Sufficient conditions for uniqueness of the weak value

Sufficient conditions for uniqueness of the weak value

J Dressel and A N Jordan

Department of Physics and Astronomy, University of Rochester, Rochester, NY 14627, USA

E-mail: jdressel@pas.rochester.edu and jordan@pas.rochester.edu

Abstract

We review and clarify the sufficient conditions for uniquely defining the generalized weak value as the weak limit of a conditioned average using the contextual values formalism introduced in Dressel, Agarwal and Jordan (2010 *Phys. Rev. Lett.* **104** 240401). We also respond to criticism of our work by Parrott (arXiv:1105.4188v1) concerning a proposed counter-example to the uniqueness of the definition of the generalized weak value. The counter-example does not satisfy our prescription in the case of an underspecified measurement context. We show that when the contextual values formalism is properly applied to this example, a natural interpretation of the measurement emerges and the unique definition in the weak limit holds. We also prove a theorem regarding the uniqueness of the definition under our sufficient conditions for the general case. Finally, a second proposed counter-example by Parrott (arXiv:1105.4188v6) is shown not to satisfy the sufficiency conditions for the provided theorem.

PACS numbers: 03.65.Ca, 03.65.Ta, 03.67.−a

1. Introduction

Since its definition in 1988 by Aharonov, Albert and Vaidman, the weak value [1] of a quantum operator $\hat{A}$ has been a source of considerable controversy. The formal weak value expression,

$$A_w = \frac{\langle \psi_f | \hat{A} | \psi_i \rangle}{\langle \psi_f | \psi_i \rangle}, \tag{1.1}$$

was originally derived as the weak coupling limit of the shift in the mean of a Gaussian momentum pointer $\hat{p}$ under a specific von Neumann interaction Hamiltonian $\hat{H}_I(t) = -g(t)\hat{q} \otimes \hat{A}$ that coupled the conjugate position pointer $\hat{q}$ to a system observable $\hat{A}$ subject to the double boundary conditions of a pure initial preparation state $|\psi_i\rangle$ and a pure final post-selection state $|\psi_f\rangle$. Though the conditioned pointer shift lent itself to a natural interpretation as a conditioned average, the weak value expression (1.1) violated such intuition by exceeding the eigenvalue range of the observable $\hat{A}$ and even being complex. Despite later experimental

confirmation of the effect [2], there was a feeling that such a strange quantity would prove to be an anomalous curiosity.

Far from being an anomaly, however, the formal weak value expression (1.1) has persisted in the literature as a relatively stable quantity in a diverse array of systems. Ironically, the same features that made its interpretation troublesome have since been fruitfully used to theoretically address a number of conceptual difficulties in quantum mechanics, including the three-box paradox, Hardy's paradox, superluminal travel, Bohmian trajectories, complementarity, macrorealism violation and contextuality [3]. More practically, the inflation from the eigenvalue range has been exploited to amplify small signals above the background noise, in polarization and interferometric experiments [4].

Given its increasingly common presence in the literature, there was considerable motivation to find a firmer foundation under which the formal expression (1.1) could be understood as a generally measurable feature related to an observable in a pre- and post-selected ensemble. Recently we provided such a foundation in the form of an article in *Physical Review Letters* [5] that indicated how the quantum weak value could be subsumed as an idealized special case of a more flexible operational formalism for the generalized measurement of observables, which we dubbed the *contextual values* formalism. Our letter indicated that a principled generalization of the weak value,

$$
{}_f\langle A\rangle_w = \frac{\mathrm{Tr}\left(\hat{E}_f^{(2)}(\hat{A}\hat{\rho} + \hat{\rho}\hat{A})\right)}{2\,\mathrm{Tr}\left(\hat{E}_f^{(2)}\hat{\rho}\right)},
\tag{1.2}
$$

could be uniquely defined as the weak measurement limit of the most general empirical conditioned average *under certain conditions* from a mixed initial state $\hat{\rho}$ and an unsharp post-selection represented by an arbitrary probability operator (or POVM element) $\hat{E}_f^{(2)}$. The generalization (1.2) reduces to the *real part* of (1.1) for pure states, clarifying the origin and significance of the formal expression (1.1) from a broader perspective. Detailed discussion on the derivation was to be saved for a longer paper reviewing and extending the full theory of contextual values, which has now also been posted [6].

The conditions under which the generalized weak value (1.2) can be uniquely defined as a limit point of a conditioned average have become recently contested in six versions of a lengthy arXiv paper [7] and a summary of the same [8]. The latter presents concise proposed counter-examples to the uniqueness of the definition (1.2) based on the understanding of our work, to which we now reply. The basic issue at hand is quite simple: under what conditions can one obtain the result (1.2) as the limit point of a conditioned average?

As we explicitly mention in [5], the conditioned average does not generally converge to (1.2) in the weak measurement limit; indeed, the limit can depend on the details of the detection setup, which we call the measurement context. We stress that our result (1.2) is thus not in contradiction to the general observation that *the weak value is not a unique limit point of a conditioned average*, which has been previously reported [10]. The sole issues being clarified here are the sufficient conditions for obtaining the context-independent special case (1.2) from the general form of the conditioned average.

This paper is devoted entirely to the subject of the uniqueness of the definition of the generalized weak value and is organized as follows. In section 2 we review some elements of the contextual value formalism. In section 3 we analyze a proposed counter-example from [8] with the contextual value formalism. In section 4 we review the motivation behind our protocol for contextual value assignment. This is followed in section 5 by a general theorem and proof of our original definition in [5] along with a precise statement of the sufficient conditions for our theorem to hold. After discussion of the theorem in section 6, we analyze

a second proposed counter-example from [9] in section 7. Finally, we give our conclusions in section 8.

2. Contextual value formalism

To keep this work self-contained, we briefly review the contextual values formalism introduced in [5] and expanded upon in [6]. The central observation of the contextual values formalism is that an observable $\hat{A}$ for a particular system can be completely measured indirectly using an imperfectly correlated detector. The formalism is powerful enough to subsume strong measurements, weak measurements, and any strength of measurement in between. Indeed, the von Neumann measurement used to derive the weak value (1.1) originally becomes a special case.

For the typical case of a detector with a pure preparation state $|d\rangle$ that is coupled to the system with any joint unitary operation $\hat{U}_{sd}$ and then subsequently measured in a detector basis $\{|j\rangle\}$, such an indirect measurement will be completely characterized by a set of measurement operators on the system $\{\hat{M}_j = \langle j|\hat{U}_{sd}|d\rangle\}$, which we call a *measurement context*. As in [5], we restrict ourselves to this typical case in what follows for simplicity; the straightforward generalization to impure detector preparation is detailed in [6].

When a system state $\hat{\rho}$ is conditioned on a particular outcome j of the detector, it becomes updated according to $\hat{\rho}_j = \hat{M}_j \hat{\rho} \hat{M}_j^\dagger / P(j)$, where the normalization probability for detecting the outcome j is given by $P(j) = \text{Tr}(\hat{\rho}\hat{E}_j)$. The positive probability operators $\{\hat{E}_j = \hat{M}_j^\dagger \hat{M}_j\}$ partition unity $\sum_j \hat{E}_j = \hat{1}$, forming a positive operator-valued measure (POVM) on the system space.

The expectation value of the observable $\hat{A}$ can be accurately measured by the imperfectly correlated detector provided that the following operator identity exists:

$$\hat{A} = \sum_j \alpha_j \hat{E}_j, \tag{2.1}$$

$$\langle A \rangle = \text{Tr}(\hat{\rho}\hat{A}) = \sum_j \alpha_j P(j), \tag{2.2}$$

which defines the *contextual values* $\{\alpha_j\}$ of the observable $\hat{A}$ with respect to the measurement context $\{\hat{M}_j\}$. As we shall explain in section 4, in the event that multiple solutions for the contextual values exist we prescribe picking the solution that places the tightest bound on the detector variance, which can be found using the pseudoinverse.

If the observable $\hat{A}$ also commutes with the entire measurement context $\forall j$, $[\hat{A}, \hat{M}_j] = 0$, then all the statistical moments of $\hat{A}$ can also be accurately measured by correlating *sequences* of measurements on the detector,

$$\langle A^n \rangle = \sum_{j_1 \cdots j_n} (\alpha_{j_1} \cdots \alpha_{j_n}) \text{Tr}(\hat{\rho}\hat{E}_{j_1} \cdots \hat{E}_{j_n}). \tag{2.3}$$

We call a detector that can measure all moments of $\hat{A}$ a *fully compatible* detector. In what follows we will concern ourselves mostly with fully compatible detectors.

For the special case of a projective detector, the measurement context $\{\hat{\Pi}_k\}$ consists of the spectral projections of $\hat{A}$, so (2.1) reduces to the spectral expansion $\hat{A} = \sum_k a_k \hat{\Pi}_k$ as a special case, where a_k are the eigenvalues of $\hat{A}$, and (2.3) reduces to the standard formula $\langle A^n \rangle = \sum_k a_k^n P(k)$ that needs only a single repeated measurement to obtain all moments. Hence, the contextual values can be considered to form a generalized spectrum for the observable that is specific to a particular measurement context.

If a second measurement is made after the first measurement of $\hat{A}$ that is characterized by an arbitrary second measurement context and associated probability operators $\{\hat{E}_f^{(2)}\}$, we can also construct the most general *conditioned averages* of the observable,

$$_f\langle A\rangle = \sum_j \alpha_j P(j|f), \tag{2.4}$$

$$P(j|f) = \frac{\text{Tr}\left(\hat{E}_f^{(2)}\hat{M}_j\hat{\rho}\hat{M}_j^\dagger\right)}{\sum_j \text{Tr}\left(\hat{E}_f^{(2)}\hat{M}_j\hat{\rho}\hat{M}_j^\dagger\right)}. \tag{2.5}$$

The post-selected conditional probabilities $P(j|f)$ are generalizations of the Aharonov–Bergmann–Lebowitz rule [3] that handle mixed states, general intermediate measurement, and unsharp post-selections. As the conditioned averages (2.4) are constructed entirely from measurable quantities, they form a principled foundation for deriving the generalization of the weak value (1.2) as a limiting value as the correlation between the system and detector vanishes.

In what follows, we stress that the contextual values formalism itself has not been challenged. Only the details of the derivation of the context-independent weak value (1.2) using the general conditioned average (2.4) are being contested.

3. Analysis of a counter-example

We now address the counter-example provided in [8]. A case where the number of POVM elements (or measurement operators in this case) exceeds the dimension of the Hilbert space for a system observable $\hat{A}$ is considered therein,

$$\hat{M}_1 = \begin{pmatrix} 1/2 + g & 0 \\ 0 & 1/2 - g \end{pmatrix},$$
$$\hat{M}_2 = \begin{pmatrix} 1/2 - g & 0 \\ 0 & 1/2 + g \end{pmatrix}, \tag{3.1}$$
$$\hat{M}_3 = \sqrt{1/2 - 2g^2}\,\hat{\mathbf{1}},$$
$$\hat{A} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \tag{3.2}$$

where the operators are expressed as matrices in the basis that diagonalizes $\hat{A}$.

To calibrate the measurement, one is then faced with determining contextual values $\{\alpha_j\}$ that satisfy the (now underspecified) equation (2.1). To see this in detail, since all operators commute and are diagonalized in the same basis, we can write (2.1) as the equivalent matrix equation, $\vec{a} = F\vec{\alpha}$, where $F_{kj} = \text{Tr}(\hat{\Pi}_k\hat{E}_j)$:

$$\begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} (1/2+g)^2 & (1/2-g)^2 & 1/2-2g^2 \\ (1/2-g)^2 & (1/2+g)^2 & 1/2-2g^2 \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix}. \tag{3.3}$$

This underspecified matrix equation is then solved in [8] by choosing $\alpha_1 = 1/g^2$ arbitrarily and then solving the resulting modified equation,

$$\begin{pmatrix} a - \dfrac{(1/2+g)^2}{g^2} \\ b - \dfrac{(1/2-g)^2}{g^2} \end{pmatrix} = \begin{pmatrix} (1/2-g)^2 & 1/2-2g^2 \\ (1/2+g)^2 & 1/2-2g^2 \end{pmatrix} \begin{pmatrix} \alpha_2 \\ \alpha_3 \end{pmatrix}, \tag{3.4}$$

which leads to the full solution,

$$\alpha_1 = \frac{1}{g^2},$$

$$\alpha_2 = \frac{1}{g^2} - \frac{a-b}{2g},$$

$$\alpha_3 = \frac{4 - g(a(1+2g)^2 - b(1-2g)^2 - 16g)}{4g^2(4g^2-1)},$$

$$= -\frac{1}{g^2} + \frac{a-b}{4g} + (a+b-8) + O(g),$$

(3.5)

which contains poles of order $1/g^2$ by construction. These poles then contribute an extra context-dependent term to the weak limit of (2.4) that is not included in (1.2). For the specific choice of the identity $a = b = 1$ considered in [8], then $\alpha_1 = \alpha_2 = 1/g^2$ and $\alpha_3 = (2g^2 + 1)/(g^2(4g^2 - 1))$.

We devote a considerable amount of space to this type of underspecified case in our four page letter [5]. We write on page 2, 'The latter case [where the number of POVM elements exceeds the dimension of the system operator] results in an infinite number of possible solutions, α_j. As such, we propose that the physically sensible choice of [contextual values] is the least redundant set uniquely related to the eigenvalues through the Moore–Penrose pseudoinverse'. All the examples we give in the paper use the pseudoinverse, and this discussion occurs immediately before the *conditioned average* section under contention.

The problem with the counter-example is that the pseudoinverse solution is not employed, and consequently the freedom in the set of underspecified equations is used to insert by hand an anomalous contextual value that diverges as $1/g^2$ in order to artificially produce an extra contribution to the result (1.2) in the $g \to 0$ weak limit. Indeed, we could go further by similarly choosing a contextual value that diverges as g^{-m}, where $m > 2$. Such a case would produce a formally divergent conditioned average in the weak limit.

However, if we solve for the contextual values using the prescription we describe in our paper, the assignment gives a clear physical interpretation to the measurement that is being done. The pseudoinverse solution is found from the singular value decomposition, $F = U\Sigma V^T$. For this example, we find

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$V = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & \dfrac{4g^2+1}{\sqrt{48g^4 - 8g^2 + 3}} & \dfrac{\sqrt{2}(4g^2-1)}{\sqrt{48g^4 - 8g^2 + 3}} \\ 1 & \dfrac{4g^2+1}{\sqrt{48g^4 - 8g^2 + 3}} & \dfrac{\sqrt{2}(4g^2-1)}{\sqrt{48g^4 - 8g^2 + 3}} \\ 0 & \dfrac{-2(4g^2-1)}{\sqrt{48g^4 - 8g^2 + 3}} & \dfrac{\sqrt{2}(4g^2+1)}{\sqrt{48g^4 - 8g^2 + 3}} \end{pmatrix},$$

(3.6)

$$\Sigma = \begin{pmatrix} 2g & 0 & 0 \\ 0 & \dfrac{1}{2}\sqrt{48g^4 - 8g^2 + 3} & 0 \end{pmatrix}.$$

The pseudoinverse is then $F^+ = V\Sigma^+ U^T$, where Σ^+ is a diagonal matrix inverting all nonzero elements of Σ^T. We can then find our prescribed solution as $\vec{\alpha} = F^+\vec{a}$, which for this example

is

$$\alpha_1 = \frac{a-b}{4g} + \frac{(a+b)(4g^2+1)}{48g^4-8g^2+3} = \frac{a-b}{4g} + \frac{a+b}{3} + O(g^2),$$

$$\alpha_2 = -\frac{a-b}{4g} + \frac{(a+b)(4g^2+1)}{48g^4-8g^2+3} = -\frac{a-b}{4g} + \frac{a+b}{3} + O(g^2), \qquad (3.7)$$

$$\alpha_3 = \frac{2(a+b)(1-4g^2)}{48g^4-8g^2+3} = \frac{2(a+b)}{3} + O(g^2).$$

The largest pole in the solution (3.7) has order $1/g$, which is the inverse of the smallest nonzero order of g in the POVM generated by (3.1)—we will show this is the general rule for pseudoinverse solutions that correctly satisfy $\vec{a} = F\vec{\alpha}$ with the lowest nonzero order in g. It is then easy to check that the generalized weak value (1.2) will be recovered from the conditioned average (2.4) in the weak limit as $g \to 0$ for any pre- and post-selection, as claimed.

For the special case of the identity, $a = b = 1$, that is considered, the solution (3.7) does not diverge as $g \to 0$, but actually converges to a constant. This behavior is intuitive because the measured system operator is the identity—the identity can always be constructed from the $g = 0$ POVM alone. In this case, the first two contextual values converge to the same value of $2/3$, while the third contextual value converges to $4/3$ and contributes twice as much to the average; this makes physical sense as the first two outcomes balance each other to produce the identity, while the third outcome directly corresponds to the identity being measured. Moreover, for the orthogonal case $a = 1, b = -1$ the first two contextual values simplify to $\pm(1/2g)$, while the third contextual value vanishes entirely; this makes physical sense since the third outcome is orthogonal to the operator being measured and can therefore be discarded. None of these physically intuitive features are present in the solution (3.5) presented in [8].

4. Pseudoinverse prescription

It is now worthwhile to review the pseudoinverse prescription, and to discuss its methodology and advantages. We recall that the equation we are solving is (2.1) in the form of the matrix equation $\vec{a} = F\vec{\alpha}$, where F is an $N \times M$ matrix (N being the dimension of the system, and M being the number of POVM elements) given by its elements, $F_{kj} = \text{Tr}(\hat{\Pi}_k \hat{E}_j)$. We can then decompose this matrix with the singular value decomposition, $F = U\Sigma V^T$, where U is an $N \times N$ orthogonal matrix, V is an $M \times M$ orthogonal matrix, and Σ is an $N \times M$ diagonal matrix of singular values. The pseudoinverse of F is then constructed as $F^+ = V\Sigma^+ U^T$, where Σ^+ is an $M \times N$ diagonal matrix formed by inverting the nonzero singular values. The pseudoinverse reduces correctly to the true inverse if one exists.

With the pseudoinverse in hand, we then find a uniquely specified solution $\vec{\alpha}_0 = F^+ \vec{a}$ that is directly related to the eigenvalues of the operator. Other solutions $\vec{\alpha} = \vec{\alpha}_0 + \vec{x}$ of (2.1) will contain additional components in the null space of F, and will thus deviate from this least redundant solution. Consequently, the solution $\vec{\alpha}_0$ has the least norm of all solutions, since $||\vec{\alpha}||^2 = ||\vec{\alpha}_0||^2 + ||\vec{x}||^2$ by the triangle inequality and the fact that $\vec{\alpha}_0$ and $\vec{x}$ live in orthogonal subspaces. Even in the case of an overdetermined set of equations (where the number of detector outcomes is less than the dimension of the system), the pseudoinverse will give the 'best fit' solution in the least-squares sense. This can be seen by solving $F^T F \vec{\alpha} = F^T \vec{a}$. One will also obtain $\vec{\alpha} = \vec{\alpha}_0 + \vec{x}$, where now $\vec{\alpha}_0$ does not solve $F\vec{\alpha} = \vec{a}$, but is the least-squares fit to it, and $\vec{x}$ is in the null space of $F^T F$. As a physical example of this last situation one could use a grid of point measurements like a pixel array to approximate measurements for a continuous variable, such as position.

In addition to the mathematical reasons for using the pseudoinverse in this context, there is an important physical one that we will now describe. As mentioned, a fully compatible detector can be used together with the contextual values to reconstruct any moment of a compatible observable. However, since the detector outcomes are imperfectly correlated with the observable, the contextual values typically lie outside of the eigenvalue range and many repetitions of the measurement must be practically performed to obtain adequate precision for the moments. Importantly, the uncertainty in the moments is controlled by the variance—not of the observable operator, but of the contextual values themselves,

$$\sigma^2 = \sum_j \alpha_j^2 P(j) - \langle A \rangle^2, \tag{4.1}$$

where $P(j)$ is the probability of outcome j. Since the mean of the contextual values is set by construction to the mean of the observable being measured, it is in the experimentalist's best interest to minimize the second moment of the contextual values. This moment has a simple upper bound of $\sum_j \alpha_j^2 P(j) < \sum_j \alpha_j^2 = ||\vec{\alpha}||^2$ because $0 < P(j) < 1$, which will also constitute an upper bound of the variance σ^2. In absence of prior knowledge about the system one is dealing with, this is a reasonable upper bound to make. Therefore, by minimizing this upper bound the pseudoinverse will choose a solution that provides rapid statistical convergence for observable measurements on the system given no prior knowledge of the system state.

For the case of the counter-example in [8], the solution (3.5) has to leading order the bound on the variance,

$$||\vec{\alpha}||^2 = \frac{3}{g^4} - \frac{3(a-b)}{2g^3} + O\left(\frac{1}{g^2}\right), \tag{4.2}$$

while the pseudoinverse solution (3.7) has to leading order the bound

$$||\vec{\alpha}||^2 = \frac{(a-b)^2}{8g^2} + \frac{2}{3}(a+b)^2 + O(g^2). \tag{4.3}$$

For any observable $\vec{a}$ the solution (3.5) has a detector variance bounded by leading order $1/g^4$, which could generally swamp any attempt to measure an observable near the weak limit. In particular, the conditioned averages (2.4) would not generally be tractable to obtain, so the anomalous weak limit derived in [8] may not be easily observable without a special initial state. However, the pseudoinverse solution (3.7) has a detector variance bounded by leading order $1/g^2$ in the worst case; moreover, for the identity, $a = b = 1$, then the bound on the noise minimizes to a constant as $g \to 0$.

5. General theorem

We now give a general proof of the result (1.2). To obtain this result, we make the following sufficient assumptions.

(i) The measurement operators $\{\hat{M}_j\}$ are analytic functions of a measurement strength parameter g, and thus have well-defined Taylor expansions around $g = 0$ such that $\forall j$, $\lim_{g \to 0} \hat{M}_j \propto \hat{1}$. This is physically reasonable because measurement operators are typically composed from matrix elements of an analytic evolution operator under an interaction Hamiltonian for which g is the coupling constant.

(ii) If a measurement operator $\hat{M}_j = \hat{U}_j \hat{E}_j^{1/2}$ is not positive, its unitary freedom $\hat{U}_j = \exp(ig\hat{G}_j)$ is generated by a Hermitian operator $\hat{G}_j$ that commutes with the density matrix $\hat{\rho}$ of the system, $\forall j$, $[\hat{G}_j, \hat{\rho}] = 0$. The reason for this assumption will become clear.

(iii) The equality $\hat{A} = \sum_j \alpha_j(g)\hat{E}_j(g)$ must be satisfied, where the contextual values $\alpha_j(g)$ are selected according to the pseudoinverse prescription.

(iv) The minimum nonzero order in g for all $\hat{E}_j(g)$ is g^n such that (iii) is satisfied. (In [5] we considered the typical case $n = 1$.)

(v) The POVM elements $\{\hat{E}_j\}$ all commute with the observable $\hat{A}$, so that they are diagonalizable in the same basis.

Then we have the following theorem: in the weak limit $g \to 0$ the context dependence of the conditioned average (2.4) vanishes and the generalized weak value (1.2) is uniquely defined.

We note before we prove this result that these are only the sufficient conditions for the unique definition (1.2) that we implied in [5]—some of the assumptions might be further weakened. For example, there may be other principled inversion schemes for the contextual values that also lead to the context-independent result (1.2).

To obtain the proof, we shall rewrite (2.4) in a useful form and then take the weak limit as $g \to 0$. Using the polar decomposition of the measurement operators $\hat{M}_j = \hat{U}_j \hat{E}_j^{1/2}$, we rewrite the probabilities that appear in (2.4) as

$$\mathrm{Tr}\left(\hat{E}_f^{(2)} \hat{M}_j \hat{\rho} \hat{M}_j^\dagger\right) = \mathrm{Tr}\left((\hat{U}_j^\dagger \hat{E}_f^{(2)} \hat{U}_j)\hat{\rho}_j'\right), \tag{5.1}$$

where the modified density operator is

$$\hat{\rho}_j' = \hat{E}_j^{1/2} \hat{\rho} \hat{E}_j^{1/2} = \tfrac{1}{2}\{\hat{E}_j, \hat{\rho}\} - \tfrac{1}{2}\left[\hat{E}_j^{1/2}, [\hat{E}_j^{1/2}, \hat{\rho}]\right], \tag{5.2}$$

and $\{a, b\} = ab + ba$ denotes the anticommutator.

From assumptions (i) and (iv) we have the lowest nonzero order expansion of the POVM $\hat{E}_j = p_j \hat{1} + g^n \hat{E}_j^{(n)} + O(g^{n+1})$ where $p_j \in (0, 1)$ are nonzero probabilities such that $\sum_j p_j = 1$. We therefore also have the expansion of the positive roots to the same order in g,

$$\hat{E}_j^{1/2}(g) = \sqrt{p_j}\hat{1} + g^n \hat{E}_j^{(n)}/2\sqrt{p_j} + O(g^{n+1}). \tag{5.3}$$

The probabilities p_j must be nonzero to satisfy assumption (i). The physical probability of outcome j is given by $P(j) = \mathrm{Tr}(\rho \hat{E}_j)$, and therefore converges to p_j in the weak limit, $g \to 0$.

Inserting expression (5.3) into (5.2), we find

$$\hat{\rho}_j' = p_j \hat{\rho} + \frac{g^n}{2}\{\hat{E}_j^{(n)}, \hat{\rho}\} - \frac{g^{2n}}{8p_j}[\hat{E}_j^{(n)}, [\hat{E}_j^{(n)}, \hat{\rho}]]. \tag{5.4}$$

This leaves the probabilities that appear in (2.4) to be

$$\mathrm{Tr}\left(\hat{E}_f^{(2)} \hat{M}_j \hat{\rho} \hat{M}_j^\dagger\right) = p_j \,\mathrm{Tr}\left((\hat{U}_j^\dagger \hat{E}_f^{(2)} \hat{U}_j)\hat{\rho}\right) + \frac{g^n}{2}\,\mathrm{Tr}\left((\hat{U}_j^\dagger \hat{E}_f^{(2)} \hat{U}_j)\{\hat{E}_j^{(n)}, \hat{\rho}\}\right), \tag{5.5}$$

plus a correction of order $O(g^{2n})$.

Invoking assumption (ii), since the generators $\hat{G}_j$ of unitaries commute with the density matrix, the unitary itself commutes with the density matrix. Consequently, the first term on the right-hand side of (5.5) simplifies to $p_j \,\mathrm{Tr}\left(\hat{E}_f^{(2)} \hat{\rho}\right)$. In the term of order $O(g^n)$, we can expand the unitary operator to first order in g, $U_j = \hat{1} + ig\hat{G}_j + O(g^2)$ to find that the second term on the right-hand side of (5.5) simplifies to $(g^n/2)\,\mathrm{Tr}\left(\hat{E}_f^{(2)}\{\hat{E}_j^{(n)}, \hat{\rho}\}\right)$ plus a correction of order $O(g^{n+1})$.

Thus, we find that the denominator of (2.4) is

$$\sum_j p_j \,\mathrm{Tr}\left(\hat{E}_f^{(2)} \hat{\rho}\right) + \sum_j (g^n/2)\,\mathrm{Tr}\left(\hat{E}_f^{(2)}\{\hat{E}_j^{(n)}, \hat{\rho}\}\right) + O(g^{n+1}). \tag{5.6}$$

However, since $\sum_j p_j = 1$ and $\sum_j \hat{E}_j^{(n)} = 0$ (the POVM condition), the denominator is simply $\text{Tr}\left(\hat{E}_f^{(2)} \hat{\rho}\right)$, with a correction of order $O(g^{n+1})$.

The numerator of (2.4) is given by summing (5.5) with the contextual values to find

$$\sum_j \alpha_j \text{Tr}\left(\hat{E}_f^{(2)}\left(\frac{1}{2}\{p_j\hat{1} + g^n\hat{E}_j^{(n)}, \hat{\rho}\}\right)\right) + O(g^{n+1}). \qquad (5.7)$$

We note that to order g^n, $\hat{A} = \sum_j \alpha_j(g)\left(p_j\hat{1} + g^n\hat{E}_j^{(n)}\right)$; since this sum exactly appears in the numerator, we recover our original result (1.2), up to a numerator correction of order $O(g^{n+1})$ times the order of each α_j. Thus, the only way the result (1.2) can be spoiled under our assumptions is if $\alpha_j(g)$ has a pole larger than $O(1/g^n)$. Hence, the last step in the proof will be to show that the pseudoinverse solution of $\alpha_j(g)$ cannot have a pole larger than $O(1/g^n)$.

To address the order of the contextual values $\alpha_j(g)$, we will first simplify notation by noting that $\hat{A}$ commutes with $\{\hat{E}_j(g)\}$ according to assumption (v). As such, we will replace all the diagonal matrices with vectors and rewrite the contextual values definition (2.1) from assumption (iii) as an equivalent matrix equation,

$$\vec{a} = F\vec{\alpha}, \qquad (5.8)$$

where

$$F = \begin{pmatrix} \vec{E}_1(g) & \vec{E}_2(g) & \ldots \end{pmatrix} = P + g^n F_n + O(g^{n+1}), \qquad (5.9)$$

and the two leading order matrices are defined as

$$P = \begin{pmatrix} p_1\vec{1} & p_2\vec{1} & \ldots \end{pmatrix}, \qquad (5.10)$$

$$F_n = \begin{pmatrix} \vec{E}_1^{(n)} & \vec{E}_2^{(n)} & \ldots \end{pmatrix}. \qquad (5.11)$$

As discussed, the minimum norm solution to (5.8) is the pseudoinverse solution $\vec{\alpha}_0 = F^+\vec{a}$. The pseudoinverse is constructed from the singular value decomposition $F = U\Sigma V^T$ as $F^+ = V\Sigma^+ U^T$, where U and V are orthogonal matrices such that $U^T U = VV^T = 1$, Σ is the singular value matrix composed of the square roots of the eigenvalues of FF^T, and Σ^+ is composed of the inverse nonzero elements in Σ^T. In order to satisfy (5.8), we then have the equivalent condition for each component of $U^T\vec{a} = \Sigma V^T\vec{\alpha}$,

$$(U^T\vec{a})_k = \Sigma_{kk}(V^T\vec{\alpha})_k. \qquad (5.12)$$

Therefore, all singular values Σ_{kk} corresponding to nonzero components of $U^T\vec{a}$ must also be nonzero; for brevity we shall call these the relevant singular values. Singular values which are not relevant will not contribute to the solution $\vec{\alpha} = V\Sigma^+ U^T\vec{a}$. Since $\alpha_j = (V\Sigma^+ U^T\vec{a})_j = \sum_k V_{jk}\Sigma_{kk}^+(U^T\vec{a})_k$, any zero element of $U^T\vec{a}$ will eliminate the inverse irrelevant singular value Σ_{kk}^+ from the solution for α_j.

Since the orthogonal matrices U and V have nonzero orthogonal limits $\lim_{g\to 0} U = U_0$ and $\lim_{g\to 0} V = V_0$, such that $U_0^T U_0 = V_0 V_0^T = 1$, and since $\vec{a}$ is g-independent, then the only poles in the solution $\vec{\alpha}_0 = F^+\vec{a} = V\Sigma^+ U^T\vec{a}$ must come from the inverses of the relevant singular values in Σ^+. If a singular value $\Sigma_{kk} = O(g^m)$, then $\Sigma_{kk}^+ = 1/\Sigma_{kk} = O(1/g^m)$; therefore, to have a pole of order higher than $O(1/g^n)$ then there must be at least one relevant singular value with a leading order greater than g^n. However, if that were the case then the expansion of F to order g^n would have a relevant singular value of zero and therefore could not satisfy (5.12), contradicting the assumption (iv) about the minimum nonzero order of the

POVM. Therefore, the pseudoinverse solution $\vec{\alpha}_0 = F^+ \vec{a}$ can have no pole with order higher than $O(1/g^n)$ and the theorem is proved.

6. Discussion

As we stated in our letter [5], 'we find that as $g \to 0$, the weak limit [of (2.4)] generally depends explicitly on $\{\hat{G}_j\}$ and $\{\alpha_j\}$, and thus will change depending on how it is measured and how the [contextual values] are chosen'. These dependences are apparent in the proof from equations (5.1) through (5.7). In other words, we find that the weak limit of the conditioned average *is not generally unique*. However, to produce any limiting value other than (1.2) one needs to violate the sufficient conditions given for our theorem. Namely, to find a different weak limit one needs either a nonanalytic or incompatible measurement context, a unitary disturbance that persists in the weak limit, a minimum nonzero order of g that does not satisfy the observable identity, or pathologically chosen CV.

In (5.2) the positive root of the POVM element $\hat{E}_j$ performs the information extraction of the measurement and modifies $\hat{\rho}$ to $\hat{\rho}_j'$, which consists of two terms: a symmetric term involving the POVM element itself, and a double commutator involving the roots. The symmetric term leads to the weak value (1.2), while the commutator term produces measurement disturbance away from (1.2). The unitary part of the POVM element in (5.1) rotates the post-selection $\hat{E}_f^{(2)}$ to a different post-selection that depends explicitly on the measurement result obtained, so this also disturbs the measurement process independently of the information extraction of the measurement.

For this reason, we consider a measurement consisting solely of positive POVM roots to be a *minimally disturbing* measurement, which is consistent with the usage of the term by Wiseman and Milburn [11]. That is, the information extraction of the measurement necessarily disturbs the system state by a minimum amount, but no additional unitary rotation occurs. Note that a weak measurement is an independent concept from a minimally disturbing measurement.

In [5] we named the limit as $g \to 0$ under the sufficiency condition $\forall j$, $[\hat{G}_j, \hat{\rho}] = 0$ placed on the unitary generators $\hat{G}_j$ the *minimal disturbance limit* since the measurement operators act like minimally disturbing POVM roots in that limit. The minimal disturbance definition $\hat{U}_j = \hat{1}$ becomes a special case.

7. Analysis of a second counter-example

Shortly after a preprint for this paper was posted, the reference [8] was updated to a sixth version [9] that adds a second proposed counter-example to our theorem, which we now address. The second proposed counter-example uses a three-outcome POVM to measure an observable in a three-dimensional Hilbert space to avoid any ambiguity related to the contextual values being underspecified. Specifically, the following measurement operators and observable are employed,

$$
\hat{M}_1 = \begin{pmatrix} \sqrt{1/2 + g} & 0 & 0 \\ 0 & \sqrt{1/2} & 0 \\ 0 & 0 & \sqrt{1/2 + g} \end{pmatrix},
$$
$$
\hat{M}_2 = \begin{pmatrix} \sqrt{1/3 + g^2} & 0 & 0 \\ 0 & \sqrt{1/3 + g} & 0 \\ 0 & 0 & \sqrt{1/3} \end{pmatrix},
$$

(7.1)

$$\hat{M}_3 = \begin{pmatrix} \sqrt{1/6 - g - g^2} & 0 & 0 \\ 0 & \sqrt{1/6 - g} & 0 \\ 0 & 0 & \sqrt{1/6 - g} \end{pmatrix},$$

$$\hat{A}_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \tag{7.2}$$

where we have corrected a minor typo in the definition of $\hat{M}_2$.[1] Computing the contextual values required to satisfy the relation $\hat{A}_p = \sum_i \alpha_i \hat{M}_i^2$ produces

$$\alpha_1 = \frac{1}{6g^2} - \frac{1}{g},$$

$$\alpha_2 = \frac{1}{6g^2} - \frac{1}{g}, \tag{7.3}$$

$$\alpha_3 = -\frac{5}{6g^2} - \frac{1}{g}.$$

The $1/g^2$ dependence of the contextual values can lead to the conditioned average (2.4) having additional context-dependent terms beyond the weak value (1.2) that are relevant in the weak limit, which seemingly contradicts our theorem.

This example, however, violates sufficiency condition (iv) for our theorem. Specifically, to first order in g—which is the lowest nonzero order—the POVM elements are

$$\hat{E}'_1 = \begin{pmatrix} 1/2 + g & 0 & 0 \\ 0 & 1/2 & 0 \\ 0 & 0 & 1/2 + g \end{pmatrix},$$

$$\hat{E}'_2 = \begin{pmatrix} 1/3 & 0 & 0 \\ 0 & 1/3 + g & 0 \\ 0 & 0 & 1/3 \end{pmatrix}, \tag{7.4}$$

$$\hat{E}'_3 = (1/6 - g)\hat{\mathbf{1}}.$$

While these first order POVM elements do satisfy the POVM condition $\hat{E}'_1 + \hat{E}'_2 + \hat{E}'_3 = \hat{\mathbf{1}}$, there is no exact solution to the required identity $\hat{A}_p = \sum_i \alpha_i \hat{E}'_i$.

We can see this fact by first noting that there is no solution for an arbitrary observable $\hat{A}$. Specifically, if we write the required identity as a matrix equation with $\vec{a} = F\vec{\alpha}$, with,

$$F = \begin{pmatrix} 1/2 + g & 1/3 & 1/6 - g \\ 1/2 & 1/3 + g & 1/6 - g \\ 1/2 + g & 1/3 & 1/6 - g \end{pmatrix}, \tag{7.5}$$

then F^{-1} does not exist since $\det(F) = 0$, so there is no general solution $\vec{\alpha} = F^{-1}\vec{a}$. However, there may still exist specific observables $\vec{a}'$ for which $\vec{a}' = F\vec{\alpha}$ is an underspecified system of equations with an infinite number of valid contextual value solutions. To rule out such a case

---

[1] The missing square root over the $1/3$ that is needed to satisfy the POVM condition and obtain the contextual values (7.3) has been restored.

for the specific observable $\vec{a}_p = (1, 0, 0)$, we compute the pseudoinverse solution $\vec{\alpha} = F^+ \vec{a}_p$,

$$F^+ \vec{a}_p = \begin{pmatrix} \dfrac{72g^2 + 30g + 11}{18g(12g^2 + 4g + 3)} \\[2ex] -\dfrac{18g^2 + 3g + 8}{9g(12g^2 + 4g + 3)} \\[2ex] -\dfrac{(6g - 1)^2}{18g(12g^2 + 4g + 3)} \end{pmatrix}, \tag{7.6}$$

and subsequently compute

$$F(F^+ \vec{a}_p) = \frac{1}{2} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}. \tag{7.7}$$

Since this does not equal $\vec{a}_p = (1, 0, 0)$, then $\vec{a}_p$ is partially in the null space of F^+ and there can be no exact solution to the required identity $\hat{A}_p = \sum_i \alpha_i \hat{E}'_i$.

Therefore, sufficiency condition (iv) for our theorem is violated and we do not expect the theorem to hold. Intuitively, the measurement (7.1) is not sufficiently correlated with the specific observable $\hat{A}_p$ as $g \to 0$ to guarantee the weak value (1.2) as the limit point of the conditioned average (2.4).

Moreover, if another observable $\hat{A}$ could be found such that $\hat{A} = \sum_i \alpha_i \hat{E}'_i$ were satisfiable to first order in g by the pseudoinverse solution, then the discussion after (5.12) in the proof of our theorem would apply. Hence, higher order poles would not appear in the contextual values, and the generalized weak value (1.2) would be obtained as the unique limit point of the conditioned average (2.4).

8. Conclusion

We have expanded upon and defended the claim made in our letter [5] that the context-independent generalized weak value (1.2) can be uniquely defined as a limit point of the conditioned average (2.4), and have given sufficient mathematical assumptions required for the definition to hold. Conceptually, the measurement context should depend on a measurement strength parameter g such that it reduces to the identity as $g \to 0$; any additional unitary disturbance in the measurement should not affect the state above and beyond the measurement being performed; the observable should be measurable to the lowest nonzero order in g; the contextual values of the measurement should be chosen to minimize an upper bound for the detector variance; and the probability operators for the measurement should commute with the observable.

We have also addressed two counter-examples to our definition that were proposed in versions 1 through 6 [2] of an arXiv post [8, 9]. In the former example our prescription for constructing contextual values in the case of a redundant detector (or underspecified measurement context) was not employed, and an anomalously divergent contextual value was inserted by hand; when our prescription for assigning contextual values is correctly applied, our theorem holds and a clear physical interpretation can be given to the measurement. In the latter example a measurement context was chosen that cannot construct the desired observable to the lowest nonzero order in g, so our theorem does not apply. Addressing these examples further demonstrates the power and utility of the contextual values formalism.

---

[2] Version 6 is the most recent version available at the time of writing, so any versions that subsequently appear are outside the scope of this paper.

Acknowledgments

References

[1] Aharonov Y, Albert D Z and Vaidman L 1988 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

[2] Ritchie N W M, Story J G and Hulet R G 1991 Realization of a measurement of a 'weak value' *Phys. Rev. Lett.* **66** 1107

Parks A D, Cullin D W and Stoudt D C 1998 Observation and measurement of an optical Aharonov–Albert–Vaidman effect *Proc. R. Soc.* **454** 2997

Brunner N, Scarani V, Wegmüller M, Legrè M and Gisin N 2004 Direct measurement of superluminal group velocity and signal velocity in an optical fiber *Phys. Rev. Lett.* **93** 203902

Resch K J, Lundeen J S and Steinberg A M 2004 Experimental realization of the quantum box problem *Phys. Lett.* A **324** 125

Pryde G J, O'Brien J L, White A G, Ralph T C and Wiseman H M 2005 Measurement of quantum weak values of photon polarization *Phys. Rev. Lett.* **94** 220405

Wang Q, Sun F W, Zhang Y S, Li J, Huang Y F and Guo G C 2006 Experimental demonstration of a method to realize weak measurement of the arrival time of a single photon *Phys. Rev.* A **73** 023814

Mir R, Lundeen J S, Mitchell M W, Steinberg A M, Garretson J L and Wiseman H M 2007 A double-slit 'which-way' experiment on the complementarity–uncertainty debate *New J. Phys.* **9** 287

Lundeen J S and Steinberg A M 2009 Experimental joint weak measurement on a photon pair as a probe of Hardy's paradox *Phys. Rev. Lett.* **102** 020404

Yokota K, Yamamoto T, Koashi M and Imoto N 2009 Direct observation of Hardy's paradox by joint weak measurement with an entangled photon pair *New J. Phys.* **11** 033011

Goggin M E, Almeida M P, Barbieri M, Lanyon B P, O'Brien J L, White A G and Pryde G J 2011 Violation of the Leggett–Garg inequality with weak measurement of photons *Proc. Natl Acad. Sci.* **108** 1256

Dressel J, Broadbent C J, Howell J C and Jordan A N 2011 Experimental violation of two-party Leggett–Garg inequalities with semiweak measurements *Phys. Rev. Lett.* **106** 040402

Kocsis S, Braverman B, Ravets S, Stevens M J, Mirin R P, Shalm L K and Steinberg A M 2011 Observing the average trajectories of single photons in a two-slit interferometer *Science* **332** 1170

Lundeen J S, Sutherland B, Patel A, Stewart C and Bamber C 2011 Direct measurement of the quantum wavefunction *Nature* **474** 188

[3] Aharonov Y, Bergmann P G and Lebowitz J L 1964 Time symmetry in the quantum process of measurement *Phys. Rev.* **134** B1410

Aharonov Y and Vaidman L 2008 The two-state vector formalism: an updated review *Lecture Notes Phys.* **734** 399

Aharonov Y, Botero A, Popescu S, Reznik B and Tollaksen J 2002 Revisiting Hardy's paradox: counterfactual statements, real measurements, entanglement and weak values *Phys. Lett.* A **301** 130

Rohrlich D and Aharonov Y 2002 Cherenkov radiation of superluminal particles *Phys. Rev.* A **66** 042102

Wiseman H M 2003 Directly observing momentum transfer in twin-slit 'which-way' experiments *Phys. Lett.* A **311** 285

Aharonov Y and Botero A 2005 Quantum averages of weak values *Phys. Rev.* A **72** 52111

Ralph T C, Bartlett S D, O'Brien J L, Pryde G J and Wiseman H M 2006 Quantum nondemolition measurements for quantum information *Phys. Rev.* A **73** 012113

Wiseman H M 2007 Grounding Bohmian mechanics in weak values and Bayesianism *New J. Phys.* **9** 165

Williams N S and Jordan A N 2008 Weak values and the Leggett–Garg inequality in solid-state qubits *Phys. Rev. Lett.* **100** 026804

Romito A, Gefen Y and Blanter Y M 2008 Weak values of electron spin in a double quantum dot *Phys. Rev. Lett.* **100** 056801

Hosoya A and Shikano Y 2010 Strange weak values *J. Phys. A: Math. Theor.* **43** 385307

[4] Hosten O and Kwiat P 2008 Observation of the spin Hall effect of light via weak measurements *Science* **319** 787

Dixon P B, Starling D J, Jordan A N and Howell J C 2009 Ultrasensitive beam deflection measurement via interferometric weak value amplification *Phys. Rev. Lett.* **102** 173601

Starling D J, Dixon P B, Jordan A N and Howell J C 2009 Optimizing the signal-to-noise ratio of a beam-deflection measurement with interferometric weak values *Phys. Rev.* A **80** 041803

Howell J C, Starling D J, Dixon P B, Vudyasetu P K and Jordan A N 2010 Interferometric weak value deflections: quantum and classical treatments *Phys. Rev.* A **81** 033813

Starling D J, Dixon P B, Williams N S, Jordan A N and Howell J C 2010 Continuous phase amplification with a Sagnac interferometer *Phys. Rev.* A **82** 011802

Starling D J, Dixon P B, Jordan A N and Howell J C 2010 Precision frequency measurements with interferometric weak values *Phys. Rev.* A **82** 063822

 [5] Dressel J, Agarwal S and Jordan A N 2010 Contextual values of observables in quantum measurements *Phys. Rev. Lett.* **104** 240401

 [6] Dressel J and Jordan A N 2011 Contextual values approach to the generalized measurement of observables arXiv:1110.0418v1

 [7] Parrott S 2011 'Contextual weak values' of quantum measurements with positive measurement operators are not limited to the traditional weak value arXiv:1102.4407v6

 [8] Parrott S 2011 Introduction to 'contextual values' and a simpler counterexample to a claim of Dressel, Agarwal, and Jordan (2010 *Phys. Rev. Lett.* **104** 240401) arXiv:1105.4188v1

 [9] Parrott S 2011 Counterexample to 'Sufficient conditions for uniqueness of the weak value' by J Dressel and A N Jordan, arXiv:1106.1871v1. arXiv:1105.4188v6

[10] Jozsa R 2007 Complex weak values in quantum measurement *Phys. Rev.* A **76** 044103

Di Lorenzo A and Egues J C 2008 Weak measurement: effect of the detector dynamics *Phys. Rev.* A **77** 042108

Parrott S 2009 What do quantum 'weak' measurements actually measure? arXiv:0908.0035v3

Parrott S 2010 Quantum weak values are not unique; what do they actually measure? arXiv:0909.0295v3

Geszti T 2010 Postselected weak measurement beyond the weak value *Phys. Rev.* A **81** 044102

Dressel J, Choi Y and Jordan A N 2011 Measuring which-path information with coupled electronic Mach–Zehnder interferometers arXiv:1105.2587

Wu S and Li Y 2011 Weak measurements beyond the Aharonov–Albert–Vaidman formalism *Phys. Rev.* A **83** 052106

Parks A D and Gray J E 2011 Variance control in weak-value measurement pointers *Phys. Rev.* A **84** 012116

Kofman A G, Ashhab S and Nori F 2011 Nonperturbative theory of weak pre- and post-selected measurements arXiv:1109.6315v1

[11] Wiseman H M and Milburn G J 2009 *Quantum Measurement and Control* (Cambridge: Cambridge University Press) p 40