Approximating incompatible von Neumann measurements simultaneously

Teiko Heinosaari,[1,*] Maria Anastasia Jivulescu,[1,2,†] Daniel Reitzner,[3,‡] and Mario Ziman[3,4,§]
[1]*Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen, Denmark*
[2]*Department of Mathematics, University Politehnica Timisoara, P-ta Victoriei Nr. 2, R-300006 Timisoara, Romania*
[3]*Institute of Physics, Slovak Academy of Sciences, Dúbravská cesta 9, 845 11 Bratislava, Slovakia*
[4]*Faculty of Informatics, Masaryk University, Botanická 68a, Brno, Czech Republic*
(Received 12 May 2010; revised manuscript received 28 July 2010; published 28 September 2010)

We study the problem of performing orthogonal qubit measurements simultaneously. Since these measurements are incompatible, one has to accept additional imprecision. An optimal joint measurement is the one with the least possible imprecision. All earlier considerations of this problem have concerned only joint measurability of observables, while in this work we also take into account conditional state transformations (i.e., instruments). We characterize the optimal joint instrument for two orthogonal von Neumann instruments as being the Lüders instrument of the optimal joint observable.

DOI: 10.1103/PhysRevA.82.032328 PACS number(s): 03.67.−a, 03.65.Ta

I. INTRODUCTION

It is a fundamental fact of quantum theory that there exist pairs of incompatible measurements. The simplest example is a pair of (ideal) spin component measurements in different directions. These measurements cannot be measured jointly using a single device.

The existence of incompatible measurements (i.e., impossibility of certain joint measurements) is linked with some other impossible tasks, such as cloning and teleportation. For each impossible device, one can study its best approximative substitute. This kind of optimal *possible* device then gives an absolute bound for the error one has to face in any attempt to build the impossible device. Evidently, this kind of quantitative bound on the error can tell us much more than just a plain statement of impossibility.

The question of approximate joint measurements of two sharp qubit observables (e.g., spin-1/2 components) was first studied in [1]. In recent years this topic has been investigated from several different aspects. The Mach-Zehnder interferometric setup was analyzed in [2,3] from the point of view of joint measurements. Various trade-off relations concerning joint approximations were derived in [4–7]. Characterizations of all jointly measurable two-outcome qubit observables were determined in [8–10]. A connection between the CHSH Bell inequality [11] and the bound on joint qubit measurements was observed in [12], and in [13] it was shown that every pair of two-outcome observables being not jointly measurable enables the violation of the CHSH Bell inequality. The relationship between cloning of observables and joint measurements was investigated in [14].

In the current work we study the question of approximate joint measurement of two sharp qubit measurements from a different perspective. In earlier works, discussion has concerned only joint measurability of *observables*. In this work we extend the problem to a joint measurability of *instruments*. In other words, we consider approximations not only to measurement outcome probabilities but also to conditional state transformations. One of our main results is the characterization of the optimal joint instrument for two orthogonal von Neumann instruments.

This paper is organized as follows. In Sec. II we explain the two different levels of compatibility. Some useful details on joint observables are presented in Sec. III. In Sec. IV a general form for joint instruments is derived. The optimal approximate joint instrument for two von Neumann instruments is then characterized in Sec. V. Finally, in Sec. VI we discuss the case of three von Neumann measurements.

II. TWO LEVELS OF INCOMPATIBILITY

Compatibility of quantum measurements has different meanings depending on what we take into consideration. In particular, two measurements can be compatible if we care only about the bare measurement outcome statistics, but fail to be compatible if we take into account the dynamics of the measurements. This fact is the motivation for the current investigation and in the following we explain this twofold meaning in detail.

In particular, let us consider two sharp observables X and Y on a qubit system. These can be, for instance, spin component measurements on a spin-$\frac{1}{2}$ system. The observables X and Y are described by the self-adjoint operators $\sigma_{\mathbf{x}} = \mathbf{x}\cdot\boldsymbol{\sigma}$ and $\sigma_{\mathbf{y}} = \mathbf{y}\cdot\boldsymbol{\sigma}$, where $\boldsymbol{\sigma} = (\sigma_1,\sigma_2,\sigma_3)$ are the Pauli matrices and $\mathbf{x},\mathbf{y}$ are unit vectors. Alternatively, and for our purposes more conveniently, these observables can be described by *projection valued measures* (PVMs). Then X and Y are identified as mappings from a set of measurement outcomes to projectors,

$$\pm 1 \mapsto \mathsf{X}(\pm 1) = \tfrac{1}{2}(I \pm \sigma_{\mathbf{x}}), \quad \pm 1 \mapsto \mathsf{Y}(\pm 1) = \tfrac{1}{2}(I \pm \sigma_{\mathbf{y}}).$$

A measurement of X (similarly Y) gives either a result up $(+1)$ or down (-1); see Fig. 1(a). For instance, if the system is in a state ϱ, then the probability of getting the outcome 1 in a measurement of X is $\mathrm{tr}[\varrho\mathsf{X}(1)]$. The operator $\sigma_{\mathbf{x}}$ gives the average value of the X measurement, which means that the formula,

$$\mathrm{tr}[\varrho\mathsf{X}(1)] - \mathrm{tr}[\varrho\mathsf{X}(-1)] = \mathrm{tr}[\varrho\sigma_{\mathbf{x}}],$$

holds for all states ϱ.

*heinosaari@nbi.dk
†jivulescu@nbi.dk
‡daniel.reitzner@savba.sk
§ziman@savba.sk

 ©2010 The American Physical Society

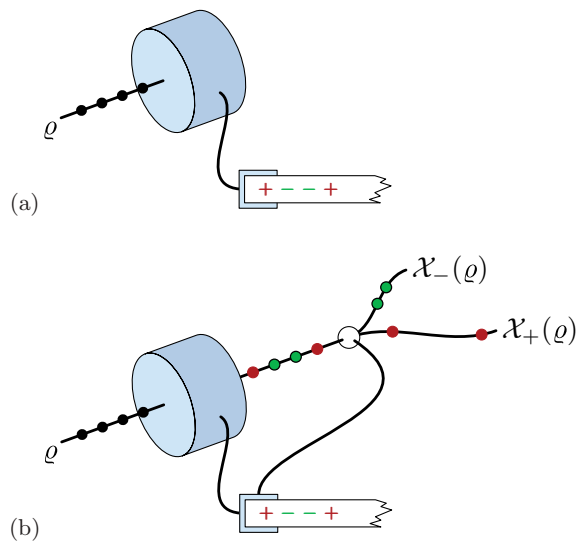

FIG. 1. (Color online) (a) An observable describes the measurement outcome statistics, whereas (b) an instrument describes both the measurement outcome statistics and the conditional output states with possible separation in accordance with the classical outcomes.

In our following investigation we assume that the unit vectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal. This is equivalent to the condition that $\mathrm{tr}[\mathsf{X}(i)\mathsf{Y}(j)] = \frac{1}{2}$ for every $i,j = \pm 1$. Hence, certain predictability of one outcome of X implies that both outcomes of Y are equally likely, and vice versa. This relation is usually referred to as "(value) complementarity" [15].

Our assumption on the orthogonality of $\mathbf{x}$ and $\mathbf{y}$ means, in particular, that the observables X and Y do not commute [i.e., $\mathsf{X}(i)\mathsf{Y}(j) \neq \mathsf{Y}(j)\mathsf{X}(i)$]. The noncommutativity implies the impossibility of performing their joint measurement. Therefore, we need to choose whether we measure X or Y, their simultaneous measurement being impossible.

It is possible to approximate X and Y with a pair of jointly measurable observables A and B described by positive operator valued measures (POVMs) [16,17]. An essential fact is that for POVMs (unlike for PVMs) commutativity is not a necessary condition for joint measurability.

Suppose we want to approximate X and Y equally well. Then a class of approximating observables, parametrized by a number $0 \leqslant \eta \leqslant 1$, is defined by

$$\mathsf{A}(\pm 1) = \tfrac{1}{2}(I \pm \eta\sigma_{\mathbf{x}}), \quad \mathsf{B}(\pm 1) = \tfrac{1}{2}(I \pm \eta\sigma_{\mathbf{y}}).$$

The number η quantifies how close A and B are to X and Y, respectively.

In the limiting case $\eta = 1$ we have $\mathsf{A} = \mathsf{X}$ and $\mathsf{B} = \mathsf{Y}$, however, in such a case A and B are not jointly measurable. It was shown in [1] that A and B have a joint measurement if and only if $\eta \leqslant \frac{1}{\sqrt{2}}$. Therefore, we fix $\eta = \frac{1}{\sqrt{2}}$ and A,B are then the optimal jointly measurable approximations to X,Y.

At this point one may wonder whether the η-parametrized class of observables leads to the best approximation, or perhaps some modification gives a better approximation (while preserving joint measurability). However, in [4,6] it has been proved that any modification to A and B leads either to a worse approximation or lack of joint measurability.

A joint observable for the observables A and B is defined as a POVM G with four outcomes corresponding to four possible pairs of A and B outcomes, $(\pm 1, \pm 1)$. It is required that the measurement outcome statistics for A (B) measured alone can be obtained from the joint observable by disregarding (summing through all possible) outcomes for B (A). Hence, the defining condition for G is that

$$\mathsf{A}(\pm 1) = \mathsf{G}(\pm 1, 1) + \mathsf{G}(\pm 1, -1), \tag{2.1a}$$

$$\mathsf{B}(\pm 1) = \mathsf{G}(1, \pm 1) + \mathsf{G}(-1, \pm 1). \tag{2.1b}$$

In other words, A and B are *marginals* of G. A possible choice is

$$\mathsf{G}(+1, \pm 1) = \frac{1}{4}\left[I + \frac{1}{\sqrt{2}}(\sigma_{\mathbf{x}} \pm \sigma_{\mathbf{y}})\right], \tag{2.2a}$$

$$\mathsf{G}(-1, \pm 1) = \frac{1}{4}\left[I - \frac{1}{\sqrt{2}}(\sigma_{\mathbf{x}} \mp \sigma_{\mathbf{y}})\right]. \tag{2.2b}$$

It is easy to verify that G indeed fulfills the requirements (2.1) and that each $\mathsf{G}(\pm 1, \pm 1)$ is a positive operator. Various ways to realize G and other related measurements have been discussed (e.g., in [2,18]).

So far, our discussion has concerned only joint measurability of observables (i.e., compatibility of measurement outcome probabilities). There is also another level of compatibility, arising from the fact that a (nontrivial) quantum measurement necessarily affects the state of the measured system. Thus, each measurement outcome has an associated operation, which is mathematically described as a completely positive trace-nonincreasing mapping on the set of states. The collection of all these operations forms an *instrument* [19].

The standard measurement for a discrete sharp observable is the so-called von Neumann measurement. The corresponding instrument, which we call the von Neumann instrument, has a very simple form. In our case, the von Neumann instruments $\mathcal{X}$ and $\mathcal{Y}$ associated with the sharp observables X and Y, respectively, are given by

$$\mathcal{X}_\pm(\varrho) = \mathsf{X}(\pm 1)\varrho\mathsf{X}(\pm 1), \tag{2.3a}$$

$$\mathcal{Y}_\pm(\varrho) = \mathsf{Y}(\pm 1)\varrho\mathsf{Y}(\pm 1). \tag{2.3b}$$

For instance, if the system is in a state ϱ and a measurement of X gives the outcome 1, then the unnormalized output state is $\mathcal{X}_+(\varrho)$; see Fig. 1(b). We can also write

$$\mathcal{X}_+(\varrho) = \mathsf{X}(1)\varrho\mathsf{X}(1) = \mathrm{tr}[\varrho\mathsf{X}(1)]\mathsf{X}(1),$$

which shows that the normalized output state is $\mathsf{X}(1)$.

Since X and Y are not jointly measurable, none of their instruments can be jointly measurable. In particular, there is no measurement scheme which would realize both $\mathcal{X}$ and $\mathcal{Y}$. Therefore, if we want to realize the instruments $\mathcal{X}$ and $\mathcal{Y}$ in a single measurement scheme, we need to approximate them.

In the case of the approximating observables A,B the von Neumann instruments are commonly replaced by Lüders instruments $\mathcal{L}^{\mathsf{A}}$ and $\mathcal{L}^{\mathsf{B}}$, defined as

$$\mathcal{L}^{\mathsf{A}}_\pm(\varrho) = \sqrt{\mathsf{A}(\pm 1)}\varrho\sqrt{\mathsf{A}(\pm 1)}, \tag{2.4a}$$

$$\mathcal{L}^{\mathsf{B}}_\pm(\varrho) = \sqrt{\mathsf{B}(\pm 1)}\varrho\sqrt{\mathsf{B}(\pm 1)}. \tag{2.4b}$$

For the operator $\mathsf{A}(1)$, the square root $\sqrt{\mathsf{A}(1)}$ takes the form,

$$\sqrt{\mathsf{A}(1)} = \sqrt{\frac{1+\eta}{2}}\,\mathsf{X}(1) + \sqrt{\frac{1-\eta}{2}}\,\mathsf{X}(-1).$$

Hence, we see that

$$\mathcal{L}^{\mathsf{A}}_{+}(\varrho) = \tfrac{1}{2}(1+\eta)\mathcal{X}_{+}(\varrho) + \text{other terms}.$$

In this way, we can understand the $\mathcal{L}^{\mathsf{A}}$ measurement as an approximate version of the $\mathcal{X}$ measurement. We refer to [20] for a convenient summary of the Lüders instrument in general.

In the limiting case $\eta = 1$ when $\mathsf{A} = \mathsf{X}$ and $\mathsf{B} = \mathsf{Y}$ the formulas (2.3) and (2.4) coincide. Hence, we would expect that the Lüders instruments of A and B are good approximations to the von Neumann instruments of X and Y. Here, however, we face a problem. It was shown in [21] that two Lüders operations $\mathcal{L}^{\mathsf{A}}_{+}$ and $\mathcal{L}^{\mathsf{B}}_{+}$ are jointly measurable if and only if[1] either $\mathsf{A}(1) + \mathsf{B}(1) \leqslant I$ or $\mathsf{B}(1) = b\mathsf{A}(1)$ for some $0 \leqslant b \leqslant 1$. Since neither of these two conditions holds in our situation, we find that the Lüders instruments $\mathcal{L}^{\mathsf{A}}$ and $\mathcal{L}^{\mathsf{B}}$ cannot be realized in a single experimental setup. Therefore, they do not provide jointly the measurable approximations that we are looking for.

We conclude that the obvious replacements for the von Neumann instruments of X and Y, namely the Lüders instruments of A and B, are not jointly measurable, although the observables A and B are. On the other hand, joint measurability of A and B implies that they have *some* jointly measurable instruments. In fact, every instrument implementing a joint observable of A and B gives instruments for A and B as its marginals. In the following we will characterize the joint instrument which gives the best approximations for the von Neumann instruments of X and Y.

III. JOINT OBSERVABLE

In this section we derive some useful properties of the joint observable G, defined in (2.2). Let us first make a general observation. Suppose that A and B would have a second joint observable G'. Then, also all the convex combinations $\lambda\mathsf{G} + (1-\lambda)\mathsf{G}'$, $0 < \lambda < 1$, defined as $(i,j = \pm 1)$

$$[\lambda\mathsf{G} + (1-\lambda)\mathsf{G}'](i,j) = \lambda\mathsf{G}(i,j) + (1-\lambda)\mathsf{G}'(i,j),$$

are joint observables of A and B. This leads to the conclusion that A and B either have a unique joint observable or uncountably many different joint observables. The case under investigation falls, luckily, into the first class. This is essential for our investigation as it crucially limits the search for optimal joint instruments.

To see that G is a unique joint observable for A and B, we first notice that any joint observable G' for A and B is completely determined by a single operator, say $\mathsf{G}'(1,1)$. The other operators are then recovered from the marginal conditions (2.1). Since $\mathsf{G}'(1,1)$ is a positive operator, we can write it as

$$\mathsf{G}'(1,1) = \tfrac{1}{2}(\gamma I + \mathbf{g}\cdot\boldsymbol{\sigma}),$$

[1]Here we make use of the usual notation that $\mathsf{A} \geqslant \mathsf{B}$ stands for the operator $\mathsf{A} - \mathsf{B}$ being positive.

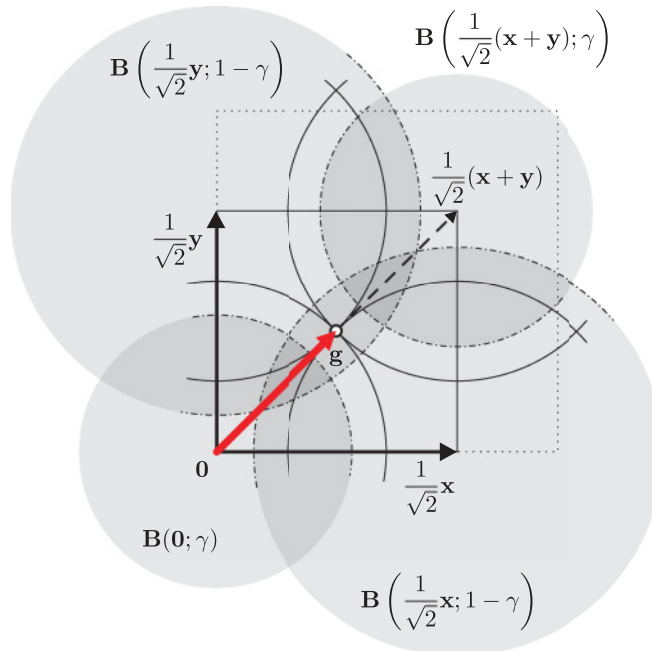

FIG. 2. (Color online) Visual representation of the four-ball intersection determining the existence of joint observable for A and B parametrized by γ and vector $\mathbf{g}$. If $\gamma = 1/2$, all four balls intersect in a single point (white circle) defining the joint observable (solid arcs). If γ differs from $1/2$ (e.g., $\gamma = 0.4$ as in the figure), two opposite circles intersect, while the other two do not; in such a case there is no joint observable with given γ (dot-dashed arcs).

where $\gamma \geqslant 0$ and $\mathbf{g} \in \mathbb{R}^3$. As noticed in [1], the conditions for $\mathsf{G}'(1,1)$ to define a joint observable for A and B are the following operator inequalities:

$$O \leqslant \mathsf{G}'(1,1) \leqslant \mathsf{A}(1),$$
$$\mathsf{A}(1) + \mathsf{B}(1) - I \leqslant \mathsf{G}'(1,1) \leqslant \mathsf{B}(1).$$

These are equivalent to the requirement that the vector $\mathbf{g}$ is in the intersection of four balls:

$$\mathbf{g} \in \mathbf{B}(\mathbf{0};\gamma) \cap \mathbf{B}\left(\frac{1}{\sqrt{2}}\mathbf{x}; 1-\gamma\right)$$
$$\cap\, \mathbf{B}\left(\frac{1}{\sqrt{2}}\mathbf{y}; 1-\gamma\right) \cap \mathbf{B}\left(\frac{1}{\sqrt{2}}(\mathbf{x}+\mathbf{y}); \gamma\right).$$

Since $\mathbf{x}$ and $\mathbf{y}$ are orthogonal unit vectors, the intersection is nonempty only if $\gamma = \frac{1}{2}$, and in that case $\mathbf{g} = \frac{1}{2\sqrt{2}}(\mathbf{x}+\mathbf{y})$; see Fig. 2. This means that $\mathsf{G}' = \mathsf{G}$, proving that there is only one joint observable for A and B.

It may be worth emphasizing that the joint observable of two observables is unique only in special cases. For instance, if instead of taking $\eta = \frac{1}{\sqrt{2}}$ we would have chosen a smaller number in the definition of A and B, then they would have infinitely many joint observables. This is as well evident from Fig. 2.

Let us denote by $\mu_{\pm\pm}(\varrho) \equiv \mathrm{tr}[\varrho\mathsf{G}(\pm 1, \pm 1)]$ the probabilities observed in the G measurement. If a state ϱ is written as $\varrho = \frac{1}{2}(I + \mathbf{r}\cdot\boldsymbol{\sigma})$, then

$$\mu_{jk}(\varrho) = \frac{1}{4}\left[1 + \frac{1}{\sqrt{2}}(j\mathbf{r}\cdot\mathbf{x} + k\mathbf{r}\cdot\mathbf{y})\right], \quad j,k = \pm.$$

It is now straightforward to see that

$$\mu_{+\mp}(\varrho) + \mu_{-\pm}(\varrho) = \tfrac{1}{2}$$

for all states ϱ. This implies that the whole probability distribution is actually determined only by two numbers [e.g., $\mu_{++}(\varrho)$ and $\mu_{+-}(\varrho)$]. We further notice that the numbers $\mu_{++}(\varrho)$ and $\mu_{+-}(\varrho)$ satisfy

$$\left(\mu_{++}(\varrho) - \frac{1}{4}\right)^2 + \left(\mu_{+-}(\varrho) - \frac{1}{4}\right)^2 \leqslant \left(\frac{\|\mathbf{r}\|}{4}\right)^2.$$

This inequality characterizes the convex set of all possible probability distributions in the range of G.

IV. JOINT INSTRUMENT

An instrument $\mathcal{G}$ implementing the joint observable G consists of four operations $\mathcal{G}_{jk}$, $j,k = \pm$, satisfying

$$\mathrm{tr}[\mathcal{G}_{jk}(\varrho)] = \mathrm{tr}[\varrho \mathsf{G}(j,k)].$$

Due to the simple structure of G, we can characterize all its instruments $\mathcal{G}$ in an uncomplicated way. Let us, for a moment, concentrate on $\mathcal{G}_{++}$, an operation associated with the outcome combination (1,1).

We first observe that $2\mathsf{G}(1,1) =: R$ is a one-dimensional projection. Let $\{K_\ell\}$ be the set of Kraus operators for $\mathcal{G}_{++}$, so that

$$\mathcal{G}_{++}(\varrho) = \sum_\ell K_\ell \varrho K_\ell^*,$$

$$\sum_\ell K_\ell^* K_\ell = \mathsf{G}(1,1) = \frac{1}{2} R.$$

The last equation implies that for each ℓ, we have $2K_\ell^* K_\ell \leqslant R$. Since R is a one-dimensional projection, there is a number $0 < k_\ell \leqslant 1$ such that $2K_\ell^* K_\ell = k_\ell R$. Clearly, $\sum_\ell k_\ell = 1$. Let $K_\ell = U_\ell |K_\ell|$ be the polar decomposition of K_ℓ. Here, U_ℓ is a unitary operator and

$$|K_\ell| = \sqrt{K_\ell^* K_\ell} = \sqrt{k_\ell/2} R.$$

For every state ϱ, we then get

$$\begin{aligned} K_\ell \varrho K_\ell^* &= \tfrac{1}{2} k_\ell U_\ell R \varrho R U_\ell^* = \tfrac{1}{2} k_\ell \mathrm{tr}[\varrho R] U_\ell R U_\ell^* \\ &= \mathrm{tr}[\varrho \mathsf{G}(1,1)] k_\ell U_\ell R U_\ell^*, \end{aligned}$$

and hence

$$\mathcal{G}_{++}(\varrho) = \mathrm{tr}[\varrho \mathsf{G}(1,1)] \sum_\ell k_\ell U_\ell R U_\ell^*.$$

Each $U_\ell R U_\ell^*$ is a one-dimensional projection and the convex sum,

$$\sum_\ell k_\ell U_\ell R U_\ell^* \equiv \xi_{++},$$

is therefore a state.

A similar calculation can be performed for the other three operations separately. Hence, we conclude that an instrument $\mathcal{G}$ implementing G is determined by four states $\xi_{\pm\pm}$, and the corresponding operations are given by

$$\mathcal{G}_{\pm\pm}(\varrho) = \mathrm{tr}[\varrho \mathsf{G}(\pm 1, \pm 1)] \xi_{\pm\pm} = \mu_{\pm\pm}(\varrho) \xi_{\pm\pm}. \tag{4.1}$$

As we have seen, this simple structure of the instruments implementing G is due to the fact that each element $\mathsf{G}(i,j)$ is a rank-1 operator.

Finally, let us emphasize that the probabilities $\mu_{\pm\pm}(\varrho)$ are fixed since $\mathcal{G}$ implements the observable G. The freedom we have is only in the choice of the four states $\xi_{\pm\pm}$.

V. OPTIMAL APPROXIMATION

A. Distance between operations

We are seeking for the best simultaneous approximation to the von Neumann instruments associated with X and Y. We therefore perform a measurement of G, which is described by an instrument $\mathcal{G}$ of the form (4.1).

In a similar way as G gives A and B as its marginals, $\mathcal{G}$ determines marginal instruments $\mathcal{A}$ and $\mathcal{B}$. Our aim is that the following approximations should be as close as possible:

$$\begin{aligned} \mathcal{A}_+ &\equiv \mathcal{G}_{++} + \mathcal{G}_{+-} \sim \mathcal{X}_+, \\ \mathcal{B}_+ &\equiv \mathcal{G}_{++} + \mathcal{G}_{-+} \sim \mathcal{Y}_+, \\ \mathcal{A}_- &\equiv \mathcal{G}_{--} + \mathcal{G}_{-+} \sim \mathcal{X}_-, \\ \mathcal{B}_- &\equiv \mathcal{G}_{--} + \mathcal{G}_{+-} \sim \mathcal{Y}_-. \end{aligned}$$

As we are already using the unique joint observable G of the optimal approximating observables A and B, the measurement outcome probabilities are set and do not depend on the choice of $\mathcal{G}$. Therefore, in order to quantify the distance between a given approximation and the corresponding von Neumann instrument, it is enough to compare the normalized output states.

There are various options for how to quantify the distance between Hilbert space operators. However, when considering the distance between density operators it is natural to choose the one induced by the trace norm. Operationally, it quantifies the optimal probability with which the states can be discriminated in a single run of the experiment (i.e., by observing a single experimental click [22]).

The distance exhibiting the difference between the output states for a given pair of operations can be utilized to induce a distance between the instruments. In particular, in what follows we will analyze the average distance over all input states (Sec. V B) and the worst-case distance (Sec. V C). Our interest is to minimize their values for all outcomes.

If we measure X and obtain the outcome 1, then the output state is $\mathsf{X}(1)$. On the other hand, if we measure A and obtain the outcome 1, then the output state is

$$\begin{aligned} \frac{\mathcal{A}_+(\varrho)}{\mathrm{tr}[\mathcal{A}_+(\varrho)]} &= \frac{\mu_{++}(\varrho)}{\mu_{++}(\varrho) + \mu_{+-}(\varrho)} \xi_{++} \\ &\quad + \frac{\mu_{+-}(\varrho)}{\mu_{++}(\varrho) + \mu_{+-}(\varrho)} \xi_{+-}. \end{aligned}$$

The trace distance of the approximation $\mathcal{A}_+$ from the desired operation $\mathcal{X}_+$, given that the input state is ϱ, is thus

$$d_+^{\mathsf{A},\mathsf{X}}(\varrho) := \left\| \frac{\mathcal{A}_+(\varrho)}{\mathrm{tr}[\mathcal{A}_+(\varrho)]} - \mathsf{X}(1) \right\|,$$

where the norm on the right-hand side is the trace norm.

B. Optimal approximations under average distance

Assigning Bloch vectors $\mathbf{q}_{\pm\pm} \in \mathbb{R}^3$ to the states $\xi_{\pm\pm}$, the distance $d_+^{\mathsf{A},\mathsf{X}}(\varrho)$ can be written in the form,

$$d_+^{\mathsf{A},\mathsf{X}}(\varrho) = \left\| \frac{\mathbf{q}_{++}+\mathbf{q}_{+-}}{2} + f(\varrho)\frac{\mathbf{q}_{++}-\mathbf{q}_{+-}}{2} - \mathbf{x} \right\|, \quad (5.1)$$

where $\mathbf{r} = (r_x, r_y, r_z)$ is the Bloch vector corresponding to ϱ, $f(\varrho) = f(r_x, r_y) = r_y/(\sqrt{2}+r_x)$ and the norm is the Euclidean norm in $\mathbb{R}^3$. Similarly, we get

$$d_+^{\mathsf{B},\mathsf{Y}}(\varrho) = \left\| \frac{\mathbf{q}_{++}+\mathbf{q}_{-+}}{2} + f(\varrho)\frac{\mathbf{q}_{++}-\mathbf{q}_{-+}}{2} - \mathbf{y} \right\|$$
$$d_-^{\mathsf{A},\mathsf{X}}(\varrho) = \left\| \frac{\mathbf{q}_{--}+\mathbf{q}_{-+}}{2} + f(\varrho)\frac{\mathbf{q}_{--}-\mathbf{q}_{-+}}{2} + \mathbf{x} \right\|$$
$$d_-^{\mathsf{B},\mathsf{Y}}(\varrho) = \left\| \frac{\mathbf{q}_{--}+\mathbf{q}_{+-}}{2} + f(\varrho)\frac{\mathbf{q}_{--}-\mathbf{q}_{+-}}{2} + \mathbf{y} \right\|.$$

Since both the vectors $\mathbf{x},\mathbf{y}$ have a vanishing z component, it follows that setting the z component of our choice of Bloch vectors $\mathbf{q}_{\pm\pm}$ to zero decreases the distances. Hence, in optimization tasks we can restrict ourselves to vectors $\mathbf{q}_{\pm\pm}$ with the vanishing z component.

We denote by $\langle\ \cdot\ \rangle_\varrho$ the normalized integration over the Bloch ball $\mathbb{B} = \{\mathbf{r} \in \mathbb{R}^3 : \|\mathbf{r}\|^2 = r_x^2 + r_y^2 + r_z^2 \leqslant 1\}$ representing the state space of a qubit. Hence, for a function F defined on the state space we have

$$\langle F(\varrho)\rangle_\varrho = \frac{3}{4\pi}\int_{\mathbb{B}} F[\rho(\mathbf{r})]dr_x dr_y dr_z.$$

We now define (for the outcome $\mathcal{A}_+$) the average distance to be[2]

$$\mathfrak{D}_+^{\mathsf{A},\mathsf{X}} := \langle [d_+^{\mathsf{A},\mathsf{X}}(\varrho)]^2\rangle_\varrho.$$

We have chosen $[d_+^{\mathsf{A},\mathsf{X}}(\varrho)]^2$ instead of $d_+^{\mathsf{A},\mathsf{X}}(\varrho)$ just to simplify the calculations.

We can write (5.1) as

$$[d_+^{\mathsf{A},\mathsf{X}}(\varrho)]^2 = \tfrac{1}{4}\|\mathbf{q}_{++}+\mathbf{q}_{+-}-2\mathbf{x}\|^2 + \tfrac{1}{4}f(\varrho)^2\|\mathbf{q}_{++}-\mathbf{q}_{+-}\|^2 + \tfrac{1}{2}f(\varrho)(\mathbf{q}_{++}+\mathbf{q}_{+-}-2\mathbf{x})\cdot(\mathbf{q}_{++}-\mathbf{q}_{+-}).$$

As $\langle f(\varrho)\rangle_\varrho = 0$ and

$$\langle f(\varrho)^2\rangle_\varrho = 2 - \frac{3\sqrt{2}}{2}\ln(1+\sqrt{2}) \equiv \alpha,$$

the average distance $\mathfrak{D}_+^{\mathsf{A},\mathsf{X}}$ is expressible in the form,

$$\mathfrak{D}_+^{\mathsf{A},\mathsf{X}} = \frac{1}{4}\|\mathbf{q}_{++}+\mathbf{q}_{+-}-2\mathbf{x}\|^2 + \frac{\alpha}{4}\|\mathbf{q}_{++}-\mathbf{q}_{+-}\|^2.$$

Analogously, we get

$$\mathfrak{D}_+^{\mathsf{B},\mathsf{Y}} = \frac{1}{4}\|\mathbf{q}_{++}+\mathbf{q}_{-+}-2\mathbf{y}\|^2 + \frac{\alpha}{4}\|\mathbf{q}_{++}-\mathbf{q}_{-+}\|^2,$$
$$\mathfrak{D}_-^{\mathsf{A},\mathsf{X}} = \frac{1}{4}\|\mathbf{q}_{--}+\mathbf{q}_{-+}+2\mathbf{x}\|^2 + \frac{\alpha}{4}\|\mathbf{q}_{--}-\mathbf{q}_{-+}\|^2,$$
$$\mathfrak{D}_-^{\mathsf{B},\mathsf{Y}} = \frac{1}{4}\|\mathbf{q}_{--}+\mathbf{q}_{+-}+2\mathbf{y}\|^2 + \frac{\alpha}{4}\|\mathbf{q}_{--}-\mathbf{q}_{+-}\|^2.$$

The distance $\mathfrak{D}_+^{\mathsf{A},\mathsf{X}}$ can be made zero by taking $\mathbf{q}_{++} = \mathbf{q}_{+-} = \pm\mathbf{x}$. This gives also $\mathfrak{D}_-^{\mathsf{A},\mathsf{X}} = 0$. But we will then have $\mathfrak{D}_\pm^{\mathsf{B},\mathsf{Y}} = 1+\alpha$ and the sum is then $\mathfrak{D} = \mathfrak{D}_+^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_-^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_+^{\mathsf{B},\mathsf{Y}} + \mathfrak{D}_-^{\mathsf{B},\mathsf{Y}} = 2(1+\alpha)$. As we shall see this choice does not achieve the minimal value of the sum $\mathfrak{D}$.

We want to make all the four distances as small as possible, under the condition that they are equal. To find the optimal instrument, we consider the sum $\mathfrak{D} = \mathfrak{D}_+^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_+^{\mathsf{B},\mathsf{Y}} + \mathfrak{D}_-^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_-^{\mathsf{B},\mathsf{Y}}$, which is obviously a convex function of vectors $\mathbf{q}_{\pm\pm}$. Thus, its minimization subject to the conditions $\|\mathbf{q}_{\pm\pm}\| \leqslant 1$ is a convex optimization problem. Since $\mathfrak{D}$ is differentiable, the optimality criterion (see, e.g., [23]) for an instrument defined by a quadruple $\bar{\mathbf{q}} = (\mathbf{q}_{++}, \mathbf{q}_{+-}, \mathbf{q}_{-+}, \mathbf{q}_{--})$ is that

$$\nabla\mathfrak{D}(\bar{\mathbf{q}})^T\cdot\bar{\mathbf{p}} \geqslant \nabla\mathfrak{D}(\bar{\mathbf{q}})^T\cdot\bar{\mathbf{q}} \quad (5.2)$$

for all $\bar{\mathbf{p}} = (\mathbf{p}_{++}, \mathbf{p}_{+-}, \mathbf{p}_{-+}, \mathbf{p}_{--})$ with $\|\mathbf{p}_{\pm\pm}\| \leqslant 1$. It is now easy to verify that the optimal solution is achieved when $\mathbf{q}_{\pm\pm} = (\pm\mathbf{x}\pm\mathbf{y})/\sqrt{2}$ and this choice gives $\mathfrak{D}_+^{\mathsf{A},\mathsf{X}} = \mathfrak{D}_+^{\mathsf{B},\mathsf{Y}} = \mathfrak{D}_-^{\mathsf{A},\mathsf{X}} = \mathfrak{D}_-^{\mathsf{B},\mathsf{Y}} = \frac{1}{2}(3-2\sqrt{2}+\alpha)$. We can also compare this solution with the example where one direction was preferred, and we see that

$$\mathfrak{D}_{\text{opt}} = 2(3-2\sqrt{2}+\alpha) < 2(1+\alpha).$$

The optimal joint instrument corresponds to the pure states $\xi_{\pm\pm}^{\text{ave}} = 2\mathsf{G}(\pm1,\pm1)$, and we recognize it being the Lüders instrument $\mathcal{L}^{\mathsf{G}}$ of G, given as

$$\mathcal{L}_{\pm\pm}^{\mathsf{G}}(\varrho) = \sqrt{\mathsf{G}(\pm1,\pm1)}\varrho\sqrt{\mathsf{G}(\pm1,\pm1)} = \text{tr}[\mathsf{G}(\pm1,\pm1)\varrho]\xi_{\pm,\pm}^{\text{ave}}. \quad (5.3)$$

In summary, we have found that the Lüders instrument $\mathcal{L}^{\mathsf{G}}$ of G gives the optimal joint approximation of $\mathcal{X}$ and $\mathcal{Y}$, the quality of approximations being quantified using the average distance.

C. Optimal approximations under the worst-case distance

The optimal joint instrument naturally depends on the quantification of the distance between two operations. We believe that the average norm studied in Sec. V B is the most relevant way to measure the distance.

If our task was to discriminate the given pair of instruments, the average norm would then quantify the average success probability under supposition of choosing the test state ρ randomly. As a comparison we take a look also on the worst-case distance, which determines test states allowing the best possible discrimination of the two instruments. In this sense the worst-case distance optimizes the distinguishability over the test states. The worst-case distance is defined as

$$\mathfrak{W}_+^{\mathsf{A},\mathsf{X}} := \sup_\varrho d_+^{\mathsf{A},\mathsf{X}}(\varrho).$$

We want, again, to find an instrument which minimizes these distances under the condition that they are all equal.

[2]Strictly speaking, we should average over states ϱ with $\text{tr}[\varrho\mathsf{X}(1)] \neq 0$. However, the complement set has measure zero and does not therefore affect our calculations.

Let us observe that the Bloch vector of a normalized outcome state $\mathcal{A}_+(\varrho)/\mathrm{tr}[\mathcal{A}_+(\varrho)]$ is of the form,

$$\tfrac{1}{2}[1+f(\varrho)]\mathbf{q}_{++}+\tfrac{1}{2}[1-f(\varrho)]\mathbf{q}_{+-}. \tag{5.4}$$

Therefore, the distance $d_+^{\mathsf{A},\mathsf{X}}(\varrho)$ is the length of a vector being a convex combination of vectors $\mathbf{q}_{++}-\mathbf{x}$ and $\mathbf{q}_{+-}-\mathbf{x}$. We thus conclude that

$$\mathfrak{W}_+^{\mathsf{A},\mathsf{X}}=\max\{\|\mathbf{q}_{++}-\mathbf{x}\|,\|\mathbf{q}_{+-}-\mathbf{x}\|\}. \tag{5.5}$$

Similarly, we get

$$\begin{aligned}\mathfrak{W}_+^{\mathsf{B},\mathsf{Y}}&=\max\{\|\mathbf{q}_{++}-\mathbf{y}\|,\|\mathbf{q}_{-+}-\mathbf{y}\|\},\\ \mathfrak{W}_-^{\mathsf{A},\mathsf{X}}&=\max\{\|\mathbf{q}_{--}+\mathbf{x}\|,\|\mathbf{q}_{-+}+\mathbf{x}\|\},\\ \mathfrak{W}_-^{\mathsf{B},\mathsf{Y}}&=\max\{\|\mathbf{q}_{--}+\mathbf{y}\|,\|\mathbf{q}_{+-}+\mathbf{y}\|\}.\end{aligned}$$

The distance $\mathfrak{W}_+^{\mathsf{A},\mathsf{X}}$ in (5.5) can be made zero, and this happens if and only if $\mathbf{q}_{++}=\mathbf{q}_{+-}=\mathbf{x}$. If $\mathbf{q}_{++}$ and $\mathbf{q}_{+-}$ are chosen in this way we are still free to choose $\mathbf{q}_{-+}$ and $\mathbf{q}_{--}$, hence we can also achieve $\mathfrak{W}_-^{\mathsf{A},\mathsf{X}}=0$ by taking $\mathbf{q}_{--}=\mathbf{q}_{-+}=-\mathbf{x}$. However, we will then have $\mathfrak{W}_\pm^{\mathsf{B},\mathsf{Y}}=\sqrt{2}$ and $\mathfrak{W}=\mathfrak{W}_+^{\mathsf{A},\mathsf{X}}+\mathfrak{W}_+^{\mathsf{B},\mathsf{Y}}+\mathfrak{W}_-^{\mathsf{A},\mathsf{X}}+\mathfrak{W}_-^{\mathsf{B},\mathsf{Y}}=2\sqrt{2}$.

The value $\mathfrak{W}=2\sqrt{2}$ for the sum is achieved also for the symmetric choice $\mathbf{q}_{\pm\pm}=(\pm\mathbf{x}\pm\mathbf{y})/2$. In this case all the distances are equal, $\mathfrak{W}_+^{\mathsf{A},\mathsf{X}}=\mathfrak{W}_+^{\mathsf{B},\mathsf{Y}}=\mathfrak{W}_-^{\mathsf{A},\mathsf{X}}=\mathfrak{W}_-^{\mathsf{B},\mathsf{Y}}=1/\sqrt{2}$.

Let us then minimize the distances under the condition that they are all equal. First, we require that $\mathfrak{W}_+^{\mathsf{A},\mathsf{X}}=\mathfrak{W}_+^{\mathsf{B},\mathsf{Y}}$ and we minimize these two distances, ignoring for a moment the other two distances. It is easy to see that in the optimal case it is necessary to choose $\mathbf{q}_{++}=(\mathbf{x}+\mathbf{y})/2$. Similarly, if we require that $\mathfrak{W}_-^{\mathsf{A},\mathsf{X}}=\mathfrak{W}_-^{\mathsf{B},\mathsf{Y}}$ and we minimize these two distances independently of the previous minimization, we see that it is necessary to put $\mathbf{q}_{--}=(-\mathbf{x}-\mathbf{y})/2$. These two optimal choices are possible simultaneously only if we set $\mathbf{q}_{+-}=(\mathbf{x}-\mathbf{y})/2$ and $\mathbf{q}_{-+}=(-\mathbf{x}+\mathbf{y})/2$. In conclusion, the symmetric choice $\mathbf{q}_{\pm\pm}=(\pm\mathbf{x}\pm\mathbf{y})/2$ is optimal.

The instrument $\mathcal{G}$ corresponding to this symmetric choice of $\xi_{\pm\pm}$ can be written as

$$\begin{aligned}\mathcal{G}_{\pm\pm}(\varrho)&=\frac{1}{\sqrt{2}}\mathcal{L}^{\mathsf{G}}_{\pm,\pm}(\varrho)+\left(1-\frac{1}{\sqrt{2}}\right)\mathrm{tr}[\varrho\mathsf{G}(\pm1,\pm1)]\tfrac{1}{2}I\\ &=\mathrm{tr}[\varrho\mathsf{G}(\pm1,\pm1)]\left[\frac{1}{\sqrt{2}}\xi^{\mathrm{ave}}_{\pm\pm}+\left(1-\frac{1}{\sqrt{2}}\right)\tfrac{1}{2}I\right]\\ &=\mathrm{tr}[\varrho\mathsf{G}(\pm1,\pm1)]\xi^{\mathrm{w.c.}}_{\pm\pm}.\end{aligned} \tag{5.6}$$

Hence, $\mathcal{G}$ is a mixture of the Lüders instrument $\mathcal{L}^{\mathsf{G}}$ and another instrument, which has very simple form. In particular, $\mathcal{G}$ can be realized by using the Lüders instrument $\mathcal{L}^{\mathsf{G}}$, accepting the measurement outcomes but ignoring the output state in $1-\frac{1}{\sqrt{2}}$ parts of the measurement and preparing the maximally mixed state $\frac{1}{2}I$ in these cases.

Previous results can be visualized as given in Fig. 3. In the case of the average distance [horizontal blue line represents possible normalized output states as a manifestation of (5.4)] we find it is natural to expect as large Bloch vectors $\mathbf{q}_{\pm\pm}$ as possible, but for the worst-case distance the additional noise in the normalized output state makes it closer to the desired state (vertical red line with red dots representing worst cases).

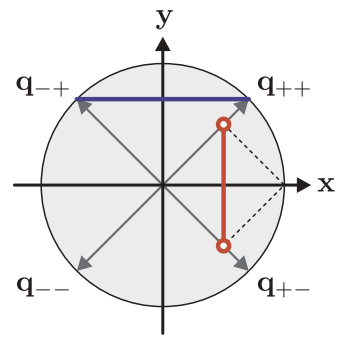

FIG. 3. (Color online) Bloch vector representation of instrument $\mathcal{G}$ gives us in the case of approximation $\mathcal{B}_+$ the resulting normalized output states lying on a (thick horizontal) blue line. The average distance is minimized for the choice of symmetrical pure states $\xi_{\pm\pm}$. In the case of approximation $\mathcal{A}_+$ the normalized output states are represented by (thick vertical) red line with red dots representing the worst cases. These states are closest to $\mathbf{x}$ when the output state is noisier than in the average case.

VI. APPROXIMATE JOINT MEASUREMENT OF THREE VON NEUMANN MEASUREMENTS

Let $\mathbf{z}$ be a unit vector which is orthogonal to both $\mathbf{x}$ and $\mathbf{y}$, and let Z be the corresponding sharp observable. Suppose we want to approximate the von Neumann instruments of X, Y, and Z. This problem of additional measurement bears some differences with the previously studied approximation task of two von Neumann measurements.

The optimal jointly measurable approximations of X, Y, and Z are given by

$$\begin{aligned}\mathsf{A}(\pm1)&=\tfrac{1}{2}(I\pm\eta\sigma_{\mathbf{x}}),\\ \mathsf{B}(\pm1)&=\tfrac{1}{2}(I\pm\eta\sigma_{\mathbf{y}}),\\ \mathsf{C}(\pm1)&=\tfrac{1}{2}(I\pm\eta\sigma_{\mathbf{z}}),\end{aligned}$$

with $\eta=\frac{1}{\sqrt{3}}$. The observables A and B are of similar form as before, but we have to decrease η from $\frac{1}{\sqrt{2}}$ to $\frac{1}{\sqrt{3}}$ to make it possible to include the additional spin component direction. It has been proved in [24] that for $\eta>\frac{1}{\sqrt{3}}$ the three observables are not jointly measurable.

Generally, a joint observable E for A, B, and C has eight outcomes. An observable defined as

$$\mathsf{E}(\pm1,\pm1,\pm1)=\frac{1}{8}\left[I+\frac{1}{\sqrt{3}}(\pm\sigma_{\mathbf{x}}\pm\sigma_{\mathbf{y}}\pm\sigma_{\mathbf{z}})\right]$$

is a joint observable for A, B, and C since it satisfies the marginal conditions,

$$\begin{aligned}\mathsf{A}(\pm1)&=\sum_{j,k\in\{-1,1\}}\mathsf{E}(\pm1,j,k),\\ \mathsf{B}(\pm1)&=\sum_{j,k\in\{-1,1\}}\mathsf{E}(j,\pm1,k),\\ \mathsf{C}(\pm1)&=\sum_{j,k\in\{-1,1\}}\mathsf{E}(j,k,\pm1).\end{aligned}$$

Unlike in the earlier situation, now we have several different joint observables. Another joint observable F for A, B, and C is given by

$$\begin{aligned}\mathsf{F}(1,1,1) &= 2\,\mathsf{E}(1,1,1),\\ \mathsf{F}(1,-1,-1) &= 2\,\mathsf{E}(1,-1,-1),\\ \mathsf{F}(-1,-1,1) &= 2\,\mathsf{E}(-1,-1,1),\\ \mathsf{F}(-1,1,-1) &= 2\,\mathsf{E}(-1,1,-1),\\ \mathsf{F}(1,1,-1) &= \mathsf{F}(1,-1,1) = \mathsf{F}(-1,1,1)\\ &= \mathsf{F}(-1,-1,-1) = 0.\end{aligned}$$

A notable feature of the observable F is that it is essentially a four-outcome observable. Although A, B, and C have infinitely many different joint observables, F is the unique joint observable having only four nonzero elements. To demonstrate this fact, suppose that F' is a joint observable of A, B, and C, and that

$$\begin{aligned}\mathsf{F}'(1,1,-1) &= \mathsf{F}'(1,-1,1) = \mathsf{F}'(-1,1,1)\\ &= \mathsf{F}'(-1,-1,-1) = 0.\end{aligned}$$

First of all, let us notice that F' is completely determined by a single nonzero element, say $\mathsf{F}'(1,1,1)$. The other operators are given by the marginal conditions. For instance, $\mathsf{F}'(1,-1-1) = \mathsf{A}(1) - \mathsf{F}'(1,1,1)$. Since F' is an observable, the operators $\mathsf{F}'(i,j,k)$ must sum up to identity. Hence, we get

$$\begin{aligned}I &= \mathsf{F}'(1,1,1) + \mathsf{F}'(1,-1,-1) + \mathsf{F}'(-1,1,-1)\\ &\quad + \mathsf{F}'(-1,-1,1)\\ &= \mathsf{F}'(1,1,1) + [\mathsf{A}(1) - \mathsf{F}'(1,1,1)] + [\mathsf{B}(1) - \mathsf{F}'(1,1,1)]\\ &\quad + [\mathsf{C}(1) - \mathsf{F}'(1,1,1)].\end{aligned}$$

Therefore, the operator $\mathsf{F}'(1,1,1)$ is determined by the equation,

$$\mathsf{F}'(1,1,1) = \tfrac{1}{2}[\mathsf{A}(1) + \mathsf{B}(1) + \mathsf{C}(1) - I].$$

It follows that $\mathsf{F}' = \mathsf{F}$, and the four-outcome joint observable is hence unique.

As done previously for the joint observable G, we can now study the instruments implementing E and F. The elements forming E and F are rank-1 operators. Therefore, our characterization for instruments implementing G in Sec. IV applies to these two observables as well.

An instrument implementing E is determined by eight states $\zeta_{\pm,\pm,\pm}$ and the corresponding operations are

$$\mathcal{E}_{\pm\pm\pm}(\varrho) = \mathrm{tr}[\varrho\,\mathsf{E}(\pm 1, \pm 1, \pm 1)]\zeta_{\pm\pm\pm}.$$

The approximations to von Neumann instruments $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ are now defined as follows:

$$\begin{aligned}\mathcal{A}_\pm &\equiv \mathcal{E}_{\pm++} + \mathcal{E}_{\pm-+} + \mathcal{E}_{\pm--} + \mathcal{E}_{\pm+-} \sim \mathcal{X}_\pm,\\ \mathcal{B}_\pm &\equiv \mathcal{E}_{+\pm+} + \mathcal{E}_{-\pm+} + \mathcal{E}_{+\pm-} + \mathcal{E}_{-\pm-} \sim \mathcal{Y}_\pm,\\ \mathcal{C}_\pm &\equiv \mathcal{E}_{++\pm} + \mathcal{E}_{--\pm} + \mathcal{E}_{+-\pm} + \mathcal{E}_{-+\pm} \sim \mathcal{Z}_\pm.\end{aligned}$$

The optimal instrument under the average distance can be deduced by following a similar procedure as the one presented in Sec. V B. We thus determine the minimum of the sum $\mathfrak{D} = \mathfrak{D}_+^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_+^{\mathsf{B},\mathsf{Y}} + \mathfrak{D}_+^{\mathsf{C},\mathsf{Z}} + \mathfrak{D}_-^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_-^{\mathsf{B},\mathsf{Y}} + \mathfrak{D}_-^{\mathsf{C},\mathsf{Z}}$ where now, for instance, the distance $\mathfrak{D}_+^{\mathsf{A},\mathsf{X}}$ is

$$\begin{aligned}\mathfrak{D}_+^{\mathsf{A},\mathsf{X}} &= \frac{1}{16}\,\|\mathbf{q}_{+++} + \mathbf{q}_{+-+} + \mathbf{q}_{+--} + \mathbf{q}_{++-} - 4\mathbf{x}\|^2\\ &\quad + \frac{\beta}{16}\,\|\mathbf{q}_{+++} - \mathbf{q}_{+-+} - \mathbf{q}_{+--} + \mathbf{q}_{++-}\|^2\\ &\quad + \frac{\beta}{16}\,\|\mathbf{q}_{+++} + \mathbf{q}_{+-+} - \mathbf{q}_{+--} - \mathbf{q}_{++-}\|^2,\end{aligned}$$

and

$$\beta \equiv \left\langle \left(\frac{r_y}{\sqrt{3} + r_x}\right)^2 \right\rangle_\varrho = \frac{1}{2}[7 - 3\sqrt{3}\ln(2 + \sqrt{3})].$$

Using the criterion (5.2), it is straightforward to verify that the optimal solution is achieved when $\mathbf{q}_{\pm\pm\pm} = \frac{1}{\sqrt{3}}(\pm\mathbf{x} \pm \mathbf{y} \pm \mathbf{z})$. All the distances are then equal and take the value $\frac{2}{3}\beta + (1 - \frac{1}{\sqrt{3}})^2$. The related joint instrument is the Lüders instrument $\mathcal{L}^{\mathsf{E}}$ of E.

An instrument $\mathcal{F}$ implementing F is, again, determined by eight states. However, the states corresponding to the zero elements play no role, so $\mathcal{F}$ is actually determined by four states only. Since F gives different measurement outcome probabilities than E, the related average distances are also different. So, if we consider F, we get

$$\mathfrak{D}_+^{\mathsf{A},\mathsf{X}} = \frac{1}{4}\,\|\mathbf{q}_{+++} + \mathbf{q}_{+--} - 2\mathbf{x}\|^2 + \frac{\gamma}{4}\,\|\mathbf{q}_{+++} - \mathbf{q}_{+--}\|^2,$$

and

$$\gamma \equiv \left\langle \left(\frac{r_y + r_z}{\sqrt{3} + r_x}\right)^2 \right\rangle_\varrho = 2\beta.$$

The sum $\mathfrak{D} = \mathfrak{D}_+^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_+^{\mathsf{B},\mathsf{Y}} + \mathfrak{D}_+^{\mathsf{C},\mathsf{Z}} + \mathfrak{D}_-^{\mathsf{A},\mathsf{X}} + \mathfrak{D}_-^{\mathsf{B},\mathsf{Y}} + \mathfrak{D}_-^{\mathsf{C},\mathsf{Z}}$ achieves its minimum when the four relevant Bloch vectors are

$$\begin{aligned}\mathbf{q}_{+++} &= (\mathbf{x} + \mathbf{y} + \mathbf{z})/\sqrt{3},\\ \mathbf{q}_{+--} &= (\mathbf{x} - \mathbf{y} - \mathbf{z})/\sqrt{3},\\ \mathbf{q}_{-+-} &= (-\mathbf{x} + \mathbf{y} - \mathbf{z})/\sqrt{3},\\ \mathbf{q}_{--+} &= (-\mathbf{x} - \mathbf{y} + \mathbf{z})/\sqrt{3}.\end{aligned}$$

This choice gives $\mathfrak{D}_+^{\mathsf{A},\mathsf{X}} = \frac{2}{3}\gamma + (1 - \frac{1}{\sqrt{3}})^2$. The related joint instrument is the Lüders instrument $\mathcal{L}^{\mathsf{F}}$ of F.

We can, again, use Bloch representation and write the normalized output states similarly as in (5.4). For the four-outcome instrument we get a similar result—the normalized output states are convex combinations of the corresponding states defining the instrument. For instance, the normalized output state, when measuring along axis x and obtaining the outcome $+1$, is a convex combination of $\mathbf{q}_{+++}$ and $\mathbf{q}_{+--}$. Hence, these Bloch vectors of output states lie on the edges of a tetrahedron as depicted on Fig. 4 (thick red line).

For the eight-outcome instrument we find that the corresponding output states form a set,

$$\frac{1}{\sqrt{3}}(\mathbf{x} + \tilde{r}_y\mathbf{y} + \tilde{r}_z\mathbf{z}), \quad \tilde{r}_y^2 + \tilde{r}_z^2 \leqslant 1.$$

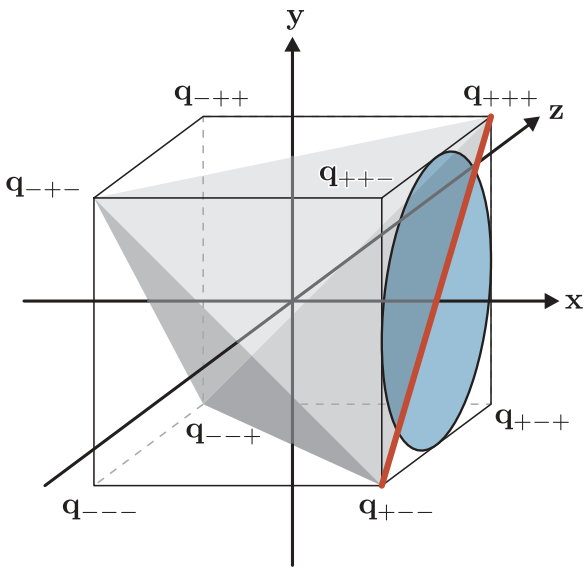

FIG. 4. (Color online) Bloch vector comparison of the four- and eight-outcome instruments. The four-outcome instrument leads to normalized output states from the edges of the tetrahedron (especially for the measurement along the x axis illustrated by the thick red diagonal). The eight-outcome instrument leads to the faces of the cube given by the vectors $\mathbf{q}_{\pm\pm\pm}$ (especially for the measurement along the x axis illustrated by the blue circle).

This set is a circle lying in the plane given by the vectors $\mathbf{q}_{+jk}$, $j,k \in \{-1,1\}$ [i.e., being inscribed into the face of the cube such as is depicted in Fig. 4 (blue circle)]. From this geometrical representation we can confirm the fact shown earlier; under the usage of the average distance the four-outcome instrument is worse that the eight-outcome instrument as it has a contribution from states being further from the reference state.

VII. CONCLUSIONS

The impossibility of joint measurements of orthogonal qubit measurements leads us to the study of their approximations. We have considered not only observables, but we have accessed the problem from the perspective of instruments. We have characterized the optimal approximation of two von Neumann instruments, and the optimal joint instrument was found to be the Lüders instrument of the optimal (unique) joint observable. This result [see Eq. (5.3)] was achieved by searching for the best approximation under the average distance between the normalized output states. When considering the worst-case norm, the resulting optimal instrument is a mixture of the corresponding Lüders instrument and the state-space contraction into the complete mixture [see Eq. (5.6)].

A similar investigation was performed in the case of three von Neumann instruments, but then we faced the problem of nonuniqueness of the optimal joint observable. Nevertheless, in the two commonly used instances the Lüders instrument was found to be optimal. Although the numerical value of the minimal average distance does not have any intrinsic meaning, one can use it in order to compare the minimal distances in the three studied cases: approximation of two von Neumann instruments and approximations of three von Neumann instruments with eight- and four-outcome measurements. The minimal average distances for these approximations are 0.15, 0.23, and 0.28, respectively. As one would expect, the average distance can be made lowest in the first case. Namely, it is certainly easier to approximate two von Neumann instruments rather than three. Furthermore, it is not surprising that in the latter two cases eight instead of four outcomes are more efficient in the approximation task. This may be explained by a broader set of outcomes to choose from, leaving less space for error.

We believe that Lüders instruments are optimal joint instruments for a more general class of situations than only those studied here. A natural extension of the two orthogonal qubit observables would be the case of two sharp observables related to mutually unbiased bases. Another interesting class is that of the continuous variable systems, where phase space observables play the role of joint observables. These generalizations merit further study.

ACKNOWLEDGMENTS

T.H. and M.A.J. acknowledge financial support from the Danish National Research Foundation Center for Quantum Optics (QUANTOP) and from the European Union projects COQUIT and QUEVADIS. D.R. and M.Z. acknowledge financial support from the European Union Project No. HIP FP7-ICT-2007-C-221889, and from Projects No. APVV-0673-07 QIAM, No. OP CE QUTE ITMS NFP 262401022, and No. CE-SAS QUTE. M.Z. also acknowledges support from Project No. MSM0021622419. The authors thank Peter Stano for useful comments.

[1] P. Busch, Phys. Rev. D **33**, 2253 (1986).
[2] P. Busch and C. Shilladay, Phys. Rep. **435**, 1 (2006).
[3] N.-L. Liu, L. Li, S. Yu, and Z.-B. Chen, Phys. Rev. A **79**, 052108 (2009).
[4] Y. Kurotani, T. Sagawa, and M. Ueda, Phys. Rev. A **76**, 022325 (2007).
[5] T. Sagawa and M. Ueda, Phys. Rev. A **77**, 012313 (2008).
[6] P. Busch and T. Heinosaari, Quantum Inf. Comput. **8**, 0797 (2008).
[7] T. Brougham, E. Andersson, and S. M. Barnett, Phys. Rev. A **80**, 042106 (2009).
[8] P. Stano, D. Reitzner, and T. Heinosaari, Phys. Rev. A **78**, 012315 (2008).
[9] S. Yu, N. Liu, L. Li, and C. H. Oh, e-print arXiv:0805.1538v1 [quant-ph].
[10] P. Busch and H.-J. Schmidt, Quant. Info. Proc. **9**, 143 (2010).
[11] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[12] E. Andersson, S. M. Barnett, and A. Aspect, Phys. Rev. A **72**, 042104 (2005).
[13] M. M. Wolf, D. Perez-Garcia, and C. Fernandez, Phys. Rev. Lett. **103**, 230402 (2009).

[14] A. Ferraro and M. G. A. Paris, Open Sys. & Information Dyn. **14**, 149 (2007).
[15] P. Busch and P. Lahti, Riv. Nuovo Cimento **18**, 1 (1995), e-print arXiv:quant-ph/0406132v1.
[16] A. S. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North-Holland Publishing, Amsterdam, 1982).
[17] P. Busch, M. Grabowski, and P. J. Lahti, *Operational Quantum Physics* (Springer-Verlag, Berlin, 1997), 2nd corrected printing.
[18] P. Busch, Found. Phys. **17**, 905 (1987).
[19] E. B. Davies, *Quantum Theory of Open Systems* (Academic Press, London, 1976).
[20] P. Busch and P. Lahti, in *Compendium of Quantum Physics*, edited by D. Greenberger, K. Hentschel, and F. Weinert (Springer, New York, 2009).
[21] T. Heinosaari, D. Reitzner, P. Stano, and M. Ziman, J. Phys. A **42**, 365302 (2009).
[22] C. W. Helstrom, *Quantum Detection and Estimation Theory* (Academic Press, New York, 1976).
[23] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
[24] T. Brougham and E. Andersson, Phys. Rev. A **76**, 052313 (2007).