New Journal of Physics

The open access journal at the forefront of physics

PAPER

CrossMark

Quantum collapse rules from the maximum relative entropy principle

Frank Hellmann[1], Wojciech Kamiński[2] and Ryszard Paweł Kostecki[3,4]

1   Potsdam Institute for Climate Impact Research, Potsdam, Germany
2   Institute for Theoretical Physics, Faculty of Physics, University of Warsaw, Warszawa, Poland
3   Perimeter Institute for Theoretical Physics, Waterloo, Canada
4   Author to whom any correspondence should be addressed.

E-mail: Hellmann@pik-potsdam.de, Wojciech.Kaminski@fuw.edu.pl and Ryszard.Kostecki@fuw.edu.pl

Abstract

We show that the von Neumann–Lüders collapse rules in quantum mechanics always select the unique state that maximises the quantum relative entropy with respect to the premeasurement state, subject to the constraint that the postmeasurement state has to be compatible with the knowledge gained in the measurement. This way we provide an information theoretic characterisation of quantum collapse rules by means of the maximum relative entropy principle.

1. Introduction

The dynamics of quantum states in the orthodox (von Neumann's) foundations of quantum mechanics consist of two different prescriptions: the unitary evolution and the so-called 'collapse' of a quantum state to a subspace encoding the knowledge gained in the outcome of a measurement. The mappings (rules) describing this collapse were originally formulated by von Neumann [1] and later improved by Lüders [2]. There are two different forms of collapse. When one knows only that a measurement corresponding to an observable (a self-adjoint operator with a discrete spectrum) O has taken place, the 'weak' rule applies. It is defined as $\rho \mapsto \sum_{i \in I} P_i \rho P_i$, where ρ is the original quantum state (in general, a density operator), while $O = \sum_{i \in I} \lambda_i P_i$ is a spectral decomposition with some countable index set I (hence, $\sum_{i \in I} P_i = \mathbb{I}$, $P_i P_j = P_i \delta_{ij}$, and $\lambda_i \in \mathbb{R} \; \forall i, j \in I$). If a measurement corresponding to O has resulted in a specific value $\lambda_k \in \{\lambda_i | i \in I\}$ associated to a projector $P_k \in \{P_i | i \in I\}$, then the 'strong' rule, $\rho \mapsto P_k \rho P_k / \mathrm{tr}(\rho P_k)$, is applied.

The negative of Umegaki's quantum relative entropy [3, 4], $D(\rho, \sigma) = -S(\rho, \sigma) := \mathrm{tr}(\rho \ln \rho - \rho \ln \sigma) \in [0, \infty]$, can be used as a measure of distinguishability, or relative information content, of the quantum state σ from the state ρ. The use of D instead of S follows Wiener's idea that the 'amount of information is the negative of the quantity defined as entropy' [5]. Note that we call $S = -D$ the relative entropy, following the convention of [6] that makes the Gibbs–Shannon and von Neumann entropies the special cases of S, after adding a constant: $S_{\mathrm{vN}}(\rho) = S(\rho, \mathbb{I}/n) + \log(n)$.

The function D can be considered as a nonsymmetric distance: in general, $D(\rho, \sigma) \neq D(\sigma, \rho)$. If a given state is σ and we believe it to be ρ, it can be easier or harder to find our error than if their roles were reversed. Say, $\sigma = P$ with P some projector and $\rho = \mathbb{I}/n$. If we measure the property corresponding to $\mathbb{I} - P$, a single measurement can tell us that the state is not σ, whereas no single measurement could reveal the same of ρ. See e.g. [7, 8] for an overview of reasons for using $D(\rho, \sigma)$ as a measure of distinguishability and relative information content.

A key information theoretic property of the strong collapse rule is that the probability of measuring the value λ_k again, after having measured it once, is 1, which follows from $\mathrm{tr}\left(P_k \frac{P_k \rho P_k}{\mathrm{tr}(\rho P_k)}\right) = 1$. Repeated measurements add no new information. Clearly, the state $P_k \rho P_k / \mathrm{tr}(\rho P_k)$ is not the only state that has this property (note that P_k is not necessary a rank 1 projector). What we demonstrate in this letter is that, among all states that have this property, the strong collapse rule selects the state that is least distinguishable from the initial state ρ, that is, it has the minimum relative information $D(\rho, \cdot)$, in a suitably regularised sense. This allows for an information theoretic characterisation of the strong collapse rule: the state after measurement is the state that is least

distinguishable from the previous state, while being compatible with the new information gained by the measurement.

In order to derive the strong collapse rule, we will need two intermediate results. First we will show that the weak collapse rule produces the least distinguishable state among the block diagonal states. We then show that a weighted version of the strong collapse rule, $\rho \mapsto \sum_i p_i P_i \rho P_i / \mathrm{tr}(\rho P_i)$, is the least distinguishable amongst the states with blocks of fixed trace. This rule can be interpreted as corresponding to a measurement where we believe that the result P_i occurred with probability p_i. This intermediate step regularises the problem of a strong collapse, which is then obtained as a limiting case, by taking $p_i \mapsto \delta_{ik}$ with $k = 1$.

Our derivation of the collapse rules from the constrained maximisation of Umegaki's quantum relative entropy is of special importance in the context of epistemic and information theoretic approaches to the foundations of quantum theory. In this context, collapse rules have been considered as analogues of the Bayes–Laplace rule [9–12]. This analogy rested on mathematical and conceptual similarity, but was not derived from any single unifying principle. In the meantime, the Bayes–Laplace rule has been shown to be a special case of the constrained maximisation of the Kullback–Leibler relative entropy [13–16]. Our result provides the missing piece of the puzzle. Both the Bayes–Laplace and von Neumann–Lüders rules are special cases of a single epistemic principle of inductive inference (or, in other words, information theoretic state updating). This issue will be discussed in more detail in section 5.

After finishing this paper, we were informed about reference [17], where it is shown that a state $\sigma = \sum_i P_i \rho_i P_i$, where P_i are rank 1 projectors, minimises the functional $D(\rho, \sigma)$. This is a special case of our result for the weak collapse rule. The generalisation to arbitrary projectors is suggested in [18], but without a proof or an indication of a method of proving this statement. The technique used by us to prove a general theorem is essentially different from one applied in [17] (and it shows that this result for the rank >1 case is more substantial and nontrivial than for the rank 1 case).

A closely related paper [26] deals with the same type of problem as addressed here, but using a different mathematical approach, allowing for treatment of the infinite dimensional case. Further conceptual and mathematical discussion associated with the results of both papers is carried out there and in [28]. A recent work [19] proves that a partial trace is also a constrained maximiser of quantum relative entropy.

2. The setup

We will consider the finite dimensional case. Hence, quantum states will be identified with non-negative matrices of trace 1, which form the convex set $\mathcal{D}$ in the space of all hermitian $n \times n$ complex matrices.

The function $D(\cdot, \cdot)$ is jointly convex in both arguments [20], which implies that $D(\rho, \cdot)$ is convex on $\mathcal{D}$ for all $\rho \in \mathcal{D}$. Due to the finite dimensionality of the problem, we can use the first order condition for the existence of a minimum of a convex function (see e.g. [21], Theorems 1.2.7 and 2.2.1): if $\mathcal{V}$ is a convex subset of a finite dimensional topological vector space, and $f \colon \mathcal{V} \to \mathbb{R}$ is convex then x is a global minimum of f on $\mathcal{V}$ if and only if all directional derivatives of f at x are non-negative.

For a function differentiable at x this condition states that if x is in the interior of $\mathcal{V}$ then the derivatives of f need to vanish. If x belongs to some strata of the boundary of $\mathcal{V}$ then all tangential derivatives need to vanish whereas derivatives in inward transversal direction need to be non-negative.

In our minimisation problem we have a subspace $\mathcal{V} \subset \mathcal{D}$ of density matrices that is defined by a linear equation and thus is a subsimplex. The function $D(\rho, \cdot)$ restricts to a convex and differentiable function on $\mathcal{V}$ and we want to find its minimum. Thus we simply differentiate in the directions preserving $\mathcal{V}$ and set the derivatives to be positive. We will denote this condition by

$$C_{\mathcal{V}}^{\rho}(\cdot) := \partial_{\mathcal{V}} D(\rho, \cdot) = \partial_{\mathcal{V}} \mathrm{tr}(\rho \ln(\cdot)) \geqslant 0. \tag{1}$$

The next two sections will be concerned with evaluating this set of equations.

3. Weak collapse

In the case of a weak collapse due to the measurement of $O = \sum_i \lambda_i P_i$, the constraint set is given by the block diagonal density matrices

$$\mathcal{V}_w := \left\{ \sigma \in \mathcal{D} \mid [P_i, \sigma] = 0 \quad \forall P_i \right\}. \tag{2}$$

The condition (2) is equivalent with $\sigma \in \mathcal{V}_w$ iff $\sigma = \sum_i P_i \sigma P_i$, as well as with $\sigma \in \mathcal{V}_w$ iff $[O, \sigma] = 0$ (see [22] for a discussion).

We can parametrise $\mathcal{V}_w$ in terms of the singular value decomposition of σ. Every element of $\mathcal{V}_w$ is of the form

$$\sigma = U\Lambda U^*, \tag{3}$$

with Λ a trace 1 diagonal matrix with positive entries, U a unitary that is a product $U = \prod_i U_i$, where U_i is an identity on the range of $\mathbb{I} - P_i$. We have that $[U, P_i] = 0$, $[\Lambda, P_i] = 0$, and thus, writing $\sigma_i = \sigma \mid_{\text{ran}(P_i)}$ we have $f(\sigma) = \oplus_i \, f(\sigma_i) = \oplus_i \, f(\sigma_i)$; that is, functions (in the sense of the functional calculus) act blockwise on the space $\mathcal{V}_w$.

Let us consider first the variation $\partial_{\mathcal{V}_w} \text{tr}(\rho \ln(\cdot)) = 0$ in the direction parametrised by the U_i. Given a function on a Lie group $f(U)$ we can take the directional derivative by looking at the parameter derivative of a one parameter group of diffeomorphisms on U. As multiplication in a Lie group is differentiable we can pick the one parameter group of diffeomorphisms generated by left multiplication with the one-dimensional subgroup $\exp(tL)$,

$$\phi_t(U) = \exp(tL)(U). \tag{4}$$

We then define the directional derivative in direction L as the derivative of the pushforward of f along ϕ_t, $\partial_L f(\cdot) = \frac{\mathrm{d}}{\mathrm{d}t} \phi_t^\sharp f(\cdot)|_{t=0}$. For a function that is the trace of U in a particular representation this can be easily evaluated:

$$\frac{\mathrm{d}}{\mathrm{d}t} \phi_t^\sharp \, \text{tr}(AU)|_{t=0} = \frac{\mathrm{d}}{\mathrm{d}t} \, \text{tr}\left(A e^{tL} U\right)|_{t=0} =$$

$$\sum_{i,j} (A)_{ij} \frac{\mathrm{d}}{\mathrm{d}t}\left(e^{tL} U\right)_{ji}|_{t=0} = \sum_{i,j} (A)_{ij} (LU)_{ji} = \text{tr}(ALU). \tag{5}$$

A straightforward calculation shows that we further have $\frac{\mathrm{d}}{\mathrm{d}t} \phi_t^\sharp \, \text{tr}(AUBU^*)|_{t=0} = \text{tr}(ALUBU^*) - \text{tr}(AUBU^*L)$.

Note that $[L_i, P_j] = 0$, and in particular $L_i P_j = \delta_{ij} L_i$. The derivative then takes the form

$$\partial_{L_i} \sum_j \text{tr}\left(P_j \rho P_j U_j \ln \Lambda_j U_j^*\right) = \text{tr}\left(\rho_i L_i \ln \sigma_i\right) - \text{tr}\left(\rho_i \ln(\sigma_i) L_i\right) = 0,$$

$$\text{tr}\left(L_i \left[\ln \sigma_i, \rho_i\right]\right) = 0. \tag{6}$$

We thus see that if σ and $\sum_i P_i \rho P_i$ are concurrently diagonalisable, the above equation vanishes. In fact, since $[\ln \sigma_i, \rho_i]$ is traceless and $\{L_i, iL_i\}$ spans the space of all traceless matrices in the i-th matrix block, this is also a necessary condition.

Let us next consider the variation in the direction of the spectrum, that is the direction of Λ. We are interested in the case where σ and $\sum_i P_i \rho P_i$ are concurrently diagonalisable. Let κ_k^σ and κ_k^ρ be the eigenvalues of σ and $\sum_i P_i \rho P_i$ respectively. If $\kappa_i^\rho \neq 0$ and $\kappa_i^\sigma = 0$ then $D(\rho, \sigma) = \infty$, so this can not be the minimum if a state with finite relative entropy exists, and we can disregard this case here.

Let us first consider the case that all $\kappa_i^\rho \neq 0$. We have the condition

$$\partial_{\Lambda_\sigma} D(\rho, \sigma) = \partial_{\Lambda_\sigma}\left(-\text{tr}\left(\Lambda_\rho \ln \Lambda_\sigma\right)\right) = \partial_{\Lambda_\sigma} \sum_k - \kappa_k^\rho \ln \kappa_k^\sigma = 0. \tag{7}$$

The derivatives $\partial_{\Lambda_\sigma}$ have to preserve the trace. An overcomplete basis of such derivatives is given by $\partial_{\kappa_k^\sigma} - \partial_{\kappa_l^\sigma}$. Thus, for all k, l, m

$$\left(\partial_{\kappa_k^\sigma} - \partial_{\kappa_l^\sigma}\right) \sum_m - \kappa_m^\rho \ln \kappa_m^\sigma = 0,$$

$$\frac{\kappa_l^\rho}{\kappa_l^\sigma} - \frac{\kappa_k^\rho}{\kappa_k^\sigma} = 0. \tag{8}$$

So, the ratios of the eigenvalues of $\sum_i P_i \rho P_i$ and σ are fixed. As they both are trace 1, this implies they are the same.

Let assume now that I is the index set of all i such that $\kappa_i^\rho = 0$. If this set is nonempty then the above conditions cannot be satisfied. However, there is still a possibility that the minimum is on the boundary. The condition for the minimum on the boundary is weaker than the above. Namely, all derivatives in directions pointing in toward the set need to be positive. Such directions can be written as a linear combination

$$\sum_{i \in I, j \notin I} \alpha_{ij}\left(\partial_{\kappa_i^\sigma} - \partial_{\kappa_j^\sigma}\right) + \sum_{i,j \notin I} \alpha_{ij}\left(\partial_{\kappa_i^\sigma} - \partial_{\kappa_j^\sigma}\right) \tag{9}$$

with $\alpha_{ij} \geqslant 0$ for $i \in I, j \notin I$ and otherwise α_{ij} arbitrary, since the derivatives with negative coefficients at $\partial_{\kappa_i^\sigma}$ would otherwise point outside the set. For that it is enough to check basis derivatives

$$\forall_{i \in I, j \notin I} \left(\partial_{\kappa_i^\sigma} - \partial_{\kappa_j^\sigma} \right) \sum_m - \kappa_m^\rho \ln \kappa_m^\sigma \geqslant 0 \Rightarrow \forall_{l \notin I} \ \frac{\kappa_l^\rho}{\kappa_l^\sigma} \geqslant 0, \tag{10}$$

$$\forall_{i,j \notin I} \left(\partial_{\kappa_i^\sigma} - \partial_{\kappa_j^\sigma} \right) \sum_m - \kappa_m^\rho \ln \kappa_m^\sigma = 0 \Rightarrow \forall_{k,l \notin I} \ \frac{\kappa_l^\rho}{\kappa_l^\sigma} - \frac{\kappa_k^\rho}{\kappa_k^\sigma} = 0, \tag{11}$$

and we see this is a global minimum.

Recall that if $\kappa_i^\rho \neq 0$ when $\kappa_i^\sigma = 0$ then $D(\rho, \sigma) = \infty$. We now also need to consider the case that $\kappa_i^\rho = 0$ when $\kappa_i^\sigma \neq 0$. In that case we would get the full derivatives in the i direction, thus the equations (11) apply, which can not be satisfied unless all $\kappa_j^\sigma = 0$, which can not occur in $\mathcal{D}$.

Combining this with the above we have that

$$C_{\mathcal{V}_w}^\rho \left(\sum_i P_i \rho P_i \right) = \partial_{\mathcal{V}_w} D \left(\rho, \sum_i P_i \rho P_i \right) \geqslant 0. \tag{12}$$

The state $\sigma = \sum_i P_i \rho P_i$ is the only state σ satisfying $C_{\mathcal{V}_w}^\rho (\sigma) \geqslant 0$. The set $\mathcal{V}_w$ is convex, so from (12) and convexity of $D(\rho, \cdot)$, this is the unique global minimum.

4. Strong collapse

The conditions defining 'strong' collapse that were specified in Introduction lead us to a troubling situation, because for such states (containing zero eigenvalues) the relative entropy is almost always infinite. We will overcome the problem by deriving a generalised version of the strong collapse rule that is a quantum counterpart of Jeffrey's rule. The ordinary strong collapse rule will be then obtained by a limiting procedure.

Consider a constraint set given in terms of $p_i \in \mathbb{R}$ such that $\sum_i p_i = 1$ by

$$\mathcal{V}_s := \left\{ \sigma \in \mathcal{D} \big| \left[P_i, \sigma \right] = 0, \quad \mathrm{tr} \left(\sigma P_i \right) = p_i \ \forall P_i \right\}, \tag{13}$$

where $\{ P_i | i \in I \}$ is again determined by the spectral decomposition of an observable $O = \sum_{i \in I} \lambda_i P_i$. The set (13) can be interpreted as encoding the knowledge that the measurement outcome λ_i corresponding to a projection P_i occurs with a probability p_i.

Here we encounter a problem. If we have a p_i nonzero but $\mathrm{tr}(\rho P_i) = 0$, then every state in $\mathcal{V}_s$ will have relative entropy $-\infty$ to ρ. Moreover, even if we subtract the infinite constant, we find that the regularised distance does not depend on the state in the block P_i and there is no unique minimum. We thus will always assume that $\mathrm{tr}(\rho P_i) \neq 0$ for $p_i \neq 0$.

The variation in the U_i direction goes through as before. However the variation in the direction of the spectrum changes in that a basis is now given in terms of $\partial_{\kappa_k^{\sigma_i}} - \partial_{\kappa_l^{\sigma_i}}$, with $\kappa_k^{\sigma_i}$ and $\kappa_l^{\sigma_i}$ belonging to the same block P_i and thus being eigenvalues of σ_i. Thus only the fractions of eigenvalues within each block are fixed. This implies that the eigenvalues of σ_i are uniformly scaled relative to the eigenvalues of ρ_i. The condition $\sum_k \kappa_k^{\sigma_i} = p_i$ fixes σ_i to be $p_i \rho_i / \mathrm{tr}(\rho_i)$.

This shows that

$$C_{\mathcal{V}_s}^\rho \left(\sum_i p_i \frac{P_i \rho P_i}{\mathrm{tr}\left(P_i \rho P_i \right)} \right) \geqslant 0. \tag{14}$$

The state $\sigma = \sum_i p_i \frac{P_i \rho P_i}{\mathrm{tr}(P_i \rho P_i)}$ is the only state σ satisfying $\partial_{\mathcal{V}_s} D(\rho, \sigma) \geqslant 0$.

The strong collapse is a limiting case of the above projection, with all p_i going to zero except of one, p_1, corresponding to a projection P_1 that, in turn, corresponds to a measurement result given by an eigenvalue λ_1. We obtain this by taking the weak continuous limit

$$\lim_{p_2, p_3, \ldots \to 0} \mathrm{arginf}_{\sigma \in \mathcal{V}_s} \{ D(\rho, \sigma) \} = \lim_{p_2, p_3, \ldots \to 0} \sum_i p_i \frac{P_i \rho P_i}{\mathrm{tr}\left(P_i \rho P_i \right)}$$

$$= \frac{P_1 \rho P_1}{\mathrm{tr}\left(P_1 \rho P_1 \right)}. \tag{15}$$

Note that in the finite dimensional case that we consider here the weak topology and norm topology coincide.

5. The foundational view

In the orthodox formulation of quantum mechanics the 'collapse rules' are *postulated*. Thus, they are not deduced from any other more fundamental principle. They can be derived from several different conditions, see

[23, 24] for a review, but none of these conditions possesses the status of a fundamental principle of quantum theory. The weak collapse rule can be derived by taking the tensor product with an auxilliary state, followed by unitary evolution and a partial trace. This may serve as a derivation independent of interpretational issues (when this procedure is interpreted as an interaction with some ontic environment, it is usually considered as an instance of decoherence). However, no such construction exists for the strong rule. This fact, as well as the unclear relationship between the strong collapse rule and unitary evolution, renders the orthodox mathematical foundations conceptually insufficient, asking for further insights.

In general, an ontic interpretation of the quantum state leads to considering quantum collapse as a change of the 'state of being' of some 'material object/thing'. On the other hand an epistemic interpretation leads to considering quantum collapse as a change of the 'state of information' of some 'experiencing user/agent'. (There also is a corresponding difference in the meaning of the term 'measurement'.) In particular, the dynamical reduction approach of [25], belongs to the former class, providing an ontic *explanation* by means of a general dynamical principle from which the quantum collapse rule is derived. On the other hand, an epistemic *interpretation* of collapse rules as quantum mechanical analogues of the Bayes–Laplace rule $p(x) \mapsto p(x)p(b|x)/p(b)$ was proposed in [9–12]. However, no epistemic *explanation*, understood as a derivation from some fundamental principle of information theory (or statistical inference theory) has been offered. Our paper (as well as the closely related paper [26]) provides such a derivation.

Following the postulates of [27, 28] (which aim at reapproaching the foundations of quantum theory in the spirit of [29–32]), we demonstrated that the mapping to the unique solution of constrained minimisation of the relative information D,

$$\rho \mapsto \mathrm{arginf}_{\sigma \in \mathcal{Q}}\{D(\rho, \sigma)\}, \tag{16}$$

can serve as the general principle of quantum state change due to the acquisition of new information (represented by the constraints $\mathcal{Q}$). This amounts to selecting the quantum state that is the least distinguishable from the original state among all states that are in a strict agreement with the new knowledge (represented by the constraints).

In order to derive the quantum collapse rules from the principle (16), we needed to identify the information theoretic constraints that *define* the situations of weak and strong collapse. The 'weak' collapse amounts to encoding the information that a specific observable O has been subjected to measurement. A quantum state σ that carries such information has to be compatible with the possibility of measuring all eigenvalues of O precisely. Such a situation can be characterised by the condition $[\sigma, O] = 0$ (or, equivalently, $[P_i, \sigma] = 0 \ \forall P_i$). The 'strong' collapse should additionally result in a state that would reproduce the result of measurement of a particular eigenvalue with certainty (that is, with probability equal 1). That is, given a projector P encoding the outcome λ of the measurement, the post-collapse density operator σ should satisfy the condition of a 'weak' collapse, as well as $\mathrm{tr}(P\sigma) = 1$. This provides an interesting general insight into a structure of quantum theory: why it is possible to use (16) in order to derive various quantum state change rules without assuming the probabilistic interpretation carried under the label of the 'Born rule', the latter seems to be required for justification of the choice of constraints leading to a specific class of rules, including 'strong' collapse.

Our results can be considered as a quantum counterpart of derivations [13–16] of the Bayes–Laplace rule from the constrained maximisation of the Kullback–Leibler relative entropy [33], $S(p, q) := -\int_{\mathcal{X}} \mu(x)p(x)\log(p(x)/q(x))$, where $x \in \mathcal{X}$, while p and q are densities of probability measures with respect to a measure μ on $\mathcal{X}$. The functional $S(p, q)$ is a special case of Umegaki's quantum relative entropy $S(\sigma, \rho)$ for discrete $\mathcal{X}$ and $[\sigma, \rho] = 0$. This strengthens the analogy between the Bayes–Laplace and the von Neumann–Lüders rules: they are just two special cases of a single general principle of inductive inference, given by (16). From the Bayesian perspective, the state ρ is a prior, while σ, satisfying the constraints and maximising $S(\rho, \sigma)$, is a posterior.

6. Remarks

It has been known for quite a long time (see e.g. [34]) that a 'weak' collapse *leads to* an increase of the *absolute* entropy $-\mathrm{tr}(\rho \log \rho)$. Our result uncovers an unexpectedly strong inverse of this fact: a 'weak' collapse *is a result of* maximisation of the *relative* entropy $-\mathrm{tr}(\rho \log \rho - \rho \log \sigma)$ under specific constraints.

All earlier results on derivation of weak and strong collapse rules from minimisation of two point functionals on the space of quantum states [22, 35–42] were obtained for (various) *symmetric* quantum information distances. The importance of our result stems from the importance of (the negative of) Umegaki's relative entropy in quantum information theory as opposed to symmetric quantum information distances, which do not carry a similar semantic significance. This statement can be approached either axiomatically or pragmatically. On the axiomatic side, $D(\rho, \sigma)$ is characterised [43] by the direct sum property, invariance under

automorphisms (so, in particular, unitaries), additive decomposition under conditional expectations (onto subalgebra), and measurability over the state space. These properties eliminate all above symmetric information distances. An extensive discussion of the reasons for (and, in particular, applications of) these properties can be found in [48]. Furthermore, $D(\rho, \sigma)$ is a direct quantum generalisation of $-S(p, q)$, which can be also characterised 'dynamically' as a unique functional $\Phi(p, q)$ such that the mapping $q \mapsto \mathrm{arginf}_{p \in C}\{\Phi(p, q)\}$ satisfies a few very reasonable desiderata for information processing [15, 32, 44]. On the pragmatic side, $D(\rho, \sigma)$ is widely used in quantum information theory as the most fundamental measure of distinguishability of quantum states (see e.g. [7, 8, 45–47]). Hence, from the perspective of quantum information theoretic approaches to foundations of quantum theory, our results provide an essential, new perspective on the mathematical form of collapse rules due to quantum measurement.

As noted by one of the referees, this leads to a question whether the results of this paper can be reproduced (or extended) in the setting of generalised probabilistic theories [49, 50]. This setting lacks a general analogue of the collapse rules, but it allows us to introduce a well defined notion of information distance [51, 52] (which reduces to the Umegaki and Kullback–Leibler distances in quantum mechanical and probabilistic case, respectively)[5]. Hence, the possible extension of our result to generalised probabilistic theories can bring in new foundational insights (in particular—as suggested by a referee—one can ask whether defining a post-measurement state as a minimiser of a specific information distance given some type of constraints preselects some type of theories). The main open technical problem is how to replace the use of block diagonal decomposition and variational analysis of the spectrum of operators by some other method. It may be possible that a restriction to a subclass of theories satisfying some sort of spectral condition (see e.g. [56, 57]) will be necessary for this. We hope to return to this problem in another paper.

Acknowledgments

References

[1] von Neumann J 1932 *Mathematische Grundlagen der Quantenmechanik* (Berlin: Springer)
[2] Lüders G 1951 Über die Zustandsänderung durch den Messprozess *Ann. Phys. Leipzig* **8** 322
[3] Umegaki H 1961 On information in operator algebras *Proc. Japan Acad.* **37** 459
[4] Umegaki H 1962 Conditional expectation in an operator algebra, IV (entropy and information) *Kōdai Math. Sem. Rep.* **14** 59
[5] Wiener N 1948 *Cybernetics or Control and Communication in the Animal and the Machine* (Cambridge: MIT Press)
[6] Bratteli O and Robinson D W 1979, 1981 *Operator Algebras and Quantum Statistical Mechanics* vol 1-2 (Berlin: Springer)
[7] Schumacher B and Westmoreland M D 2002 Relative entropy in quantum information theory *Quantum Computation and Information* ed S J Lomonaco Jr and H E Brandt (Providence, RI: American Mathematical Society) p 265 (arXiv:quant-ph/0004045)
[8] Vedral V 2002 The role of relative entropy in quantum information theory *Rev. Mod. Phys.* **74** 197
[9] Bub J 1977 Von Neumann's projection postulate as a probability conditionalization rule in quantum mechanics *J. Phil. Logic* **6** 381
[10] Bub J 1979 The measurement problem in quantum mechanics *Problems in the Foundations of Physics* ed G Toraldo di Francia (Amsterdam: North-Holland) p 71
[11] Fuchs C A 2002 Quantum mechanics as quantum information (and only a little more) (arXiv:quant-ph/0205039)
[12] Jacobs K 2002 How do two observers pool their knowledge about a quantum system? *Quantum Inf. Process.* **1** 73
[13] Williams P M 1980 Bayesian conditionalisation and the principle of minimum information *Br. J. Phil. Sci.* **31** 131
[14] Warmuth M K 2005 A Bayes rule for density matrices *Advances in Neural Information Processing Systems 18* (*NIPS 05*) ed Y Weiss *et al* (Cambridge, MA: MIT Press) p 1457
[15] Caticha A and Giffin A 2006 Updating probabilities *Bayesian Inference and Maximum Entropy Methods in Science and Engineering* ed A Mohammad-Djafari (Melville, NY: American Institute of Physics) p 31
[16] Douven I and Romeijn J-W 2012 A new resolution of the Judy Benjamin problem *Mind* **479** 637
[17] Modi K, Paterek T, Son W, Vedral V and Williamson M 2010 Unified view of quantum and classical correlations *Phys. Rev. Lett.* **104** 080501
[18] Coles P J 2012 Unification of different views of decoherence and discord *Phys. Rev. A* **85** 042103

---

[5] This problem seems to reflect quite similar issue in quantum logic, where the collapse rules are also not a part of the basic framework, so they require an additional justification or derivation (see e.g. [53]), pointing towards a possibility of some more fundamental principle. Such derivations were provided (under some assumptions) by means of minimisation of suitable symmetric distances (see e.g. [38, 39, 42]), corresponding (in some cases) to symmetric transition probability functionals. It was shown in [54] that nonsymmetric transition probability [55] plays more fundamental role in quantum logic, but its relationship to nonsymmetric distances and to the problem of derivation of collapse rules was not investigated.

[19] Munk-Nielsen M I 2015 Quantum measurements from entropic projections *MSc Thesis* Perimeter Institute for Theoretical Physics, Waterloo

[20] Lindblad G 1974 Expectations and entropy inequalities for finite quantum systems *Commun. Math. Phys.* **39** 111

[21] Borwein J M and Vanderwerff J D 2010 *Convex Functions: Constructions, Characterizations and Counterexamples* (Cambridge: Cambridge University Press)

[22] Herbut F 1969 Derivation of the change of state in measurement from the concept of minimal measurement *Ann. Phys.* **55** 271

[23] Busch P, Lahti P J and Mittelstaedt P 1991 *The Quantum Theory of Measurement* (Berlin: Springer)

[24] Busch P, Grabowski M and Lahti P J 1995 Repeatable measurements in quantum theory: their role and feasibility *Found. Phys.* **25** 1239

[25] Ghirardi G, Rimini A and Weber T 1986 Unified dynamics for microscopic and macroscopic systems *Phys. Rev.* D **34** 470

[26] Kostecki R P 2014 Lüders' and quantum Jeffrey's rules as entropic projections (arXiv:1408.3502)

[27] Kostecki R P 2010 Quantum theory as inductive inference *Proc. 30th Int. Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering* ed J-F Bercher and P Bessière (Melville, NY: American Institute of Physics) p 24 (arXiv:1009.2423)

[28] Kostecki R P 2016 Quantum information kinematics and dynamics beyond quantum mechanics (in preparation, based on talk http://pirsa.org/15050090)

[29] Jaynes E T 1957 Information theory and statistical mechanics *Phys. Rev.* **106** 620

[30] Kullback S 1959 *Information Theory and Statistics* (New York: Wiley)

[31] Jaynes E T 1979 Where do we stand on maximum entropy? *The Maximum Entropy Formalism* ed R D Levine and M Tribus (Cambridge, MA: MIT Press) p 15

[32] Shore J E and Johnson R W 1980 Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy *IEEE Trans. Inf. Theory* **26** 26

[33] Kullback S and Leibler R A 1951 On information and sufficiency *Ann. Math. Stat.* **22** 79

[34] Wehrl A 1978 General properties of entropy *Rev. Mod. Phys.* **50** 221

[35] Marchand J-P and Wyss W 1977 Statistical inference and entropy *J. Stat. Phys.* **16** 349

[36] Benoist R W, Marchand J-P and Yourgrau W 1977 Statistical inference and quantum mechanical measurement *Found. Phys.* **7** 827 (addend. 1978 *Found. Phys.* **8** 117)

[37] Gudder S P, Marchand J-P and Wyss W 1979 Bures distance and relative entropy *J. Math. Phys.* **20** 1963

[38] Hadjisavvas N 1981 Distance between states and statistical inference in quantum theory *Ann. Inst. Henri Poincaré* A **38** 167

[39] Dieks D and Veltkamp P 1983 Distance between quantum states, statistical inference and the projection postulate *Phys. Lett.* A **97** 24

[40] Marchand J-P 1983 Statistical inference by minimal Bures distance *Old and New Questions in Physics, Cosmology, Philosophy, and Theoretical Biology. Essays in Honor of Wolfgang Yourgreau* ed A van der Merwe (New York: Plenum) p 275

[41] Marchand J-P 1983 Statistical inference in non-commutative probability *Rend. Sem. Math. Fis. Milano* **52** 551

[42] Raggio G A 1984 Generalized transition probabilities and applications *Quantum Probability and Appplications to Quantum Theory of Irreversible Processes* ed L Accardi (Berlin: Springer) p 327

[43] Petz D 1992 Characterization of the relative entropy of states of matrix algebras *Acta Math. Hung.* **59** 449

[44] Csiszár I 1991 Why last squares and maximum entropy? An axiomatic approach to inference for linear inverse problems *Ann. Stat.* **19** 2032

[45] Vedral V, Plenio M B, Rippin M A and Knight P L 1997 Quantifying entanglement *Phys. Rev. Lett.* **78** 2275

[46] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quantum Information* (Cambridge: Cambridge University Press)

[47] Keyl M 2002 Fundamentals of quantum information theory *Phys. Rep.* **369** 431

[48] Ohya M and Petz D 1993 *Quantum Entropy and its Use* (Berlin: Springer)

[49] Barrett J 2007 Information processing in generalized probabilistic theories *Phys. Rev.* A **75** 032304

[50] Barnum H and Wilce A 2011 Information processing in convex operational theories *Electr. Not. Theor. Comp. Sci.* **270** 3–15

[51] Short A J and Wehner S 2010 Entropy in general physical theories *New J. Phys.* **12** 033023

[52] Barnum H, Barrett J, Orloff Clark L, Leifer M, Spekkens R W, Stepanik N, Wilce A and Wilke R 2010 Entropy and information causality in general probabilistic theories *New J. Phys.* **12** 033024

[53] Stairs A 1982 Discussion: quantum logic and the Lüders rule *Phil. Sci.* **49** 422

[54] Guz W 1980 A non-symmetric transition probability in quantum mechanics *Rep. Math. Phys.* **17** 385

[55] Mielnik B 1969 Theory of filters *Commun. Math. Phys.* **15** 1

[56] Alfsen E M and Shultz F W 2003 *Geometry of State Spaces of Operator Algebras* (Basel: Birkhäuser)

[57] Krumm M 2015 Thermodynamics and the structure of quantum theory as a generalized probabilistic theory *MSc Thesis* University of Heidelberg, Heidelberg (arXiv:1508.03299)