : L^2(\mu_B^\psi) \rightarrow \mathfrak{E}_\psi(B)$, it is natural to conjecture the validity of the equality $P_\alpha(A|B; \psi) = \mathbb{E}^\alpha[A|B; \psi]$, where the l. h. s. is the orthogonal projection of A onto the space $\mathfrak{E}_\psi(B)$ with respect to the (parameter-dependent) inner product $\langle \cdot, \cdot \rangle_{\psi, \alpha}$, whereas the r. h. s. denotes the image of the (parameter-dependent) conditional quasi-expectation of A given B by the unitary map Φ .

Proposition 7.7 (Orthogonal Projection and Conditional Quasi-Expectation). *Let $A, B \in L(\mathcal{H})$ be self-adjoint, $|\psi\rangle \in \mathcal{H}$ and $-1 \leq \alpha \leq 1$. Then, the conditional quasi-expectation $\mathbb{E}^\alpha[A|B; \psi]$ introduced in (4.77) is μ_B^ψ -square-integrable, which could thus be identified with the equivalence class of bounded operators*

$$\mathbb{E}^\alpha[A|B; \psi] := \Phi(\mathbb{E}^\alpha[A|B; \psi]) \quad (7.43)$$

by means of the embedding $\Phi : L^2(\mu_B^\psi) \rightarrow \mathfrak{E}_\psi(B)$ defined in (7.41). Then, the orthogonal projection of A onto the subspace $\mathfrak{E}_\psi(B)$ generated by B , defined in (7.42), reads

$$P_\alpha(A|B; \psi) = \mathbb{E}^\alpha[A|B; \psi], \quad (7.44)$$

which is to say that orthogonal projections are equivalent to conditional quasi-expectations.

Proof. Let $f \in L^2(\mu_B^\psi)$, and let $f(B) := \Phi(f)$ denote the embedding of f into the space $\mathfrak{E}_\psi(B)$. By definition of orthogonal projections, one readily finds

$$\langle f(B), A \rangle_{\psi, \alpha} = \langle f(B), P_\alpha(A|B; \psi) \rangle_{\psi, \alpha}. \quad (7.45)$$

On the other hand, observe that

$$\begin{aligned}
\langle f, \mathbb{E}^\alpha[A|B; \psi] \rangle_{\mu_B^\psi} &:= \int_{\mathbb{R}} f^*(b) \cdot \mathbb{E}^\alpha[A|B = b; \psi] d\mu_B^\psi(b) \\
&= \frac{1 + \alpha}{2} \int_{\mathbb{R}} f^*(b) d \frac{\langle \psi, E_B(b)A\psi \rangle}{\|\psi\|^2} + \frac{1 - \alpha}{2} \int_{\mathbb{R}} f^*(b) d \frac{\langle \psi, AE_B(b)\psi \rangle}{\|\psi\|^2} \\
&= \frac{1 + \alpha}{2} \cdot \frac{\langle f(B)\psi, A\psi \rangle}{\|\psi\|^2} + \frac{1 - \alpha}{2} \cdot \frac{\langle A^*\psi, f(B)^*\psi \rangle}{\|\psi\|^2} \\
&=: \langle\langle f(B), A \rangle\rangle_{\psi, \alpha}
\end{aligned} \tag{7.46}$$

Combining this with the unitarity of the embedding $\Phi : L^2(\mu_B^\psi) \rightarrow \mathfrak{E}_\psi(B)$, one thus has

$$\langle\langle f(B), A \rangle\rangle_{\psi, \alpha} = \langle\langle f(B), \mathbb{E}^\alpha[A|B; \psi] \rangle\rangle_\psi. \tag{7.47}$$

The positive-definiteness of the inner product applied to the two results (7.45) and (7.47) proves our statement. \blacksquare

Just as we have seen for the classical case, this result provides a geometric interpretation of conditional quasi-expectations as orthogonal projections. As a corollary, one has a geometric interpretation of Aharonov's weak value.

Corollary 7.8 (Geometric Interpretation of Aharonov's Weak Value). *Under the same conditions, let*

$$A_w(B) := \Phi(A_w) \tag{7.48}$$

denote the embedding of the Aharonov's weak value $A_w := \mathbb{E}^1[A|B; \psi]$ introduced in (4.81). Then, the weak value

$$A_w(B) = P_1(A|B; \psi) = \mathbb{E}^1[A|B; \psi] \tag{7.49}$$

could be interpreted as the orthogonal projection of A onto the subspace $\mathfrak{E}_\psi(B)$ generated by B .

Topic: Weak Value as Optimal Approximation. As a direct consequence of Proposition 7.8 (specifically, Corollary 7.8), we note an interesting result regarding conditional quasi-expectations (specifically, the weak value) and optimal approximation. As orthogonal projections, observe that conditional quasi-expectations furnish the optimal proxy function for A minimising the distance

$$\|A - \mathbb{E}^\alpha[A|B; \psi]\|_{\psi, \alpha} = \min_f \|A - f(B)\|_{\psi, \alpha} \tag{7.50}$$

from an observable A to the space of normal observables $\mathfrak{E}_\psi(B)$ generated by another observable B . An equivalent expression to this is the equality

$$\|A - f(B)\|_{\psi, \alpha} = \|A - \mathbb{E}^\alpha[A|B; \psi]\|_{\psi, \alpha} + \|\mathbb{E}^\alpha[A|B; \psi] - f(B)\|_{\psi, \alpha}, \tag{7.51}$$

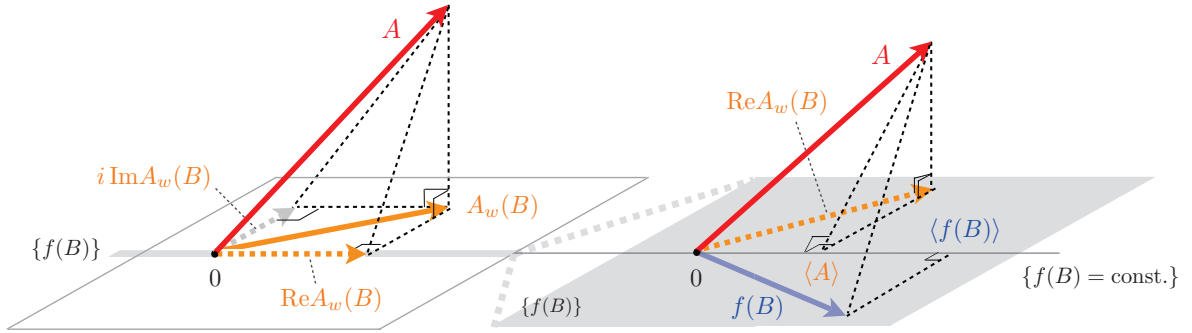


Fig. 2 Geometric relations among the operators involved for the choice $\alpha = 1$. The left illustrates how the operator A is projected onto the subspace of normal operators $\mathfrak{E}_\psi(B)$ generated by B , with the center line representing the space of self-adjoint operators $\{f(B)\}$. The right elaborates the projection onto the space $\{f(B)\}$, where now the center line represents the space of constant functions $\{f(B) = \text{const.}\}$ (more precisely, functions proportional to the identity operator I) including $f(B) = \langle A \rangle$.

which is nothing but the ‘Pythagorean identity’ valid for orthogonal projections²⁹ in Hilbert spaces. The interpretation of conditional quasi-expectations as orthogonal projections provide the core geometric observations why the weak value appears as the optimal choice for the proxy functions in the novel uncertainty relations for approximation/estimation [14].

7.3.2. Statistical Interpretation of Conditional Quasi-expectations. In classical probability theory, conditional expectations not only admitted geometric interpretation as orthogonal projections, but also statistical interpretation as conditioned averages. We next seek to provide a quantum analogue of this observation, namely, to provide a statistical interpretation of the conditional quasi-expectations as ‘conditional averages’ with respect to QJP distributions. To this, we first introduce a general term:

Definition (Conditional Quasi-expectation of Quantum Observables). *Let $\mu \in \mathfrak{M}_{A,B}^\psi$ be a QJP distribution of a pair of quantum observables A and B on the state $|\psi\rangle \in \mathcal{H}$, such that it admits representation by quasi-probability measures, and suppose that the expectation value $\mathbb{E}[\pi_A; \mu]$ exists. Denoting the measurable functions representing the behaviour of the measurement outcomes of A and B by $\pi_A(a, b) = a$ and $\pi_B(a, b) = b$, respectively, we then*

²⁹ Specifically, observing that $\mathbb{E}^\alpha[A|B; \psi] = \text{Re}A_w(B) + i\alpha\text{Im}A_w(B)$, we have $\mathbb{E}^\alpha[A|B; \psi] = \text{Re}A_w(B)$ for the choice $\alpha = 0$. This gives

$$\begin{aligned}
\|(A - \text{Re}A_w(B))\psi\| &= \|A - \text{Re}A_w(B)\|_{\psi,0} \\
&\leq \|A - f(B)\|_{\psi,0} \\
&= \|(A - f(B))\psi\|, \quad f \in L^2(\mu_B^\psi), \quad f \text{ is real,}
\end{aligned} \tag{7.52}$$

as a special case. This specific form is known by [45, 46], although proven from a different perspective than directly utilising the geometric observation made in this paper.

define the conditional quasi-expectation of A given B under the QJP distribution μ by

$$\mathbb{E}[A|B; \mu] := \mathbb{E}[\pi_A|\pi_B; \mu], \quad (7.53)$$

where the definition of the r. h. s. is given in Section 5.2.1, whenever the Radon-Nikodým derivatives concerned exist.

We next see that the definition of conditional quasi-expectations agree with those introduced earlier (4.77).

Proposition 7.9 (Statistical Interpretation of Conditional Quasi-expectations). *Let $A, B \in L(\mathcal{H})$ be self-adjoint, $|\psi\rangle \in \mathcal{H}$, and let $\mu_{\text{add}}^{\psi, \alpha}$ be a member of the additive complex-parametrised sub-family of the QJP distributions of A and B for the choice $\alpha \in \mathbb{C}$ defined as in (7.22). Then, the conditional quasi-expectation $\mathbb{E}[A|B; \mu_{\text{add}}^{\psi, \alpha}]$ of A given B under $\mu_{\text{add}}^{\psi, \alpha}$ is well-defined, which reads*

$$\mathbb{E}^\alpha[A|B; \psi] = \mathbb{E}[A|B; \mu_{\text{add}}^{\psi, \alpha}], \quad (7.54)$$

where the l. h. s. is the α -parametrised conditional quasi-expectation introduced earlier in (4.77).

Proof. We first demonstrate the well-definedness of $\mathbb{E}[A|B; \mu_{\text{add}}^{\psi, \alpha}]$, and to this, let

$$\begin{aligned} \nu(\Delta) &:= \int_{\pi_B^{-1}(\Delta)} \pi_A(a, b) d\mu_{\text{add}}^{\psi, \alpha}(a, b) \\ &= \int_{\mathbb{R} \times \Delta} a d\mu_{\text{add}}^{\psi, \alpha}(a, b) \\ &= \frac{1 + \alpha}{2} \cdot \frac{\langle \psi, E_B(\Delta)A\psi \rangle}{\|\psi\|^2} + \frac{1 - \alpha}{2} \cdot \frac{\langle \psi, AE_B(\Delta)\psi \rangle}{\|\psi\|^2}. \end{aligned} \quad (7.55)$$

Since $\nu \ll \mu_B^\psi$, the Radon-Nikodým derivative $\mathbb{E}[A|B; \mu_{\text{add}}^{\psi, \alpha}] := d\nu/d\mu_B^\psi$ exists. The validity of the equality $\mathbb{E}^\alpha[A|B; \psi] = \mathbb{E}[A|B; \mu_{\text{add}}^{\psi, \alpha}]$ is immediate by definition. \blacksquare

This result provides a statistical interpretation of conditional quasi-expectations as ‘conditional averages’ of an observable A given another observable B with respect to the QJP distributions concerned. As a corollary, one also has a statistical interpretation of Aharonov’s weak value.

Corollary 7.10 (Statistical Interpretation of Aharonov’s Weak Value). *Under the same conditions, let $A_w := \mathbb{E}^1[A|B; \psi]$ denote the Aharonov’s weak value introduced in (4.81). Then, the weak value*

$$A_w(B) = \mathbb{E}[A|B; \mu_{\text{add}}^{\psi, 1}] = \mathbb{E}^1[A|B; \psi] \quad (7.56)$$

admits interpretation as the ‘conditional averages’ of an observable A given another observable B with respect to the additive subfamily of QJP distributions for the choice $\alpha = 1$.

8. Summary and Discussion

Now we present a recap of our results obtained in this paper before going over to our discussions.

8.1. Summary

The underlying motivation for our study was to obtain a coherent understanding to the formalism of quasi-joint-probabilities (QJP) of quantum observables, and to find the interpretation of Aharonov's weak value within this framework. The main body, starting from Section 2 to 7, was devoted to the discussion of three logical groups of mutually interrelated topics, namely (i) an heuristic construction of QJP distributions (Section 2 to 5), (ii) formal definition of QJP distributions (Section 6), and (iii) its application to the interpretation of the weak value (Section 7). Each of these sections will be summarised concisely below.

8.1.1. QJP: Heuristic Construction. Four sections starting from Section 2 to Section 5 were devoted to some careful analyses on the quantum measurement models that we called the unconditioned and the conditioned measurement (UM and CM) schemes. By inspecting each of the measurement models in terms of statistical averages and raw probability distributions, we confirmed that one may obtain the desired information of the target system by either looking into the strong or weak regions of the intensity of the interaction parameter. Specifically, we saw that the study on the CM scheme naturally lead us to the concept of quasi-joint-probability (QJP) distributions of generally non-commuting pair of observables.

Section 2 (UM I). Section 2 was devoted to a review on the UM scheme from a standard operator-centric approach. The quantity of interest was the statistical average of the meter observable after the interaction, from which we confirmed the well known fact that the information of both the target observable A and the target state $|\phi\rangle$ can be retrieved in the form of the expectation value $\mathbb{E}[A; \phi]$ of A . The expectation value $\mathbb{E}[A; \phi]$ was shown to be obtained from the measurement outcome of the meter observable, either by probing the strong limit $g \rightarrow \pm\infty$ or the weak limit $g \rightarrow 0$ of the interaction parameter.

Section 3 (UM II). In Section 3, we took a closer look at the UM scheme in the level of probabilities, where the quantity of interest was now not just the statistical average but the 'raw' probability measure describing the probabilistic behaviour of the measurement outcomes of the meter observable. We saw that the outcome of the meter observable after the interaction was given by a convolution of both the initial profiles of the target and the meter observables. As for the retrieval of the target information, we found that, in a parallel manner to the previous section, the full profile of the target observable can be reclaimed by either probing the strong or the weak limit of the interaction.

Section 4 (CM I). In Section 4, we conducted an analysis of the conditioned measurement scheme in the operator level, where the quantity of interest became the *conditional expectation* of the meter observable under another given conditioning observable B of the target system. Some relevant topics, including a review and comments on the recent theoretical analyses on the alleged technical advantages of employing conditioning for precision measurements were presented, along with a measure theoretic approach to the possible limit for

‘amplification’ by conditioning, and a systematical method (with an example) to analytically evaluate the conditional expectation in the case where A has a spectrum consisting of finite points. As for the retrieval of the target information, we exclusively studied the behaviour of the meter outcome in the weak region of the interaction parameter, and observed that the obtained value can be understood as a quantum analogue of conditional expectations, which we termed *conditional quasi-expectations*, of the target observable A given the conditioning observable B , to which Aharonov’s weak value belongs as a special case. It was also revealed that there exists some qualitative difference on the properties between the classical conditional expectations and the quantum analogue discussed here.

Section 5 (CM II). In Section 5, the study of the conditioned measurement scheme was given a probabilistic approach, where the quantity of interest now became the *QJP distribution* of a pair of canonically conjugate observables on the meter system, conditioned by the outcome of the conditioning observable B of the target system. For definiteness, this was accomplished in view of the Wigner-Ville distribution, which was primarily chosen as a convenient realisation among the various candidates of the quasi-probability distributions of the canonically conjugate pair that may be naturally associated with the quantum state of the meter system. It was then argued that, in parallel to the UM case, one can recover the information of the target system by examining either the strong or the weak region of the interaction parameter, and that the information obtained can be understood as a quantum analogue of conditional probabilities, which we termed *conditional quasi-probabilities*, of the target observable A given the conditioning observable B on the initial state $|\phi\rangle$. We then found that the conditional quasi-probability shares similar properties with the classical counterpart, while it admits complex values unlike the classical one. We subsequently confirmed that, given B , the ‘statistical average’ of the conditional quasi-probability of A coincides with the conditional quasi-expectation of A obtained in the preceding section. This is precisely the same as the relation between classical conditional probabilities and conditional expectations.

8.1.2. QJP: Formal Definition. Inspired by the heuristic arguments from the bottom-up and operational analyses given in the preceding four sections, in Section 6 we provided the top-down discussion on QJP distributions defined for arbitrary pairs of generally non-commutative quantum observables.

Section 6 (QJP of Quantum Observables). Based on the results of the spectral theorem for self-adjoint/normal operators on Hilbert spaces and their Fourier transforms, we proposed a general prescription for defining distributions describing the ‘joint behaviour’ of a pair of generally non-commuting quantum observables, which serves as a natural generalisation to that defined for a pair of simultaneously measurable observables. We then observed that the QJP defined this way for a non-commutative pair of observables admits arbitrariness, that is, there exists a multitude of candidates that all share in common certain desirable properties to be qualified as QJP. We subsequently concentrated on a special sub-family of the class of QJP distributions parametrised by a single complex number for the ease of further discussions, such that it includes both the Wigner-Ville type and the Kirkwood-Dirac type of QJP distributions which are among the most familiar examples considered in the

literature. We then summarised our results obtained up to Section 5 from a relatively aerial viewpoint gained here, and discussed where the heuristic arguments and observations in the foregoing sections find their places in this broader framework.

8.1.3. Application. As the final topic, we gave an example of application of our observations on QJP distributions of quantum observables.

Section 7 (Application: Interpretation of the Weak Value). To discuss where the mathematical observations on QJP distributions may find their use, we studied on the quantum analogue of ‘correlations’ (inner products) that can be defined even for a pair of non-commuting observables. As is well known, due to the non-commutative nature of quantum observables, there is no unique way to introduce a ‘natural inner product’ on the space of quantum observables. We showed that the ambiguity of the possible geometries that can be introduced on the space corresponds precisely to the ambiguity of the definition of QJP distributions, and that the QJP distributions provides a convenient representation of the geometries in terms of ‘integration’ (statistics). We then concentrated on a special sub-family of all possible QJP distributions parametrised by a single complex number and observed that the geometric concept of *orthogonal projection* may be endowed with a statistical interpretation as *conditioning*. This fact is analogous to the classical case, while the difference lying in the fact that, for the quantum case, there could be multiple orthogonal projections due to the non-uniqueness of the inner product. The main finding is that, Aharonov’s weak value may be understood as a special realisation of the possible orthogonal projections of a quantum observable A onto the space of all normal operators generated by another observable B , and at the same time, as a conditioning of A when the outcomes of B is given. The former is a geometric interpretation of the weak value, while the latter is its statistical interpretation, but since QJP distributions tie them together, both interpretations are equivalent.

8.2. Discussion

Since the advent of quantum theory founded nearly a century ago, non-commutativity of quantum observables has undoubtedly been one of the major sources of troubles we face when we try to interpret their measurement outcomes in a sensible manner. This has naturally led to various attempts of ‘quasi-classical’ interpretation of quantum observables in terms of commuting quantities familiar to us in classical theory. Wigner, Weyl and Moyal were among the prominent figures who have made much contribution in this effort, bearing most notably the theory of Wigner-Weyl transform [47] and Weyl-Groenewold-Moyal product [48, 49]. In particular, the theory of Wigner-Weyl transform provides an invertible mapping between functions defined on a phase space and operators on a Hilbert space, in which the mapping from functions to operators is called the Weyl transform, whereas the inverse is called the Wigner transform. It is notable in this respect that the Wigner-Ville distributions arise as the Wigner transform of density operators, and from this follows the fact that the expectation values of quantum observables can be expressed as the statistical average by integration of their Wigner transforms with respect to the Wigner-Ville distribution defined on the phase space.

Viewed from the broader context of these quasi-classical transforms, the mathematical methods developed in this paper may be understood as another functional analytic approach

to this problem. Recall that, in functional analysis, a map that assigns a ring of functions onto a *commutative* sub-algebra of the algebra of quantum observables is known as the functional calculus, which in turn is known to be uniquely represented by a spectral measure. The family of quasi-joint-spectral distributions (QJSDs) introduced in Section 6 are non-commutative analogues of spectral measures, which induce maps that assign functions to generally *non-commutative* sets of quantum observables. Due to the possible non-commutativity of the chosen combination of observables, QJSDs are in general highly non-unique, and this leads to various candidates of quasi-classical transforms. In fact, the Wigner-Weyl transform can be understood as a special case in this framework, namely, the quasi-classical transform corresponding to the member of our complex parametrised convolutive sub-family of QJSDs mentioned in the text for the particular choice $\alpha = 0$. The method of ‘hashing’ presented in this paper thus exemplifies a procedure for constructing a broad class of candidates of quasi-classical transforms.

The method of quasi-classical transforms, to which the Wigner-Weyl transform belongs as a special case, not only offers a statistical interpretation of the behaviour of a combination of non-commuting quantum observables, but also sheds new light on the physical analysis in quantum mechanics pertaining to that process. It should be obvious that one can draw an analogy to various concepts and results in classical probability theory when one considers the quantum counterparts obtained by this method, which allows for an intuitive treatment of the latter based on the geometric structure present in the probability theory. Besides, transformation of Hilbert space operators into functions or quasi-probability distributions has its own technical merit in the mathematical analysis, since familiar results in measure and integration theory, including various convergence theorems, integral inequalities and representation theorems, are readily available.

One of the direct applications taking advantage of these properties is the geometric/statistical interpretation of the weak value discussed in Section 7. There, we found that the weak value can be regarded as one of the possible quantum analogues of conditional expectations, which are indeed fundamental quantities in quantum mechanics as much as the standard conditional expectations are in classical probability theory. This interpretation also leads to novel inequalities of uncertainty relations for approximation and estimation which are capable of treating both the position-momentum inequality and the time-energy inequality [14] within a unified framework.

Finally, we wish to note that, in any conditioned quantum measurement such as the weak measurement, non-commutative observables must be dealt with in one way or another in the context of probability theory when one tries to make sense of the measurement outcome. Given this, we expect that our method of quasi-classical transforms, which is established on a rigorous mathematical basis, may offer a fundamental and practical scheme in which issues involving measurement results of non-commuting observables are analysed.

Acknowledgment

The authors appreciate Professor A. Hosoya and S. Tanimura for helpful discussions and insightful comments. This work was supported in part by JSPS KAKENHI No. 25400423, No. 26011506, and by the Center for the Promotion of Integrated Sciences (CPIS) of SOKENDAI.

References

- [1] E. Wigner. On the Quantum Correction For Thermodynamic Equilibrium. *Phys. Rev.*, 40:749, 1932.
- [2] J. Ville. Théorie et Applications de la Notion de Signal Analytique. *Câbles et Transmission*, 2:61–74, 1948.
- [3] J. G. Kirkwood. Quantum Statistics of Almost Classical Assemblies. *Phys. Rev.*, 44:31, 1933.
- [4] P. A. M. Dirac. On the Analogy Between Classical and Quantum Mechanics. *Rev. Mod. Phys.*, 17:195–199, 1945.
- [5] Y. Aharonov, D. Z. Albert, and L. Vaidman. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Phys. Rev. Lett.*, 60:1351, 1988.
- [6] Y. Aharonov, P. G. Bergmann, and L. Lebowitz. Time Symmetry in the Quantum Process of Measurement. *Phys. Rev.*, 134:B1410, 1964.
- [7] J. S. Lundeen, B. Sutherland, A. Patel, C. Stewart, and C. Bamber. Direct measurement of the quantum wavefunction. *Nature*, 474:188–191, 2011.
- [8] T. Mori and I. Tsutsui. Weak value and the wave–particle duality. *Quantum Stud.: Math. Found.*, 2:371, 2015.
- [9] Y. Aharonov and L. Vaidman. Complete description of a quantum system at a given time. *J. Phys. A: Math. Gen.*, 24:2315, 1991.
- [10] K. Yokota, T. Yamamoto, M. Koashi, and N. Imoto. Direct observation of Hardy’s paradox by joint weak measurement with an entangled photon pair. *New Jour. Phys.*, 11:033011, 2009.
- [11] Y. Aharonov and D. Rohrlich. *Quantum Paradoxes: Quantum Theory for the Perplexed*. Wiley-VCH, 2005.
- [12] M. Ozawa. Universal uncertainty principle, simultaneous measurability, and weak values. *AIP Conf. Proc.*, 1363:53–62, 2011.
- [13] H. F. Hofmann. Reasonable conditions for joint probabilities of non-commuting observables. *Quantum Stud.: Math. Found.*, 1:39, 2014.
- [14] J. Lee and I. Tsutsui. Uncertainty relations for approximation and estimation. *Phys. Lett. A*, 380:2045, 2016.
- [15] J. L. Kelley. *General Topology*, volume 27 of *Graduate Texts in Mathematics*. Springer-Verlag, 1975.
- [16] J. R. Munkres. *Topology*. Prentice Hall, 2 edition, 2000.
- [17] B. v. Querenburg. *Mengentheoretische Topologie*. Springer-Lehrbuch. Springer-Verlag, 3 edition, 2001.
- [18] J. Elstrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. Springer-Verlag, Berlin, 7 edition, 2011.
- [19] H. Amann and J. Escher. *Analysis I*. Grundstudium Mathematik. Birkhäuser Verlag, 2 edition, 2006.
- [20] H. Amann and J. Escher. *Analysis II*. Grundstudium Mathematik. Birkhäuser Verlag, 2 edition, 2006.
- [21] H. Amann and J. Escher. *Analysis III*. Grundstudium Mathematik. Birkhäuser Verlag, 2 edition, 2008.
- [22] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill, 1976.
- [23] W. Rudin. *Real and Complex Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill Publishing Company, 3 edition, 1986.
- [24] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 2 edition, 1991.
- [25] D. Werner. *Funktionalanalysis*. Springer-Lehrbuch. Springer-Verlag, Berlin, 7 edition, 2011.
- [26] M. Reed and B. Simon. *Fourier Analysis, Self-adjointness*, volume 2 of *Methods of Modern Mathematical Physics*. Academic Press, 1975.
- [27] M. Reed and B. Simon. *Functional Analysis*, volume 1 of *Methods of Modern Mathematical Physics*. Academic Press, 1980.
- [28] K. H. Goldhorn, H. P. Heinz, and M. Kraus. *Moderne mathematische Methoden der Physik*, volume 1. Springer-Verlag, 2009.
- [29] K. H. Goldhorn, H. P. Heinz, and M. Kraus. *Moderne mathematische Methoden der Physik*, volume 2. Springer-Verlag, 2010.
- [30] S. Roman. *Advanced Linear Algebra*, volume 135 of *Graduate Texts in Mathematics*. Springer-Verlag, 3 edition, 2008.
- [31] F. Hausdorff. Summationsmethoden und Momentfolgen. I. *Mathematische Zeitschrift*, 9:74–109, 1921.
- [32] F. Hausdorff. Summationsmethoden und Momentfolgen. I. *Mathematische Zeitschrift*, 9:280–199, 1921.
- [33] O. Hosten and P. Kwiak. Observation of the Spin Hall Effect of Light via Weak Measurements. *SCIENCE*, 319:787–790, 2008.
- [34] P. B. Dixon, D. J. Starling, A. N. Jordan, and J. C. Howell. Ultrasensitive Beam Deflection Measurement via Interferometric Weak Value Amplification. *Phys. Rev. Lett.*, 102(173601), 2009.
- [35] T. Koike and S. Tanaka. Limits on amplification by Aharonov-Albert-Vaidman weak measurement. *Phys. Rev. A*, 84(062106), 2011.
- [36] Y. Susa, Y. Shikano, and A. Hosoya. Optimal probe wave function of weak-value amplification. *Phys.*

- Rev. A*, 85(052110), 2012.
- [37] K. Nakamura, A. Nishizawa, and M. K. Fujimoto. Evaluation of weak measurements to all orders. *Phys. Rev. A*, 85(012113), 2012.
 - [38] G. C. Knee, J. Combes, C. Ferrie, and E. M. Gauger. Weak-value amplification: state of play. *arXiv*, (1410.6252), 2014.
 - [39] S. Tanaka and N. Yamamoto. Information amplification via postselection: A parameter-estimation perspective. *Phys. Rev. A*, 88(042116), 2013.
 - [40] G. C. Knee, G. A. D. Briggs, S. C. Benjamin, and E. M. Gauger. Quantum sensors based on weak-value amplification cannot overcome decoherence. *Phys. Rev. A*, 87(012115), 2013.
 - [41] G. C. Knee and E. M. Gauger. When Amplification with Weak Values Fails to Suppress Technical Noise. *Phys. Rev. X*, 4(011032), 2014.
 - [42] C. Ferrie and J. Combes. Weak Value Amplification is Suboptimal for Estimation and Detection. *Phys. Rev. Lett.*, 112(040406), 2014.
 - [43] J. Lee and I. Tsutsui. Merit of amplification by weak measurement in view of measurement uncertainty. *Quantum Studies: Mathematics and Foundations*, 1:65–78, 2014.
 - [44] T. Morita, T. Sasaki, and I. Tsutsui. Complex probability measure and Aharonov’s weak value. *Prog. Theor. Exp. Phys.*, (053A02), 2013.
 - [45] M. J. W. Hall. Exact uncertainty relations. *Phys. Rev. A*, 64(052103), 2001.
 - [46] L. M. Johansen. What is the value of an observable between pre- and postselection? *Phys. Lett. A*, 322:298–300, 2004.
 - [47] H. Weyl. Quantenmechanik und Gruppentheorie. *Zeitschrift für Physik*, 46(1):1–46, 1927.
 - [48] H. J. Groenewold. On the principles of elementary quantum mechanics. *Physica*, 12:405–460, 1946.
 - [49] J. E. Moyal. Quantum mechanics as a statistical theory. *Mathematical Proceedings of the Cambridge Philosophical Society*, 45:99–124, 1949.
 - [50] S. Wu and Y. Li. Weak measurements beyond the Aharonov-Albert-Vaidman formalism. *Phys. Rev. A*, 83(052106), 2011.

A. Post-selected Measurement

Given that the conditioning observable B has a spectrum of finite cardinality, we have seen in (4.48) that the conditional expectation $\mathbb{E}[X|B; \Psi^g]$ admits an explicit expression, of which value reduces to

$$\begin{aligned}\mathbb{E}[X|B = b; \Psi^g] &= \frac{\mathbb{E}[\Pi_b \otimes X; \Psi^g]}{\|(\Pi_b \otimes I)\Psi^g\|^2} \\ &= \mathbb{E}[X|\Pi_b = 1; \Psi^g]\end{aligned}\quad (\text{A1})$$

for the choice $b \in \sigma(B)$ such that the probability of observing it is non-vanishing. This is roughly to say that the description of a conditioning by a general observable B , hence the study of conditioned measurement scheme, essentially reduces to that given by a projection. Of course, this should be intuitively clear, since each self-adjoint operator admits a unique spectral decomposition. As the extreme case, the choice of the conditioning observable

$$B = |\phi'\rangle\langle\phi'| \quad (\text{A2})$$

given by a projection on a one-dimensional subspace of \mathcal{H} spanned by a unit vector $|\phi'\rangle \in \mathcal{H}$ becomes of special interest for our study. The vast majority of literatures with similar interest to this paper is devoted to the study of this special type of conditional measurement, and the act of measuring the conditional expectation

$$\mathbb{E}[X||\phi'\rangle\langle\phi'| = 1; \Psi^g] \quad (\text{A3})$$

is mostly referred to as the ‘post-selected measurement’ or the ‘weak measurement’. In such context, the unit vector $|\phi'\rangle$ is occasionally called the *final state*, denoted as $|\phi_f\rangle := |\phi'\rangle$, in order to contrast it with the initial state denoted as $|\phi_i\rangle := |\phi\rangle$.

A.1. Example: Analytic Model

We are now interested in the construction of a model in which the conditional expectation can be analytically computed for all range of the interaction parameter $g \in \mathbb{R}$. To this end, we assume that the target observable A has a spectrum of finite cardinality. One readily sees that the ‘conditional’ composite state essentially reduces to the computation of the vector

$$(\Pi_f \otimes I)|\Psi^g\rangle = \sum_{n=1}^N \langle\phi_f, \Pi_{a_n} \phi_i\rangle \cdot |\phi_f \otimes e^{-iga_n Y} \psi\rangle \quad (\text{A4})$$

for the special choice of the conditioning observable $B = \Pi_f := |\phi_f\rangle\langle\phi_f|$ defined by some final state $|\phi_f\rangle \in \mathcal{H}$. A careful observation reveals that the ‘conditional’ meter state (5.63), which is in general a mixed state for the general conditioning observable B , in fact becomes a pure state

$$|\psi_{\Pi_f=1}^g\rangle = \frac{\sum_{n=1}^N \langle\phi_f, \Pi_{a_n} \phi_i\rangle \cdot |e^{-iga_n Y} \psi\rangle}{\left\| \sum_{n=1}^N \langle\phi_f, \Pi_{a_n} \phi_i\rangle \cdot |e^{-iga_n Y} \psi\rangle \right\|} \quad (\text{A5})$$

for the post-selected measurement case, in which the conditional expectation reads

$$\begin{aligned}\mathbb{E}[X|\Pi_f = 1; \Psi^g] &= \mathbb{E}[X; \psi_{\Pi_f=1}^g] \\ &= \frac{\sum_{m=1}^N \sum_{n=1}^N \langle\phi_i, \Pi_{a_m} \phi_f\rangle \langle\phi_f, \Pi_{a_n} \phi_i\rangle \langle e^{-iga_m Y} \psi, X e^{-iga_n Y} \psi\rangle}{\sum_{m=1}^N \sum_{n=1}^N \langle\phi_i, \Pi_{a_m} \phi_f\rangle \langle\phi_f, \Pi_{a_n} \phi_i\rangle \langle e^{-iga_m Y} \psi, e^{-iga_n Y} \psi\rangle},\end{aligned}\quad (\text{A6})$$

whenever the denominator is non-vanishing, *i.e.*, when the ‘conditional’ meter state is not a zero vector. One thus learns that the computation of the conditional expectation essentially

reduces to the that of the quantity of the form

$$\langle e^{-iga_m Y} \psi, Z e^{-iga_n Y} \psi \rangle, \quad 1 \leq m, n \leq N \quad (\text{A7})$$

for the choice $Z = I, Q, P$.

Gaussian Example. For our demonstration, we consider the simplest non-trivial model in which the target observable A is dichotomic, that is, it has a discrete spectrum consisting of only two distinct eigenvalues $\{a_1, a_2\}$. For concreteness, we now assume that the meter system is described by the Schrödinger representation of the CCR $\{L^2(\mathbb{R}), \mathcal{S}(\mathbb{R}), \{\hat{x}, \hat{p}\}\}$, and choose $Y = \hat{p}$ without loss of generality. Despite its simplicity, this model should retain its usefulness in the sense that it covers the situations in recent experiments of weak measurement [33, 34]. We also note that the condition $A^2 = I$, under which the previous works [35, 37, 50] performed a full order calculation, is in fact a special case ($\{a_1, a_2\} = \{-1, 1\}$) of our setting.

Now, by recalling that the subspace $\mathcal{S}(\mathbb{R}) \subset L^2(\mathbb{R})$ is in particular invariant under the operations \hat{x} and \hat{p} , hence $\mathcal{S}(\mathbb{R}) \subset \mathcal{D}$, we see from our previous general argument that for any choice of the initial meter state $\psi \in \mathcal{S}(\mathbb{R})$ and the pair of pre- and post-selections satisfying the non-orthogonality condition $|\langle \phi_f | \phi_i \rangle| \neq 0$, the conditional expectation (4.54) should be well-defined on an appropriate neighbourhood of $g = 0$. For both definiteness and practicality, we shall choose the initial meter state $\psi \in \mathcal{S}(\mathbb{R})$ to be a real Gaussian wave function

$$\psi(x) := \pi^{-1/4} \exp\left(-\frac{x^2}{2}\right) \quad (\text{A8})$$

centred at $x = 0$ with normalisation $\|\psi\|_2 = 1$. In order to see how the choice of the parameter g and that of the initial meter state affects the result of the measurement, we consider the family of states $\{\psi_{(h)}\}_{h>0}$ scaled from the Gaussian state defined by

$$\psi_{(h)}(x) := (\pi h^2)^{-1/4} \exp\left(-\frac{x^2}{2h^2}\right), \quad (\text{A9})$$

where now the parameter h specifies the ‘width’ of the initial Gaussian profile of ψ (*cf.* (3.123)). One then finds

$$\begin{aligned} \langle e^{-iga_m Y} \psi, Z e^{-iga_n Y} \psi \rangle &= \int_{\mathbb{R}} \psi_{(h)}^*(x - ga_m) Z \psi_{(h)}(x - ga_n) d\beta(x) \\ &= \begin{cases} \exp\left(-\frac{g^2}{h^2} \left(\frac{a_m - a_n}{2}\right)^2\right), & Z = I, \\ \left(g \cdot \frac{a_m + a_n}{2}\right) \exp\left(-\frac{g^2}{h^2} \left(\frac{a_m - a_n}{2}\right)^2\right), & Z = \hat{x}, \\ \left(i \frac{g}{h^2} \cdot \frac{a_m - a_n}{2}\right) \exp\left(-\frac{g^2}{h^2} \left(\frac{a_m - a_n}{2}\right)^2\right), & Z = \hat{p}. \end{cases} \end{aligned} \quad (\text{A10})$$

Given the spectral decomposition $A = a_1 \Pi_{a_1} + a_2 \Pi_{a_2}$ for our case, we introduce the shorthand

$$\Lambda_m := \frac{a_1 + a_2}{2}, \quad \Lambda_r := \frac{a_2 - a_1}{2}, \quad A_w^0 := A_w - \Lambda_m \quad (\text{A11})$$

for later convenience, which respectively represents the barycentre of the two eigenvalues, the half-width of the numerical range, and the ‘centralised’ weak value of A defined by

$$A_w := \frac{\langle \phi_f, A \phi_i \rangle}{\langle \phi_f, \phi_i \rangle}. \quad (\text{A12})$$

One then finds through routine computation (see Appendix A.2 for computational detail) the following results

$$\mathbb{E}[\hat{x}|\Pi_f = 1; \Psi^g] = g \cdot \frac{\operatorname{Re}[A_w^0]}{1 + a(1 - e^{-g^2\Lambda_r^2/h^2})} + g\Lambda_m, \quad (\text{A13})$$

$$\mathbb{E}[\hat{p}|\Pi_f = 1; \Psi^g] = \frac{g}{2h^2} \cdot \frac{\operatorname{Im}[A_w^0]e^{-g^2\Lambda_r^2/h^2}}{1 + a(1 - e^{-g^2\Lambda_r^2/h^2})}, \quad (\text{A14})$$

where we have used the quantity

$$a := \frac{1}{2} \left(\left| \frac{A_w^0}{\Lambda_r} \right|^2 - 1 \right), \quad (\text{A15})$$

which is to be understood as a parameter corresponding to the ‘amplification rate’ of the ‘centralised’ weak value A_w^0 of A to the half-width of its numerical range Λ_r . We mention again that the result of the previous works in which $A^2 = 1$ is assumed is indeed a special case of the above formulae: we just put $\Lambda_m = 0$, $\Lambda_r = 1$, $a = \frac{1}{2}(|A_w|^2 - 1)$ to reproduce it.

Some Observations. While the general argument only assures that the shift of the conditional expectation values are well-defined on an appropriate neighbourhood U_0 of $g = 0$ for a given non-orthogonal choice of pre- and post-selections, the above result shows that it is in fact well-defined on the whole real line (hence $U_0 = \mathbb{R}$) for our case. Moreover, we also find that the shifts are indeed bounded for any choice of the pair of states of the target system due to the presence of the term $|A_w^0|^2$ hidden in the quantity a in the denominator.

As for the recovery of the weak value A_w , one realises that, since the present choice of the meter state implies $\psi \in \mathcal{S}(\mathbb{R}) \subset \mathcal{D}$, the general argument in the previous subsection guarantees the differentiability of the shift, and by noting that $\operatorname{CV}_S[\hat{x}, \hat{p}; \psi] = 0$ and $\operatorname{CV}[\hat{p}, \hat{p}; \psi] = \mathbb{V}[\hat{p}; \psi] = (2h^2)^{-1}$, one should have

$$\left. \frac{d}{dg} \mathbb{E}[X|\Pi_f = 1; \Psi^g] \right|_{g=0} = \begin{cases} \operatorname{Re}[A_w], & X = \hat{x}, \\ \operatorname{Im}[A_w] \cdot (2h^2)^{-1}, & X = \hat{p}, \end{cases} \quad (\text{A16})$$

based on the result (4.50). Indeed, observing that $\mathbb{E}[X|\Pi_f = 1; \Psi^0] = 0$ for both choices $X \in \{\hat{x}, \hat{p}\}$, one may directly verify this as

$$\begin{aligned} \left. \frac{d}{dg} \mathbb{E}[\hat{x}|\Pi_f = 1; \Psi^g] \right|_{g=0} &= \lim_{g \rightarrow 0} \frac{\mathbb{E}[\hat{x}|\Pi_f = 1; \Psi^g]}{g} \\ &= \lim_{g \rightarrow 0} \frac{\operatorname{Re}[A_w^0]}{1 + a(1 - e^{-g^2\Lambda_r^2/h^2})} + \Lambda_m \\ &= \operatorname{Re}[A_w^0] + \Lambda_m \\ &= \operatorname{Re}[A_w] \end{aligned} \quad (\text{A17})$$

and

$$\begin{aligned} \left. \frac{d}{dg} \mathbb{E}[\hat{p}|\Pi_f = 1; \Psi^g] \right|_{g=0} &= \lim_{g \rightarrow 0} \frac{\mathbb{E}[\hat{p}|\Pi_f = 1; \Psi^g]}{g} \\ &= \lim_{g \rightarrow 0} \frac{1}{2h^2} \cdot \frac{\operatorname{Im}[A_w^0]e^{-g^2\Lambda_r^2/h^2}}{1 + a(1 - e^{-g^2\Lambda_r^2/h^2})} \\ &= \operatorname{Im}[A_w] \cdot (2h^2)^{-1} \end{aligned} \quad (\text{A18})$$

as expected

Another observation worthy of note is that the scaled outputs $\mathbb{E}[\hat{x}|\Pi_f = 1; \Psi^g]/g$ and $\mathbb{E}[\hat{p}|\Pi_f = 1; \Psi^g]/(g/(2h^2))$ are dependent on the parameters g and h only through the combination hg^{-1} , and that both tend to their respective desired values

$$\lim_{hg^{-1} \rightarrow \infty} \frac{\mathbb{E}[\hat{x}|\Pi_f = 1; \Psi^g]}{g} = \lim_{hg^{-1} \rightarrow \infty} \frac{\text{Re}[A_w^0]}{1 + a(1 - e^{-g^2 \Lambda_r^2/h^2})} + \Lambda_m = \text{Re}[A_w], \quad (\text{A19})$$

$$\lim_{hg^{-1} \rightarrow \infty} \frac{\mathbb{E}[\hat{p}|\Pi_f = 1; \Psi^g]}{g/(2h^2)} = \lim_{hg^{-1} \rightarrow \infty} \frac{\text{Im}[A_w^0]e^{-g^2 \Lambda_r^2/h^2}}{1 + a(1 - e^{-g^2 \Lambda_r^2/h^2})} = \text{Im}[A_w] \quad (\text{A20})$$

by taking the limit of the combination $hg^{-1} \rightarrow \infty$. Observe that the manner in which we take the limit of the combination hg^{-1} to recover the desired information is the opposite between the unconditioned case (‘strong’/‘sharp’ measurement) (3.125) and the post-selected case above. Namely, here we may either take the interaction $g \rightarrow 0$ to the ‘weak’ limit, broaden the wave-function $h \rightarrow \infty$ to the ‘unsharp’ limit, or appropriately balancing the combination thereof and let $hg^{-1} \rightarrow \infty$ as a whole.

A.2. Computation of the Gaussian Example

For better readability, we write

$$c_n := \langle \phi_f, \Pi_{a_n} \phi_i \rangle, \quad n = 1, 2. \quad (\text{A21})$$

Observing that $\langle \phi_f, A\phi_i \rangle = a_1 c_1 + a_2 c_2$ and $\langle \phi_f, \phi_i \rangle = c_1 + c_2$, one has

$$\begin{aligned} A_r &:= \frac{\langle \phi_f, A\phi_i \rangle}{\langle \phi_f, \phi_i \rangle} - \Lambda_m \\ &= \frac{a_1 c_1 + a_2 c_2}{c_1 + c_2} - \Lambda_m \\ &= \frac{a_1 |c_1|^2 + a_2 |c_2|^2 + a_1 c_1 c_2^* + a_2 c_1^* c_2}{|c_1|^2 + |c_2|^2 + 2 \text{Re}[c_1^* c_2]} - \Lambda_m \\ &= \Lambda_r \cdot \frac{-|c_1|^2 + |c_2|^2 + 2i \text{Im}[c_1^* c_2]}{|c_1|^2 + |c_2|^2 + 2 \text{Re}[c_1^* c_2]}, \end{aligned} \quad (\text{A22})$$

whereby one obtains

$$\text{Re}[A_r] = \Lambda_r \cdot \frac{-|c_1|^2 + |c_2|^2}{|c_1|^2 + |c_2|^2 + 2 \text{Re}[c_1^* c_2]}, \quad (\text{A23})$$

$$\text{Im}[A_r] = \Lambda_r \cdot \frac{2 \text{Im}[c_1^* c_2]}{|c_1|^2 + |c_2|^2 + 2 \text{Re}[c_1^* c_2]} \quad (\text{A24})$$

and

$$\begin{aligned}
a &:= \frac{1}{2} \left(\frac{|A_r|^2}{\Lambda_r^2} - 1 \right) \\
&= \frac{1}{2} \left(\frac{(-|c_1|^2 + |c_2|^2)^2 + 4(\operatorname{Im}[c_1^* c_2])^2}{(|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2])^2} - 1 \right) \\
&= \frac{1}{2} \left(\frac{(|c_1|^2 + |c_2|^2)^2 - 4|c_1^* c_2|^2 + 4(\operatorname{Im}[c_1^* c_2])^2}{(|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2])^2} - 1 \right) \\
&= \frac{1}{2} \left(\frac{(|c_1|^2 + |c_2|^2)^2 - 4(\operatorname{Re}[c_1^* c_2])^2}{(|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2])^2} - 1 \right) \\
&= -\frac{2 \operatorname{Re}[c_1^* c_2]}{|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2]} \tag{A25}
\end{aligned}$$

in terms of $\{c_n\}$. Then, based on (A6) and (A10), one has

$$\langle \psi^g, I\psi^g \rangle = |c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2] e^{-g^2 \Lambda_r^2 / d^2}, \tag{A26}$$

$$\langle \psi^g, \hat{x}\psi^g \rangle = |c_1|^2 \cdot g a_1 + |c_2|^2 \cdot g a_2 + 2 \operatorname{Re}[c_1^* c_2] \cdot g \Lambda_m e^{-g^2 \Lambda_r^2 / d^2}, \tag{A27}$$

$$\langle \psi^g, \hat{p}\psi^g \rangle = 2 \operatorname{Im}[c_1^* c_2] \cdot \frac{g}{d^2} \Lambda_r e^{-g^2 \Lambda_r^2 / d^2}, \tag{A28}$$

which in turn yields

$$\begin{aligned}
\mathbb{E}[\hat{x}; \psi^g] &= \frac{\langle \psi^g, \hat{x}\psi^g \rangle}{\langle \psi^g, I\psi^g \rangle} \\
&= g \frac{|c_1|^2 \cdot a_1 + |c_2|^2 \cdot a_2 + 2 \operatorname{Re}[c_1^* c_2] \cdot \Lambda_m e^{-g^2 \Lambda_r^2 / d^2}}{|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2] e^{-g^2 \Lambda_r^2 / d^2}} \\
&= g \frac{|c_1|^2 \cdot a_1 + |c_2|^2 \cdot a_2 - (|c_1|^2 + |c_2|^2) \Lambda_m}{|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2] e^{-g^2 \Lambda_r^2 / d^2}} + g \Lambda_m \\
&= g \frac{\Lambda_r (-|c_1|^2 + |c_2|^2)}{|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2] e^{-g^2 \Lambda_r^2 / d^2}} + g \Lambda_m \\
&= g \frac{\Lambda_r (-|c_1|^2 + |c_2|^2)}{(|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2]) - 2 \operatorname{Re}[c_1^* c_2] (1 - e^{-g^2 \Lambda_r^2 / d^2})} + g \Lambda_m \\
&= g \cdot \frac{\operatorname{Re}[A_r]}{1 + a(1 - e^{-g^2 \Lambda_r^2 / d^2})} + g \Lambda_m \tag{A29}
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}[\hat{p}; \psi^g] &= \frac{\langle \psi^g, \hat{p}\psi^g \rangle}{\langle \psi^g, I\psi^g \rangle} \\
&= \frac{2 \operatorname{Im}[c_1^* c_2] \cdot \frac{g}{d^2} \Lambda_r e^{-g^2 \Lambda_r^2 / d^2}}{|c_1|^2 + |c_2|^2 + 2 \operatorname{Re}[c_1^* c_2] e^{-g^2 \Lambda_r^2 / d^2}} \\
&= \frac{g}{d^2} \cdot \frac{\operatorname{Im}[A_r] e^{-g^2 \Lambda_r^2 / d^2}}{1 + a(1 - e^{-g^2 \Lambda_r^2 / d^2})}. \tag{A30}
\end{aligned}$$