Measurement uncertainty relations — **Source link** ⬈

Paul Busch, Pekka Lahti, Reinhard F. Werner

Related papers:

- Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik

- Proof of Heisenberg's Error-Disturbance Relation

- Colloquium : Quantum root-mean-square error and measurement uncertainty relations

- The Uncertainty Principle

- Zur Quantenmechanik einfacher Bewegungstypen

Measurement uncertainty relations

Paul Busch, Pekka Lahti, and Reinhard F. Werner

View Online   Export Citation   CrossMark

Measurement uncertainty relations

Paul Busch,[1,a)] Pekka Lahti,[2,b)] and Reinhard F. Werner[3,c)]

[1]*Department of Mathematics, University of York, York, United Kingdom*
[2]*Turku Centre for Quantum Physics, Department of Physics and Astronomy,*
University of Turku, FI-20014 Turku, Finland
[3]*Institut für Theoretische Physik, Leibniz Universität, Hannover, Germany*

Measurement uncertainty relations are quantitative bounds on the errors in an approximate joint measurement of two observables. They can be seen as a generalization of the error/disturbance tradeoff first discussed heuristically by Heisenberg. Here we prove such relations for the case of two canonically conjugate observables like position and momentum, and establish a close connection with the more familiar preparation uncertainty relations constraining the sharpness of the distributions of the two observables in the same state. Both sets of relations are generalized to means of order α rather than the usual quadratic means, and we show that the optimal constants are the same for preparation and for measurement uncertainty. The constants are determined numerically and compared with some bounds in the literature. In both cases, the near-saturation of the inequalities entails that the state (resp. observable) is uniformly close to a minimizing one. © *2014 AIP Publishing LLC.*
[http://dx.doi.org/10.1063/1.4871444]

I. INTRODUCTION

Following Heisenberg's ground-breaking paper[1] from 1927 uncertainty relations have become an indispensable tool of quantum mechanics. Often they are used in a qualitative heuristic way rather than in the form of proven inequalities. The paradigm of the latter kind is the Kennard-Robertson-Weyl inequality,[2–4] which is proved in virtually every quantum mechanics course. This relation shows that by turning the uncertainty relations into a theorem, one also reaches a higher level of conceptual precision. Heisenberg talks rather vaguely of quantities being "known" to some precision. With Kennard, Robertson, and Weyl, we are given a precise physical setting: a position and a momentum measurement applied to distinct instances of the same preparation (state), with the uncertainties interpreted as the root of the variances of the probability distributions obtained. However, this mathematical elucidation does not cover all quantitative aspects of uncertainty. There have been several papers in recent years formalizing and proving further instances of quantitative uncertainty. This is perhaps part of a trend which has become necessary as more and more experiments approach the uncertainty-limited regime. Moreover, uncertainty relations play an important role in some proofs of quantum cryptographic security. For a review of uncertainty up to 2006, we recommend to consult Ref. 5.

The Kennard-Robertson-Weyl inequality can be modified in several ways: On the one hand, we can stick to the same physical scenario, and apply different definitions of "spread" of a probability distribution. This is one of the routes taken in this paper: We replace the quadratic mean by one based on powers α and β for the position and the momentum distributions, respectively. One could go further here and introduce measures of "peakedness" for probability distributions such as entropies.[6]

---

a)Electronic mail: paul.busch@york.ac.uk
b)Electronic mail: pekka.lahti@utu.fi
c)Electronic mail: reinhard.werner@itp.uni-hannover.de

Our main goal, however, is a modification of the basic scenario, much closer to the original discussion of Heisenberg. In his discussion of the γ-ray microscope, the uncertainty relations concern the resolution of the microscope and the momentum kick imparted by the Compton scattering. Due to this momentum kick a momentum measurement after the observation gives different results from a direct momentum measurement. We generalize this scenario by taking the microscope with a subsequent momentum measurement as a "phase space measurement," i.e., as an observable which produces a position value and a momentum value in every single shot. Measurement uncertainty relations then constrain the accuracy of the marginal measurements: the resolution of the microscope is a benchmark parameter for the position output of the device, and the deviation from an ideal position measurement. Similarly, the disturbance is a quantity characterizing the accuracy of the momentum output, once again as compared to an ideal momentum measurement. Then, by definition, a measurement uncertainty relation is an inequality implying that these two error quantities cannot both be small. It has been disputed recently in a series of papers[7–11] that such a relation holds. We believe that this claim is based on badly chosen definitions of uncertainty, and we will give a detailed critique of these claims elsewhere.[12]

We offer two ways of quantifying the error quantities in the measurement uncertainty relation: on the one hand by a calibration process, based on worst case deviations of the output distribution, when the device is presented with states of known and sharp position (resp. momentum). On the other hand, we introduce a distance of observables based on transportation metrics. In both definitions, a power α can naturally be used. Thus for preparation uncertainty as well as both definitions of errors for measurement uncertainty, we will prove relations of the form

$$\Delta_\alpha(P)\Delta_\beta(Q) \geq c_{\alpha\beta}\,\hbar, \tag{1}$$

of course with quite different interpretations of the Δ-quantities. The constants $c_{\alpha\beta}$, however, will be the same and best possible in all three cases. The cases of equality can be characterized precisely (they depend on α and β). Moreover, there is a second constant $c'_{\alpha\beta}$ such that an uncertainty product u with $c_{\alpha\beta}\hbar \leq u < c'_{\alpha\beta}\hbar$ implies that the state (in the case of preparation uncertainty) or the observable (in the case of measurement uncertainty) is uniformly close to one with strictly minimal uncertainty, with an error bound going to zero as (1) becomes sharp.

The basic idea of measurement uncertainty and the idea of calibration errors was presented, together with a sketch of the proof, in a recent Letter.[13] Here we give a full version of the proof. At the same time we lift the restriction $\alpha = \beta = 2$, thereby covering also the previously studied case $\alpha = \beta = 1$ based only on the metric uncertainty definition.[14]

Our paper is organized as follows: In Sec. II, we review the concept of an observable as an object that assigns to all states the probability measures for the outcomes of a given measurement. We use this language to describe the idea of an approximate joint measurement of noncommuting observables, such as position and momentum, for which the traditional perspective of an observable as a self-adjoint operator is inadequate. We then recall the definition and some relevant properties of a covariant phase space observable, which constitutes a fundamentally important special case of an approximate joint measurement. Section III presents measures of measurement errors, understood as the difference between the observable actually measured in a measurement scheme and the target observable. This difference is quantified in terms of the so-called Wasserstein distance of order α for probability measures. Maximising this distance over all pairs of probability distributions of the estimator and target observables for the same state yields a metric distance between observables. If the maximization is performed over calibrating states only, we obtain a measure of calibration error for observables.

In Sec. IV, we formulate and prove our main result—the joint measurement and error-disturbance relations for position and momentum. These inequalities are obtained as consequences of preparation uncertainty relations for these quantities, where the usual product of standard deviations is replaced with more general choices of α-deviations for the position and momentum, respectively. The tight lower bound, which will be determined explicitly and investigated numerically in Sec. V, is given by Planck's constant $\hbar$ multiplied by a constant that depends on the choice of deviation measure. In

Sec. VI, some possible extensions and generalizations are briefly discussed. Section VII concludes with a summary and outlook.

II. OBSERVABLES

The setting of this paper is standard quantum mechanics of a single canonical degree of freedom. In this section, we fix some notation and terminology.

An *observable* is the mathematical object describing a measuring device. This description must provide, for every input state given by a density operator ρ, the probability distribution of the measurement outcomes. The set Ω of possible outcomes is part of the basic description of the device, and in order to talk about probabilities it must come equipped with a σ-algebra of "measurable" subsets of Ω, which we suppress in the notation. The only cases needed in this paper are $\Omega = \mathbb{R}$ (position or momentum) and $\mathbb{R}^2$ (phase space), or subspaces thereof, each with the Borel σ-algebra. For the observable F, we denote the outcome probability measure on Ω in the state ρ by F_ρ. Since, for every measurable $X \subset \Omega$, $\rho \mapsto F_\rho(X)$ must be a bounded linear functional, there is a positive operator $F(X)$ such that $F_\rho(X) = \operatorname{tr} \rho F(X)$. The measure property of F_ρ then implies an operator version thereof, i.e., for a sequence of disjoint X_i we have $F(\bigcup X_i) = \sum_i F(X_i)$, where the sum converges in the weak, strong, and ultraweak operator topologies. We take all observables to be normalized, i.e., $F(\Omega) = \mathbb{1}$. An observable is thus (given by) a normalized positive operator valued measure on the outcome space of a measurement.

When all the operators $F(X)$ are projection operators, we say that F is a *sharp* observable. The prime example is the unique spectral measure on $\mathbb{R}$ associated with a self-adjoint operator A, which we denote by E^A, and which is uniquely determined by A through the resolution formula $A = \int x\, dE^A(x)$. Most textbooks use the term "observable" synonymously with "self-adjoint operator" and go on to explain how to determine the outcome probabilities by using the spectral measure E^A or, equivalently, how expectation values of functions of the outcomes are to be computed as the expectations of functions of the operator in the functional calculus. For the purposes of this paper (and many other purposes), this view is too narrow, since the phase space observables describing an approximate joint measurement of position and momentum cannot be sharp. However, the "ideal" position and momentum measurements will be described by the usual sharp observables E^Q and E^P.

The principal object we study are joint measurements of position and momentum. These are observables with two real valued outcomes (i.e., $\Omega = \mathbb{R}^2$), where the first outcome is called the position outcome and the second the momentum outcome. These labels have no significance, except that we will compare the first outcomes with those of a standard position measurement and the second to those of a standard momentum observable. More precisely, we denote by M^Q the first "position" marginal of the observable (i.e., $M^Q(X) = M(X \times \mathbb{R})$) and ask to what extent we can have $M^Q \approx E^Q$, and at the same time $M^P \approx E^P$ for the second marginal M^P. The precise interpretation of the approximation will be discussed in Sec. III D. This rather abstract approach covers many concrete implementations, including, of course, the scenario of Heisenberg's microscope, where an approximate position measurement is followed by a standard momentum measurement. The quality of the approximation $M^Q \approx E^Q$ is then quantified by the "error" of the measurement or, put positively, by the resolution of the microscope. The approximation $M^P \approx E^P$ compares the momentum after the measurement with the direct momentum measurement, i.e., the measurement without the microscope. The approximation error here quantifies the "momentum disturbance." However, we emphasize that the joint measurement need not be constructed in this simple way. For example, we could make any suitable measurement on the particle after the position measurement, designed to make the approximation $M^P \approx E^P$ as good as possible. For this we could use the position outputs and everything we know about the construction of the microscope, correcting as much as possible the systematic errors introduced by this device. So "momentum disturbance" is not just a question whether the direct momentum measurement after the microscope still gives a good result, but whether there is any way at all of retrieving the momentum from the post-measurement state. Needless to say, the joint measurement perspective also restores the symmetry between position and momentum, i.e., the results apply equally to an approximate momentum measurement disturbing

the position information. In any case, the results will be quantitative bounds expressing that $M^Q \approx E^Q$ and $M^P \approx E^P$ cannot both hold with high accuracy.

A. Covariant phase space observables

Covariant phase space observables are of fundamental importance for this study. Though extensively studied in the literature we also recall briefly their definitions and characteristic properties. For details, see, for instance, Refs. 15–20.

In his famous paper, Heisenberg announced that he would show "a direct mathematical connection" between the uncertainty relation and the commutation relations of position and momentum (and henceforth forgets this announcement). Of course, this is what we will do. Due to von Neumann's uniqueness theorem for the commutation relations we may as well start from the usual form of the operators: Q is the operator of multiplication by x on $\mathcal{L}^2(\mathbb{R}, dx)$, and P is the differentiation operator $P\psi = -i\psi'$. Here and in the sequel we will set $\hbar = 1$. The joint translation by q in position and by p in momentum are implemented by the Weyl operators (also known as Glauber translations) $W(q, p) = \exp(ipQ - iqP)$ acting explicitly as

$$(W(q, p)\psi)(x) = e^{-iqp/2 + ipx} \, \psi(x - q). \tag{2}$$

Of course, these commute only up to phases, which are again equivalent expression of the commutation relations.

A *covariant observable* is defined as an observable M with phase space outcomes $(\Omega = \mathbb{R}^2)$ such that, for any measurable $Z \subset \mathbb{R}^2$,

$$M\big(Z - (q, p)\big) = W(q, p)^* M(Z) W(q, p). \tag{3}$$

This implies[16,17,19,20] that the measure M has an operator valued density with respect to Lebesgue measure, that the densities at different points are connected by the appropriate phase space translations, and that the density at the origin is actually itself a density operator σ, i.e., a positive operator with unit trace. Explicitly, we get the formula

$$M(Z) = \int_{(q,p) \in Z} W(q, p) \Pi \sigma \Pi W(q, p)^* \, \frac{dq \, dp}{2\pi\hbar} \, . \tag{4}$$

Here we deviated from the announcement to set $\hbar = 1$ to emphasize that the measure for which the density of a normalized observable has trace 1 is the usual volume normalization in units of Planck's "unreduced" constant $h = 2\pi\hbar = 2\pi$. The operator Π is the parity operator $(\Pi\psi)(x) = \psi(-x)$, and merely changes the parametrization of observables in terms of σ. The reason for this will become clear presently.

We need to compute the expectations M_ρ^Q of the Q-marginal of a covariant observable. For the sake of this computation, we may set $\rho = |\psi\rangle\langle\psi|$ and $\sigma = |\phi\rangle\langle\phi|$, and later extend by linearity. Then the probability density of M_ρ^Q at the point q is obtained from (4) by taking the expectation with ρ and leaving out the integral over q while retaining the one over p. This gives

$$\frac{1}{2\pi} \int \langle \psi, W(q, p)\Pi\phi \rangle \langle W(q, p)\Pi\phi, \psi \rangle \, dp$$

$$= \frac{1}{2\pi} \int \overline{\psi(x)} \psi(y) \phi(q - x) \overline{\phi(q - y)} \, e^{i(px - py)} \, dp \, dx \, dy = \int |\psi(x)|^2 \, |\phi(q - x)|^2 \, dx \, . \tag{5}$$

Here we used $\int e^{ipx} dp = 2\pi\delta(x)$. The result is the convolution of the position distributions of ρ and σ. Together with the analogous relation for momentum, we can write this as

$$M_\rho^Q = E_\rho^Q * E_\sigma^Q \quad \text{and} \quad M_\rho^P = E_\rho^P * E_\sigma^P \, , \tag{6}$$

where the star denotes the convolution of measures or their density functions. Thus we arrive at the key feature of covariant measurements for our study: *The marginal distributions of a covariant measurement are the same as those of the corresponding ideal measurement with some added noise,*

which is statistically independent of the input state. The noise distributions are just E_σ^Q and E_σ^P, so they are constrained exactly by the preparation uncertainty of σ.

Remark 1. There is a converse to Eq. (6). Rather than asking how we can approximately measure the standard position and momentum observables together, we can ask under which conditions approximate position and momentum observables can exactly be measured together. Here by an approximate position measurement we mean an observable F, which is covariant for position shifts, and commutes with momentum, so that $F(X - q) = W(q, p)F(X)W^*(q, p)$. These are necessarily of the form $F_\rho = \mu * E_\rho^Q$ for some measure μ (see, e.g., Ref. 21). Suppose that we have such an approximate position measurement and, similarly, an approximate momentum measurement given by a noise measure ν. Then, using the averaging technique developed in Ref. 14 (reviewed below), it can also be shown that these two are jointly measurable, i.e., they are the marginals of some phase space observable M, if and only if there is a covariant observable M with these marginal, i.e., if and only if $\mu = E_\sigma^Q$ and $\nu = E_\sigma^P$ for some density operator σ.[22]

Remark 2. The parity operator appears either in (4) or in (6). The convention we chose is in agreement with the extension of the convolution operation from classical measures on phase space to density operators and measures. Indeed, a convolution can be read as the average over the translates of one factor weighted with the other. Therefore the convolution of a density operator ρ and a probability measure μ on phase space is naturally defined as the density operator

$$\mu * \rho = \rho * \mu = \int W(q, p)\rho W^*(q, p) \, d\mu(q, p) . \tag{7}$$

This sort of definition would work for any group representation. A special property of the Weyl operators ("square integrability") allows us to define[17] a convolution also of two density operators giving the probability density

$$(\rho * \sigma)(q, p) = \operatorname{tr} \rho \, W(q, p)\Pi\sigma \Pi W^*(q, p) . \tag{8}$$

It turns out that the integrable functions on phase space together with the trace class then form a commutative and associative Banach algebra or, more precisely, a $\mathbb{Z}_2$-graded Banach algebra, where functions have grade 0, operators have grade 1, and the grade of a product is the sum of the grades mod 2. Since this algebra is commutative, it can be represented as a function algebra, which is done by the Fourier transform for functions and by the Fourier-Weyl transform $(\mathcal{F}\rho)(q, p) = \operatorname{tr} \rho W(q, p)$ for operators. The Wigner function of ρ is then the function (not usually integrable) that has the same Fourier transform as ρ. This explains why the convolution of two Wigner functions is positive: this is just the convolution of the density operators in the sense of (8). A similar argument will be used in the proof of Proposition 21.

With this background the appearance of convolutions in (6) is easily understood: It is just the equation $M_\rho = \rho * \sigma$ for the phase space density, integrated over p and q, respectively. Similarly, it becomes clear that any kind of variance of the observed joint distribution M_ρ will be the sum of two terms, one coming from the preparation ρ and one coming from the measurement defined by σ, but that these two play interchangeable roles.

III. QUANTIFYING MEASUREMENT ERRORS

In a fundamentally statistical theory like quantum mechanics the results of individual measurements tell us almost nothing: It is always the probability distribution of outcomes for a fixed experimental arrangement which can properly be called the result of an experiment. The fact that even for a pure state ($\rho = |\psi\rangle\langle\psi|$), the probabilities $\langle \psi | F(X)\psi \rangle$ usually take values other than 0 or 1 is not a bug but a feature of quantum mechanics. Therefore the variance of position in a particular state has nothing to do with an "error" of measurement. There is no "true value" around which the outcomes are scattering, and which the measurement is designed to uncover. The variance merely provides some partial information about the probability distribution and a careful experimenter will

record as much information about this distribution as can be reliably inferred from his finite sample of individual measurements.

Nevertheless, experimental errors occur in this process. However, they cannot be detected from just one distribution. Instead they are related to a difference between the observable the experimenter tries to measure and the one that she actually measures. When the state is fixed, this amounts to a difference between two probability distributions. In this section, we will review some ways of quantifying the distance between probability distributions. For the sake of discussion let us take two probability measures μ and ν on some set Ω.

Remark 3. For a probability measure μ on the real line one may determine its moments $\mu[x^n]$ = $\int x^n \, d\mu(x)$, $n = 0, 1, 2, \ldots$. Even if all the moment integrals exist and are finite they do not necessarily determine the probability measure. Thus, on the statistical level of moments, two probability measures μ and ν may be indistinguishable even when $\mu \neq \nu$. This is not a mathematical artefact but a rather common quantum mechanical situation. Indeed, consider, for instance, the double-slit states defined by the functions $\psi_\delta = \frac{1}{\sqrt{2}}(\psi_1 + e^{i\delta}\psi_2)$, $\delta \in \mathbb{R}$, where ψ_1, ψ_2 are smooth functions with disjoint compact supports (in the position representation). A direct computation shows that the moments of the momentum distribution $E^P_{\psi_\delta}$ are independent of δ although the distribution $p \mapsto |\hat{\psi}_\delta(p)|^2 = \frac{1}{2}[|\hat{\psi}_1(p)|^2 + |\hat{\psi}_2(p)|^2 + 2\text{Re}(\overline{\hat{\psi}_1(p)}\hat{\psi}_2(p)e^{i\delta})]$ is δ-dependent. Therefore, the moments do not distinguish between the different distributions $E^P_{\psi_\delta}$ and $E^P_{\psi_\gamma}$, $\delta \neq \gamma \,(\text{mod}\, 2\pi)$. This is to remind us that a discrimination between two probability measures cannot, in general, be obtained from moments alone; in particular, expectations $\mu[x]$ and variances $\mu[x^2] - \mu[x]^2$ are not enough. If μ is compactly supported or exponentially bounded, then the moments $(\mu[x^n])_{n \geq 0}$ uniquely determine μ (see, for instance, Ref. 23).

A. Variation norm

The most straightforward distance measure is the *variation norm*, which is equal to the $\mathcal{L}^1$ distance of the probability densities when the two measures are represented by densities with respect to a reference measure. Operationally, it is twice the largest difference in probabilities:

$$\|\mu - \nu\|_1 = 2 \sup_{X \subset \Omega} |\mu(X) - \nu(X)| = \sup\left\{ \left| \int f(x)d\mu(x) - \int f(x)d\nu(x) \right| \,\Big|\, -1 \leq f(x) \leq 1 \right\}. \quad (9)$$

For observables E and F, we consider the corresponding norm

$$\|E - F\| = \sup_{\rho} \|E_\rho - F_\rho\|_1 = \sup_{\rho, f} \text{tr} \, \rho\left(\int f(x)dE(x) - \int f(x)dF(x) \right), \quad (10)$$

where the sup runs over all density operators ρ and all measurable functions f with $-1 \leq f(x) \leq 1$. Thus the statement "$\|E - F\| \leq \varepsilon$" is equivalent to the rather strong claim that no matter what input state and outcome event we look at, the probability predictions from E and F will never differ by more than $\varepsilon/2$.

However, this measure of distance is not satisfactory for quantifying measurement errors of continuous variables. Indeed there is no reference to the distance of underlying points in Ω. Thus two point measures of nearby points will have distance 2, no matter how close the points are. Another way of putting this is to say that variation distance is dimensionless like a probability. What we often want, however, is a distance of probability distributions, say, on position space, which is measured in meters. The distance of two point measures would then be the distance of the points, and shifting a probability distribution by δx would introduce an "error" of no more than $|\delta x|$. It is clear that this requires a metric on the underlying space Ω, so from now on we assume a metric $D : \Omega \times \Omega \to \mathbb{R}_+$ to be given. Of course, in the case of $\mathbb{R}$ or $\mathbb{R}^n$ we just take the Euclidean distance $D(x, y) = |x - y|$.

B. Metric distance from a point measure

Let us begin with a simple case, which is actually already sufficient for preparation uncertainty and for the calibration definition of measurement uncertainty: We assume that one of the measures is

a point measure, say $\nu = \delta_y$. Then, for $1 \leq \alpha < \infty$, we define the *deviation of order* α, or α-*deviation*, of μ from y as

$$D_\alpha(\mu, \delta_y) = \left(\int D(x, y)^\alpha \, d\mu(x) \right)^{\frac{1}{\alpha}}. \tag{11}$$

The letter D is intentionally chosen to be the same: This definition will be an instance of the general extension of the underlying metric D from points to probability measures. Note that, in particular, we have $D_\alpha(\delta_x, \delta_y) = D(x, y)$ for all α, x, y. We will also consider the limiting case $\alpha = \infty$, for which we have to set

$$D_\infty(\mu, \delta_y) = \mu - \mathrm{ess} \sup\{D(x, y) | x \in \Omega\} = \inf\left\{ t \geq 0 \,\Big|\, \mu\{(x, y) | D(x, y) > t\} = 0 \right\}. \tag{12}$$

Of course, any one of the expressions (11), (12) may turn out to be infinite.

Connected to the α-deviations are the α-*spreads* or minimal deviations of order α, namely,

$$\Delta_\alpha(\mu) = \inf_{y \in \mathbb{R}} D_\alpha(\mu, \delta_y) = \inf_y \left(\int |x - y|^\alpha \, d\mu(x), \right)^{\frac{1}{\alpha}}, \tag{13}$$

where the second expression (valid for $1 \leq \alpha < \infty$) just inserts the definition of D_α for the only metric space ($\Omega = \mathbb{R}$, absolute value distance), which we actually need in this paper. When (11) is interpreted as distance, (13) represents the smallest distance of μ to the set of point measures. In the case of $\Omega = \mathbb{R}$ and $\alpha = 2$, we recover the ordinary standard deviation, since the infimum is attained for y equal to the mean. The point y to which a given measure is "closest" depends on α. For the absolute deviation ($\alpha = 1$) this is the median, for $\alpha = 2$ it is the mean value, and for $\alpha = \infty$ it is the midpoint of the smallest interval containing the support $\mathrm{supp}\,(\mu)$ of μ, the smallest closed set of full measure.

The interpretation of (11)/(12) as "distance to a point measure" hinges on the possibility to extend this definition to a metric proper on the set of probability measures. This is done in Sec. III C.

C. Metric distance for probability distributions

The standard distance function with the properties described above is known as the Monge-Kantorovich-Wasserstein-Rubinstein-Ornstein-Gini-Dall'Aglio-Mallows-Tanaka "earth mover's" or "transportation" distance, or some combination of these names (see Ref. 24 for background and theory, for $\alpha = \infty$ we refer to Refs. 25 and 26). For purpose of assessing the accuracy of quantum measurements it was apparently first used by Wiseman.[27] The natural setting for this definition is an outcome space Ω, which is a complete separable metric space with metric $D : \Omega \times \Omega \to \mathbb{R}_+$. For any two probability measures μ, ν on Ω, we define a *coupling* to be a probability measure γ on $\Omega \times \Omega$ with μ and ν as the marginals. We denote by $\Gamma(\mu, \nu)$ the set of couplings between μ and ν. Then, for any $\alpha, 1 \leq \alpha < \infty$ we define the α-distance (also Wasserstein α-distance[24]) of μ and ν as

$$D_\alpha(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} D_\alpha^\gamma(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \left(\int D(x, y)^\alpha \, d\gamma(x, y) \right)^{\frac{1}{\alpha}}. \tag{14}$$

For $\alpha = \infty$, one again defines $D_\infty^\gamma(\mu, \nu) = \gamma - \mathrm{ess}\ \sup\{D(x, y) \,|\, (x, y) \in \Omega \times \Omega\}$ and thus

$$D_\infty(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} D_\infty^\gamma(\mu, \nu). \tag{15}$$

Actually, $D_\infty^\gamma(\mu, \nu)$ depends only on the support of γ, i.e., $D_\infty^\gamma(\mu, \nu) = \sup\{D(x, y) \,|\, (x, y) \in \mathrm{supp}\,(\gamma)\}$.

The existence of an optimal coupling is known, for $1 \leq \alpha < \infty$, see Theorem 4.1 of Ref. 24, the case $\alpha = \infty$ is shown in Theorem 2.6 of Ref. 26, but it does not imply that $D_\alpha(\mu, \nu)$ is finite.

When $\nu = \delta_y$ is a point measure, there is only one coupling between μ and ν, namely, the product measure $\gamma = \mu \times \delta_y$. Hence (14)/(15) reduces to (11)/(12). In particular, we indeed get an extension of the given metric for points, interpreted as point measures. The metric can become

infinite, but the triangle inequality still holds[24][after Example 6.3]. The proof relies on Minkowski's inequality and the use of a "Gluing Lemma,"[24] which builds a coupling from μ to ζ out of couplings from μ to ν and from ν to ζ. It also covers the case $\alpha = \infty$, which is not otherwise treated in Ref. 24.

Thus the space breaks up into equivalence classes of measures, which have finite distance from each other. In view of Sec. III B, it is natural to consider the class containing all point measures, i.e., the measures of finite spread. However, the restriction to measures of finite spread is not necessary. In fact, for measures on $\mathbb{R}$ each equivalence is closed under translations, and the distance of two translates is bounded by the size of the translation.

The metric also has the right *scaling*: For $\Omega = \mathbb{R}$ let us denote the scaling of measures by s_λ, so that for $\lambda > 0$ and measurable $X \subset \mathbb{R}$, $s_\lambda(\mu)(X) = \mu(\lambda^{-1}X)$. Then $D_\alpha(s_\lambda\mu, s_\lambda\nu) = \lambda D_\alpha(\mu, \nu)$, so this metric is compatible with a change of units. Of course, the metric is also unchanged when both measures are shifted by the same translation.

$D_\alpha(\mu, \nu)$ is also known as *transport distance*, due to the following interpretation: Suppose that an initial distribution μ of some "stuff" (earth or probability) has to be converted to another distribution ν by moving the stuff around. The measure γ then encodes how much stuff originally at x is moved to y. If the transport cost per unit is $D(x, y)^\alpha$, the integral represents the total transport cost. The minimum then is the minimal cost of converting μ to ν. The root is taken to ensure the right scaling behaviour.

For *convexity properties*, note that the function $t \mapsto t^{1/\alpha}$ is concave, so by Jensen's inequality $D_\alpha(\mu, \delta_y)$ is concave in μ, and so is $\Delta_\alpha(\mu)$, as the infimum of concave functions. This is expected, since it is exactly the point measures that have zero spread, and all other measures are convex combinations of these. The metric is neither concave nor convex in its arguments. However, the function $\gamma \mapsto D_\alpha^\gamma(\mu, \nu)^\alpha$ is linear, and since $\lambda\gamma_1 + (1 - \lambda)\gamma_2$ is a coupling between the respective convex combinations of marginals, $(\mu, \nu) \mapsto D_\alpha(\mu, \nu)^\alpha$ is convex. Therefore, the level sets $\{(\mu, \nu)|D_\alpha(\mu, \nu) \leq c\}$ are convex, and D_α is "pseudoconvex."

Suppose that $\Omega = \mathbb{R}$ and we "add independent noise" to a real-valued random variable with distribution μ by a random translation with distribution η. This leads to the convolution $\eta * \mu$ for the new distribution. The following bounds govern this operation

Lemma 4. Let μ, ν, η be probability measures on $\mathbb{R}$, and $1 \leq \alpha \leq \infty$. Then

$$\Delta_\alpha(\mu) \leq \Delta_\alpha(\eta * \mu) \leq \Delta_\alpha(\eta) + \Delta_\alpha(\mu), \tag{16}$$

$$D_\alpha(\eta * \mu, \eta * \nu) \leq D_\alpha(\mu, \nu), \tag{17}$$

$$D_\alpha(\eta * \mu, \mu) \leq D_\alpha(\eta, \delta_0). \tag{18}$$

The first inequality says that noise increases spread, but not by more than the spread of the noise. The second says that adding noise washes out the features of two distributions, making them more similar. Finally the third inequality, which will be crucial for us, says that adding a little noise only changes a measure by little. Note that it is not only the spread of the noise which counts here, but also the absolute displacement. That is, a special case of the last relation is that $D_\alpha(\mu^y, \mu) \leq |y|$, where $\mu^y = \delta_y * \mu$ is the shift of μ by y. We emphasize that (18) does not require $\Delta_\alpha(\mu) < \infty$.

Proof. We note that $D_\alpha(\mu, \delta_y)$ is a standard p-norm, or rather α-norm $\|\cdot\|_{\mu,\alpha}$ in $\mathcal{L}^\alpha(\Omega, \mu)$ of the function $x \mapsto (x - y)$. We denote this function as $x - y\mathbf{1}$ to indicate that y is considered as a constant. That is

$$D_\alpha(\mu, \delta_y) = \|x - y\mathbf{1}\|_{\mu,\alpha}. \tag{19}$$

This equation is also valid for $\alpha = \infty$, so we need not consider this case separately.

For the first inequality in (16), we use translation invariance and concavity of Δ_α by considering $\eta * \mu$ as a convex combination of translates of μ with weight η. For the second inequality in (16)

consider the expression

$$\|x + y - x' - y'\|_{\eta \times \mu, \alpha} = D_\alpha(\mu * \nu, \delta_{x'+y'}).$$

This is larger than the infimum over all choices of $(x' + y')$, i.e., $\Delta_\alpha(\mu * \nu)$. Using the Minkowski inequality (triangle inequality for the α-norm), we conclude

$$\begin{aligned}
\Delta_\alpha(\mu * \nu) &\le \|x - x' + y - y'\|_{\eta \times \mu, \alpha} \\
&\le \|x - x'\|_{\eta \times \mu, \alpha} + \|y - y'\|_{\eta \times \mu, \alpha} \\
&= \|x - x'\|_{\eta, \alpha} + \|y - y'\|_{\mu, \alpha} \\
&= D_\alpha(\mu, \delta_{x'}) + D_\alpha(\nu, \delta_{y'}).
\end{aligned}$$

Here at the last but one equality we used that in the relevant integrals the integrand depends only on one of the two variables x, y and the other is integrated over by a probability measure. The desired inequality (16) now follows by minimizing over x' and y'.

Then any coupling γ between μ and ν provides a coupling $\tilde{\gamma}(X \times Y) = \int \gamma(X - x, Y - x) \, d\eta(x)$ between $\eta * \mu$ and $\eta * \nu$, for which we get $\int |x - y|^\alpha \, d\tilde{\gamma}(x, y) = \int |x - y|^\alpha \, d\gamma(x, y)$. Since the infimum may be attained at a coupling different from $\tilde{\gamma}$ (17) follows.

Finally, to prove (18), we note that it is a special case of (17) since $D_\alpha(\eta * \mu, \mu) = D_\alpha(\mu * \eta, \mu * \delta_0)$. □

A powerful tool for working with the distance functions is a dual expression of the infimum over couplings as a supremum over certain other functions. A nice interpretation is in terms of transportation prices. We describe it here to motivate the expressions, and refer to the excellent book,[24] from where we took this interpretation, for the mathematical details. In this context, we have to exclude the case $\alpha = \infty$.

Suppose the "stuff" to be moved is bread going from the bakeries in town to cafés. The transport costs $D(x, y)^\alpha$ per unit, and the bakeries and cafés consider hiring a company to take care of the task. The company will pay a price of $\Psi(x)$ per unit to the bakery at x and charges $\Phi(y)$ from the café at y. Clearly this makes sense if each transport becomes cheaper, i.e.,

$$\Phi(y) - \Psi(x) \le D(x, y)^\alpha. \tag{20}$$

Pricing schemes satisfying this inequality are called competitive. We are now asking what the maximal gain of a company under a competitive pricing scheme can be, given the productivity μ of the bakeries and the demand ν at the cafés. This will be

$$\int \Phi(y) \, d\nu(y) - \int \Psi(x) \, d\mu(x) \le \int D(x, y)^\alpha \, d\gamma(x, y). \tag{21}$$

This inequality is trivial from (20), and holds for any pricing scheme (Ψ, Φ) and any transport plan γ. Optimizing the pricing scheme maximizes the left-hand side and optimizing the transport plan minimizes the right-hand side. The Kantorovich Duality Theorem asserts that for these optimal choices the gap closes, and equality holds in (21), i.e.,

$$D_\alpha(\mu, \nu)^\alpha = \sup_{\Phi, \Psi} \int \Phi(y) \, d\nu(y) - \int \Psi(x) \, d\mu(x), \tag{22}$$

where Φ and Ψ satisfy (20).

When maximizing the left-hand side of (21), one can naturally choose Φ as large as possible under the constraint (20), i.e., $\Phi(y) = \inf_x \{\Psi(x) + D(x, y)^\alpha\}$, and similarly for Ψ. Hence one can choose just one variable Φ or Ψ and determine the other by this formula. In case $\alpha = 1$, the triangle inequality for the metric D entails that one can take $\Phi = \Psi$. In this case (20) just asserts that this function be Lipshitz continuous with respect to the metric D, with constant 1. The left-hand side of (21) is thus a difference of expectation values of the given measures μ, ν.

For later purposes we also have to make sure that the duality gap closes if we restrict the set of functions Φ, Ψ. The natural condition is, of course, that $\Psi \in \mathcal{L}^1(\mu)$. The statement of Kantorovich

Duality in Theorem. 5.10 of Ref. 24 includes that in the supremum (22) one can restrict to bounded continuous functions. In the same spirit, we add

Lemma 5. In (22), the supremum can be restricted to positive continuous functions of compact support without changing its value.

Proof. Suppose that some bounded continuous functions Ψ, Φ are given, which satisfy (20). Since we can add the same constant to each, we may also assume them to be positive. Our aim is to find compactly supported functions Ψ_ε, Φ_ε such that $0 \le \Psi_\varepsilon \le \Psi$, $\int (\Psi(x) - \Psi_\varepsilon(x))d\mu(x) \le \varepsilon$, and similarly for Φ_ε. The problem is to find such functions so that (20) still holds.

Pick a compact region U so that $\int_{y \notin U}\Phi(y)\,d\nu(y) < \varepsilon$, and some continuous function $0 \le \Phi_\varepsilon \le \Phi$ coinciding with Φ on U and vanishing outside a compact set $\widehat{U} \supset U$. Let V be a set on which $\Psi d\mu$ similarly achieves its integral to within ε, and which also contains $\widehat{U}$ and all points of distance at most $\|\Phi\|_\infty^{1/\alpha}$ from it. Construct a compactly supported function Ψ_ε coinciding with Ψ on V. Consider now the inequality

$$\Phi_\varepsilon(y) - \Psi_\varepsilon(x) \le D(x, y)^\alpha.$$

Clearly this holds for all $x \in V$, because then $\Psi_\varepsilon(x) = \Psi(x)$ and $\Phi_\varepsilon(y) \le \Phi(y)$. For $y \in \widehat{U}$ and $x \notin V$, we have

$$D(x, y)^\alpha - \Phi_\varepsilon(y) \ge D(x, y)^\alpha - \Phi(y) \ge D(x, y)^\alpha - \|\Phi\|_\infty \ge 0.$$

Hence the inequality follows from $\Psi_\varepsilon \ge 0$. Finally for $y \notin \widehat{U}$ we have $\Phi_\varepsilon(y) = 0$, and the inequality is once again trivial. To summarize, we have shown it on $(\Omega \times V) \cup (\widehat{U} \times V^c) \cup (\widehat{U}^c \times \Omega) = \Omega \times \Omega$.

Of the properties of $\Omega = \mathbb{R}$ we only used (for the compactness of V) that closed balls of the metric are compact. $\qquad\square$

Example 6. 2-distance and affine families.

It is instructive to see just how far one can go by taking Φ, Ψ in (21) to be quadratic expressions in the case $\alpha = 2$. So let $\Psi(x) = (a - 1)x^2 + 2bx$ with $a > 0$. Then

$$\Phi(y) = \inf_x \{\Psi(x) + (x - y)^2\} = \left(1 - \frac{1}{a}\right)y^2 + \frac{2by}{a} - \frac{b^2}{a}. \tag{23}$$

Now let μ, ν be probability measures with finite second moments, say means $m(\mu) = \mu[x]$, $m(\nu) = \nu[x]$ and variances $s(\mu)^2$, $s(\nu)^2$. Then the left-hand side of (21) can be entirely expressed by these moments, and we get a lower bound

$$D_2(\mu, \nu)^2 \ge (s(\mu) - s(\nu))^2 + (m(\mu) - m(\nu))^2 \tag{24}$$

$$\ge \frac{a - 1}{a}(s(\mu)^2 + m(\mu))^2 + \frac{2b}{a}m(\mu) - \frac{b^2}{a} - (a - 1)(s(\nu)^2 + m(\nu)^2) - b\,m(\nu).$$

Here the second expression is what one gets by just inserting the moments (e.g., $m(\mu)$ and $s(\mu)^2 + m(\mu)^2$) into (21), and the first is the result of maximizing over b and $a > 0$. Turning to the upper bound, it is not a surprise that the maximization for getting Ψ leaves a quadratic expression for the difference

$$(x - y)^2 + \Psi(x) - \Phi(y) = \frac{1}{a}(ax + b - y)^2. \tag{25}$$

Hence for equality in (21), and therefore in both inequalities of (24) to hold we must have a coupling γ which is concentrated on the line $y = ax + b$, so $\int g(x, y)\gamma(dx\,dy) = \int g(x, ax + b)d\mu(x)$ for any test function g. In particular, ν must arise from μ by translation and dilatation. But if that is the case the moments just have to come out right, so the converse is also true. Hence we have a very simple formula for D_2 on any orbit of the affine group, for example, the set of Gaussian probability distributions. The argument also allows us to find the shortest D_2-distance from a measure μ to a set of measures with fixed first and second moments: The closest point will be the appropriate affine

argument transformation applied to μ. This, with much further information about D_2 geodesics and the connection with Legendre transforms is to be found in Theorem 3.1 of Ref. 28.

With the above assumption of finite second moments we also get an upper bound, so that

$$(s(\mu) - s(\nu))^2 + (m(\mu) - m(\nu))^2 \leq D_2(\mu, \nu)^2 \leq (s(\mu) + s(\nu))^2 + (m(\mu) - m(\nu))^2,$$

with the bounds obtained if there is a γ giving strong negative, resp. positive correlation for μ and ν, making them linearly dependent.

D. Metric distance of observables

Given an α-deviation for probability distributions we can directly define an α-deviation for observables E, F with the same metric outcome space (Ω, D). We set, for $1 \leq \alpha \leq \infty$,

$$D_\alpha(F, E) := \sup_\rho D_\alpha(F_\rho, E_\rho). \tag{26}$$

Note that we are taking here the *worst case* with respect to the input states. Indeed we consider the deviation of an observable F from an "ideal" reference observable E as a figure of merit for F, which a company might advertise: No matter what the input state, the distribution obtained by F will be ε-close to what you would get with E. When closeness of distributions is measured by D_α, then (26) is the best ε for which this is true. In contrast, the individual deviations $D_\alpha(F_\rho, E_\rho)$ are practically useless as a benchmark. Indeed, a testing lab, which is known to always use the same input state for its tests, is easily fooled. Their benchmark could be met by any fake device, which does not make any measurement, but instead produces random numbers with the expected distribution. Put in colloquial terms: Nobody would buy a meter stick which is advertised as "very precise, provided the length measured is 50 cm," or a watch which "shows the correct time twice a day."

The additional maximization in (26) leads to some simplifications, and in particular to an explicit expression for the difference between a sharp observable and the same observable with added noise.

Lemma 7. Let E be a sharp observable on $\mathbb{R}$, η some probability measure on $\mathbb{R}$, and $F = \eta * E$, i.e., $F_\rho = \eta * E_\rho$ for all ρ. Then

$$D_\alpha(F, E) = D_\alpha(\eta, \delta_0). \tag{27}$$

Proof. By (18), we have $D_\alpha(F_\rho, E_\rho) \leq D_\alpha(\eta, \delta_0)$. We claim that this upper bound is nearly attained whenever E_ρ is sharply concentrated, say, $D_\alpha(E_\rho, \delta_q) \leq \varepsilon$; this is possible, because E was assumed to be sharp. Indeed we then have $D_\alpha(\eta, \delta_0) = D_\alpha(\eta * \delta_q, \delta_q) \leq D_\alpha(\eta * \delta_q, \eta * E_\rho) + D_\alpha(\eta * E_\rho, E_\rho) + D_\alpha(E_\rho, \delta_q) \leq 2\varepsilon + D_\alpha(\eta * E_\rho, E_\rho) = 2\varepsilon + D_\alpha(F_\rho, E_\rho)$. $\square$

Example 8. The standard model for measuring (approximately) a sharp observable associated with the self-adjoint operator A consists in coupling the system with a probe, with the Hilbert space $\mathcal{L}^2(\mathbb{R})$, using the direct interaction $e^{i\lambda A \otimes P_p}$, and monitoring the shifts in the probe's position Q_p. If the probe is initially prepared in a state σ, then the actually measured observable F is a smearing of E^A, with the (λ-scaled) probability density of the probe position in state $\Pi \sigma \Pi$. Thus we get $D_\alpha(F, E^A) = D_\alpha(E_{\Pi\sigma\Pi}^{Q_p/\lambda}, \delta_0)$. This shows that the error in measuring E^A with the standard model can be made arbitrarily small with an appropriate choice of the initial probe state σ but can never be made equal to 0.

Example 9. If the standard measurement of a sharp observable A is followed immediately (in the sense that any free evolution in between can be neglected) by a measurement of another sharp observable B, then the resulting (sequential) joint measurement constitutes an approximate joint measurement of A and B, with the first marginal observable M_1 being a smearing of A, as given in Example 8, and the second marginal M_2 is a distorted version of B,

$$M_2(Y) = \mathcal{I}(\mathbb{R})^*(E^B(Y)) = \int_\mathbb{R} K_x^* E^B(Y) K_x \, dx,$$

where $K_x = \int \sqrt{\lambda}\phi(-\lambda(y-x))\,dE^A(y) = \sqrt{\lambda}\phi(-\lambda(A-x))$ for all $x \in \mathbb{R}$; for simplicity, we have assumed here that the initial probe state σ is a pure state given by a function $\phi \in L^2(\mathbb{R})$ of unit norm.

While $D_\alpha(M_1, E^A)$ is easily computed, the error $D_\beta(M_2, E^B)$ can be determined only if A and B are explicitly given.

For instance, if $A = Q$ and $B = P$, then M_2 is a smearing of E^P, with the convolving probability measure being the ($1/\lambda$-scaled) momentum distribution of the probe in the state $\Pi\sigma\Pi$. A standard position measurement followed by a momentum measurement turns out to be an implementation of a covariant phase space observable M^τ, τ depending on σ. In this case, $D_\alpha(M_1, E^Q)D_\beta(M_2, E^P)$ $= D_\alpha(\mu_\tau, \delta_0)D_\beta(\nu_\tau, \delta_0)$, which reduces to the generalized Kennard-Robertson-Weyl inequality of Proposition 12. For $\alpha = \beta = 2$, one thus gets $D_2(M_1, E^Q)D_2(M_2, E^P) \geq \hbar/2$, where the lower bound is reached exactly when τ is a centered minimal uncertainty state, that is, τ is a pure state given by a real valued Gaussian wave function ϕ whose position and momentum distributions are centered at the origin.

For $\alpha = \infty$, say, the finiteness of the uncertainty product implies that $\beta < \infty$ since there is no τ for which both the position and momentum distributions would have bounded supports.

Remark 10. As seen above, any covariant phase space observable M^σ can be realized, for instance, as a standard (approximate) position measurement followed by a momentum measurement, the generating operator σ depending on the initial probe state. A more realistic implementation of an M^σ can be obtained as the high amplitude limit of the signal observable measured by an eight-port homodyne detector; for details, see Ref. 29.

E. Calibration error

The idea of looking especially at states for which E_ρ is sharply concentrated can be used also in a more general setting, and even gives a possible alternative definition of the error quantities. The idea is that the supremum (26) over all states is not easily accessible to experimental implementation. It seems more reasonable to just calibrate the performance of a measurement F as an approximate measurement of E by looking at the distributions F_ρ for preparations for which E_ρ is nearly a point measure, i.e., those for which E "has a sharp value." The idea of calibration error was formalized in Ref. 30 as a measure of *error bar width* which was shown to obey a measurement uncertainty relation using the method developed in Ref. 14 and applied here.

For $\varepsilon > 0$, we define the ε-*calibration error*, resp. the *calibration error* of F with respect to E as

$$\Delta_\alpha^\varepsilon(F, E) = \sup_{\rho, x}\Big\{ D_\alpha(F_\rho, \delta_x) \,\Big|\, D_\alpha(E_\rho, \delta_x) \leq \varepsilon \Big\}, \tag{28}$$

$$\Delta_\alpha^c(F, E) = \lim_{\varepsilon \to 0} \Delta_\alpha^\varepsilon(F, E). \tag{29}$$

Here the limit in (29) exists because (28) is a monotonely decreasing function. By the triangle inequality, we have $D_\alpha(F_\rho, \delta_x) \leq D_\alpha(F_\rho, E_\rho) + D(E_\rho, \delta_x)$ and, taking the supremum over ρ and x as in (28)

$$\Delta_\alpha^\varepsilon(F, E) \leq D_\alpha(F, E) + \varepsilon \quad \text{and} \quad \Delta_\alpha^c(F, E) \leq D_\alpha(F, E). \tag{30}$$

When F just adds independent noise, there is also the complementary inequality, the direct analog of Lemma 7.

Lemma 11. Let E be a sharp observable on $\mathbb{R}$, η some probability measure on $\mathbb{R}$, and $F = \eta * E$. Then

$$D_\alpha(\eta, \delta_0) - \varepsilon \leq \Delta_\alpha^\varepsilon(F, E) \leq D_\alpha(\eta, \delta_0) + \varepsilon\,, \tag{31}$$

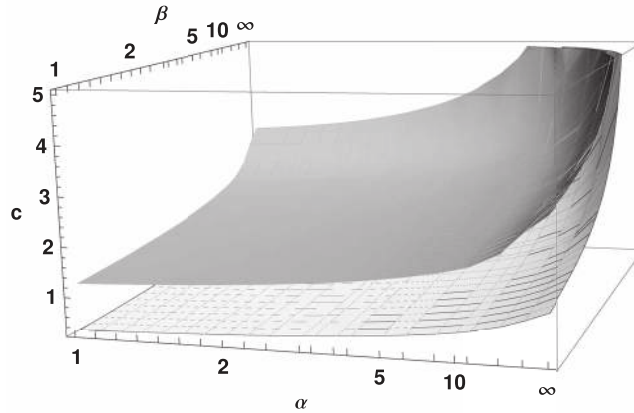*so that letting $\varepsilon \to 0$ yields $\Delta_\alpha^c(F, E) = D_\alpha(\eta, \delta_0)$.*

FIG. 1. The constants $c_{\alpha\beta}$ and $c'_{\alpha\beta}$ appearing in Propositions 13 and 12, as determined numerically in Sec. V. The α and β axes have been scaled according to $\alpha \mapsto (\alpha - 1)/(\alpha + 1)$ to include the whole range $1 \leq \alpha \leq \infty$.

Proof. For any calibration state ρ, i.e., $D_\alpha(E_\rho, \delta_x) \leq \varepsilon$, we have the upper bound $D_\alpha(F_\rho, \delta_x) = D_\alpha(\eta * E_\rho, \delta_x) \leq D_\alpha(\eta * E_\rho, E_\rho) + D_\alpha(E_\rho, \delta_x) \leq D_\alpha(\eta, \delta_0) + \varepsilon$. For the complementary bound, we use $D_\alpha(\eta, \delta_0) = D_\alpha(\eta * \delta_x, \delta_x) \leq D_\alpha(\eta * \delta_x, \eta * E_\rho) + D_\alpha(\eta * E_\rho, \delta_x) \leq \varepsilon + D_\alpha(\eta * E_\rho, \delta_x)$. Hence

$$D_\alpha(\eta, \delta_0) - \varepsilon \leq D_\alpha(F_\rho, \delta_x) \leq D_\alpha(\eta, \delta_0) + \varepsilon,$$

$$D_\alpha(\eta, \delta_0) - \varepsilon \ \leq \Delta^\varepsilon_\alpha(F, E) \leq \ D_\alpha(\eta, \delta_0) + \varepsilon, \tag{32}$$

where the second row is the supremum of the first over all x and all calibrating states. $\square$

Hence, in the case of convolution observables $F = \eta * E$, we have $D_\alpha(F, E) = \Delta^c_\alpha(F, E)$. In general, however, the inequality (30) is strict, as is readily seen by choosing a discrete metric on two points ($\Omega = \{0, 1\}$). Then $D_\alpha(F, E) = \sup_\rho |\operatorname{tr} \rho(F(\{1\}) - E(\{1\}))|/2$, but $\Delta^c_\alpha(F, E)$ is a similar expression with ρ constrained to diagonal pure states.

IV. ERROR-DISTURBANCE RELATIONS: POSITION AND MOMENTUM

A. Preparation uncertainty

Before we can formulate the main result of our paper we state a generalization of the Kennard-Robertson-Weyl inequality for the α-spreads introduced in (13). This result improves the inequality derived by Hirschman,[6] see also Refs. 31 and 32, in that we now have an optimal lower bound. The details of the constants $c_{\alpha\beta}$ as a function of α and β will be studied numerically in Sec. V, with an overview given in Fig. 1.

Proposition 12 (Preparation uncertainty). Let E^Q and E^P be canonically conjugate position and momentum observables, and ρ a density operator. Then, for any $1 \leq \alpha, \beta < \infty$,

$$\Delta_\alpha(E^Q_\rho)\Delta_\beta(E^P_\rho) \geq c_{\alpha\beta}\hbar. \tag{33}$$

The constant $c_{\alpha\beta}$ is connected to the ground state energy $g_{\alpha\beta}$ of the Hamiltonian $H_{\alpha\beta} = |Q|^\alpha + |P|^\beta$ by the equation

$$c_{\alpha\beta} = \alpha^{\frac{1}{\beta}} \beta^{\frac{1}{\alpha}} \left(\frac{g_{\alpha\beta}}{\alpha + \beta} \right)^{\frac{1}{\alpha} + \frac{1}{\beta}}. \tag{34}$$

The lower bound is attained exactly when ρ arises from the ground state of the operator $H_{\alpha\beta}$ by phase space translation and dilatation. For $\alpha = \beta = 2$, $H_{\alpha\beta}$ is twice the harmonic oscillator Hamiltonian with ground state energy $g_{22} = 1$, and $c_{22} = 1/2$.

The Hamiltonian $H_{\alpha\beta}$ appears here mainly through the quadratic from $\langle\psi|H_{\alpha\beta}|\psi\rangle$ where ψ runs over, say, the unit vectors in the Schwartz space of tempered functions. The inequality (33) depends only on the lower bound $g_{\alpha\beta}$ of this form.

This makes sense also for $\alpha = \infty$, when $\langle\psi||q|^\infty|\psi\rangle$ is interpreted by the limit $\alpha \to \infty$, i.e., as ∞ unless ψ vanishes almost everywhere outside $[-1, 1]$, in which case the expectation is zero. The effect of this singular "potential" is to confine the particle to the box $[-1, 1]$. Note that in this case (34) simplifies to $c_{\infty\beta} = g_{\infty\beta}^{1/\beta}$. Of course, since ψ cannot be compactly supported in both position and momentum, we have $c_{\infty\infty} = \infty$.

Proof. Consider the family of operators

$$H_{\alpha\beta}(p, q, \lambda) = \lambda^\alpha|Q - q\,\mathbb{I}|^\alpha + \lambda^{-\beta}|P - p\,\mathbb{I}|^\beta \;\geq g_{\alpha\beta}\,\mathbb{I}, \tag{35}$$

obtained from $H_{\alpha\beta}$ by shifting in phase space by (q, p), and by a dilatation $(Q, P) \mapsto (\lambda Q, \lambda^{-1}P)$. Since these operations are unitarily implemented, the lower bound $g_{\alpha\beta}$ for all these operators is independent of p, q, λ. Now, for a given ρ, we may assume that $\Delta_\alpha(E_\rho^Q)$ and $\Delta_\beta(E_\rho^P)$ are both finite, since these uncertainties do not vanish for any density operator, and one infinite factor hence renders the inequality trivial. Let q be the point for which $D_\alpha(E_\rho^Q, \delta_q)$ attains its minimum $\Delta_\alpha(E_\rho^Q)$, and choose p similarly for P. Then by taking the expectation of (35) with ρ we obtain the inequality

$$\lambda^\alpha \Delta_\alpha(E_\rho^Q)^\alpha + \lambda^{-\beta}\Delta_\beta(E_\rho^P)^\beta \geq g_{\alpha\beta}. \tag{36}$$

The minimum of the left-hand side with respect to λ is attained at

$$\lambda = \left(\frac{\beta\,\Delta_\beta(E_\rho^P)^\beta}{\alpha\,\Delta_\alpha(E_\rho^Q)^\alpha}\right)^{1/(\alpha+\beta)}. \tag{37}$$

Inserting this into (36) gives an expression that depends only on the uncertainty product $u = \Delta_\alpha(E_\rho^Q)\Delta_\beta(E_\rho^P)$, namely,

$$u^{\alpha\beta/(\alpha+\beta)}\,\alpha^{-\alpha/(\alpha+\beta)}\,\beta^{-\beta/(\alpha+\beta)}\,(\alpha + \beta) \geq g_{\alpha\beta}. \tag{38}$$

Now solving for u gives the uncertainty inequality. Moreover, since the left-hand side is still nothing but the expectation of $H_{\alpha\beta}(p, q, \lambda)$ with a suitable choice of parameters, equality holds exactly if ρ is the ground state density operator of $H_{\alpha\beta}(p, q, \lambda)$. But since this operator arises by dilatation and shifts from $H_{\alpha\beta}$, its ground state must arise by the same operations from the ground state of $H_{\alpha\beta}$. $\qquad\square$

For the statements about equality and near equality, which are the subject of the following theorem, we need more information about the operator $H_{\alpha\beta}$, particularly its low-lying eigenvalues. Thus we have to turn the quadratic form into a bona fide self-adjoint operator by the Friedrichs extension. This approach also regulates how to understand the case $\alpha = \infty$; the resulting operator lives on $\mathcal{L}^2([-1, 1], dq)$, with the domain chosen so that the extension of the function to the whole line is in the domain of $|P|^\alpha$. This requires the function and some derivatives to vanish at the boundary. Are there eigenvalues at the bottom of the spectrum? Intuitively, $H_{\alpha\beta}$ is the quantization of a phase space function diverging at infinity, so should have a purely discrete spectrum with eigenvalues going to infinity. This can be verified by the Golden-Thompson inequality according to which, for any $\lambda \geq 0$,

$$\operatorname{tr} e^{-\lambda H_{\alpha\beta}} \leq \operatorname{tr} e^{-\lambda|Q|^\alpha} e^{-\lambda|P|^\beta} = \int e^{-\lambda|q|^\alpha}\,dq \int e^{-\lambda|p|^\beta}\,dp < \infty, \tag{39}$$

see, for instance, p. 94 of Ref. 33. Thus, the positive operator on the left is trace class (and thus compact), and, therefore, the spectrum of the generator of the semigroup $e^{-\lambda H_{\alpha\beta}}$, $\lambda > 0$, consists of a countable discrete set of eigenvalues each of finite multiplicity (Theorem 2.20 of Ref. 34). Since

each of the terms in $H_{\alpha\beta}$ already has strictly positive expectation in any state, the lowest eigenvalue $g_{\alpha\beta}$ is strictly positive. Although this might be an interesting task, we do not prove more fine points about the spectrum of $H_{\alpha\beta}$ in this paper. Supported by the numerical calculations (on which we anyhow have to rely for the concrete values) and some solvable cases, we take for granted that the ground state is non-degenerate and lies in the symmetric subspace, and the first excited state has strictly higher energy and lies in the odd subspace. The gap $g'_{\alpha\beta} - g_{\alpha\beta}$ plays a crucial role in showing the stability of the minimizing states: A state with near-minimal uncertainty product must be close to a state with exactly minimal uncertainty product. The precise statement is as follows:

Proposition 13 (Near-minimal preparation uncertainty). Under the conditions of Proposition 12 consider the case of equality in (33). Suppose that this is only nearly the case, i.e., the uncertainty product is

$$c_{\alpha\beta}\hbar \leq u = \Delta_\alpha(E_\rho^Q)\Delta_\beta(E_\rho^P) < c'_{\alpha\beta}\hbar, \tag{40}$$

where $c'_{\alpha\beta}$ is related to the energy $g'_{\alpha\beta}$ of the first excited state of $H_{\alpha\beta}$ by (34). Then there is a state ρ' minimizing (33) exactly, such that with $\gamma = \alpha\beta/(\alpha + \beta)$,

$$\|\rho - \rho'\|_1 \leq 2\sqrt{\frac{u^\gamma - c_{\alpha\beta}^\gamma}{c'^\gamma_{\alpha\beta} - c_{\alpha\beta}^\gamma}}. \tag{41}$$

The bound in (41) approaches zero as $u \to c_{\alpha\beta}$, and becomes vacuous for $u > c'_{\alpha\beta}$. The constants $c_{\alpha\beta}, c'_{\alpha\beta}$ are shown in Fig. 1, and indications how to compute them will be given in Sec. V.

The assumption (40) entails that the left-hand side of (38) is below the next eigenvalue $g'_{\alpha\beta}$. In this case, we get information about how close ρ must be to the ground state ρ'. This is obtained via a simple and well-known Lemma, whose straightforward application to (38) then gives the inequality (41).

Lemma 14. Let H be a self-adjoint operator with non-degenerate ground state ψ_0 with energy E_0 such that the rest of the spectrum lies above $E_1 > E_0$. Then for any density operator ρ we have $\langle\psi_0, \rho\psi_0\rangle \geq (E_1 - \text{tr}\,\rho H)/(E_1 - E_0)$ and the trace norm bound

$$\left\|\rho - |\psi_0\rangle\langle\psi_0|\right\|_1 \leq 2\sqrt{\frac{\text{tr}\,\rho H - E_0}{E_1 - E_0}}. \tag{42}$$

Proof. The statements about the spectrum of H are equivalent to the operator inequality

$$H \geq E_0|\psi_0\rangle\langle\psi_0| + E_1\big(\mathbb{I} - |\psi_0\rangle\langle\psi_0|\big). \tag{43}$$

Taking the trace with ρ gives the bound on the fidelity $f = \langle\psi_0, \rho\psi_0\rangle$. The bound $\|\rho - |\psi_0\rangle\langle\psi_0|\|_1 \leq 2\sqrt{1 - f}$ holds in general, and is proved easily for pure states ρ and extended to mixed ones by Jensen's inequality for the concave square root function. $\square$

B. The covariant case

Here we will study the simple case of covariant observables described in Sec. II A. By virtue of (4), these are parameterized by a density operator σ, and we will see that both measurement uncertainty quantities (by distance of observables and by calibration) for the marginals of the covariant observable are simply equal to the corresponding preparation uncertainties for the density operator σ, considered as a state. This explains why the constants for our measurement and preparation uncertainties are also just the same. The formal statement is as follows:

Proposition 15. Let M be a covariant phase space observable, generated by a density operator σ, with position and momentum marginals M^Q and M^P. Then for all $\alpha, \beta \in [1, \infty]$:

$$D_\alpha(M^Q, E^Q) = \Delta_\alpha^c(M^Q, E^Q) = D_\alpha(E_\sigma^Q, \delta_0), \tag{44}$$

$$D_\beta(M^P, E^P) = \Delta_\beta^c(M^P, E^P) = D_\beta(E_\sigma^P, \delta_0). \tag{45}$$

Suppose that the product u of these uncertainties is close to its minimum $c_{\alpha\beta}$. Then there is another covariant observable M′ with exactly minimal uncertainty product such that

$$\|M - M'\| \leq 2\sqrt{\frac{u^\gamma - c_{\alpha\beta}^\gamma}{c_{\alpha\beta}'^\gamma - c_{\alpha\beta}^\gamma}}. \tag{46}$$

The finiteness of the uncertainty product implies that E_σ^Q, E_σ^P have finite moments of degree α, $\beta < \infty$. If $\alpha = \infty$ then $\mathrm{supp}\,(E_\sigma^Q)$ is bounded, $\beta < \infty$ and $\int |p|^\beta \, dE_\sigma^P(p) < \infty$.

Proof. The equalities are direct applications of Lemmas 7 and 11. Therefore, for a near-minimal uncertainty product, we can apply Proposition 13 to conclude that there is a density operator σ' with exactly minimal uncertainty product, which is norm close to σ. Then the corresponding covariant observable is also close to M. It remains to show the norm estimate $\|M - M'\| \leq \|\sigma - \sigma'\|_1$. This follows because, for any input state ρ, we have $\|M_\rho - M'_\rho\|_1 = \|\rho * (\sigma - \sigma')\|_1 \leq \|\sigma - \sigma'\|_1$. □

C. The general case

The main result of this paper is the following measurement uncertainty relation.

Theorem 16. *Let M be a phase space observable and $1 \leq \alpha, \beta \leq \infty$. Then*

$$D_\alpha(M^Q, E^Q)\, D_\beta(M^P, E^P) \geq c_{\alpha\beta}\hbar \quad and \tag{47}$$

$$\Delta_\alpha^c(M^Q, E^Q)\, \Delta_\beta^c(M^P, E^P) \geq c_{\alpha\beta}\hbar, \tag{48}$$

provided that in each inequality the quantities on the left-hand side are finite. The constants $c_{\alpha\beta}$ are the same as in Proposition 12.

Note that the proviso rules out the indefinite product $0 \cdot \infty$, along with the utterly uninteresting case $\infty \cdot \infty$. Examples for the indefinite case can be given quite easily. It suffices to combine an ideal position measurement with a random momentum output. Although the statement given here seemingly excludes the indefinite case, it is actually the best one can say about it: If the uncertainty relation is to express quantitatively that not both $D_\alpha(M^Q, E^Q)$ and $D_\beta(M^P, E^P)$ can become small, then we should also have that if one is zero, the other must be infinite. But this statement is implied by the Theorem, which shows that the case $0 \cdot \infty$ does not occur. Of course, we can also conclude that if in some process one uncertainty tends to zero the other has to diverge in keeping with the theorem. That is, the indefinite case as approached from less idealized situations is also covered and interpreted as "$0 \cdot \infty \geq c_{\alpha\beta}$."

The reason this indefinite case does not occur for preparation uncertainty is that we have restricted ourselves to states given by density operators, for which $\Delta_\alpha(E_\rho^{P,Q})$ cannot vanish. Among the so-called singular states (positive normalized expectation value functional on the bounded operators which are not given by density operators) one also finds examples of the indefinite case. Singular states with sharp position assign zero probability to every finite interval of momenta. The momentum is thus literally infinite with probability one, not just a distribution on $\mathbb{R}$ with diverging moments. An example from the literature is the (mathematically cleaned up version of) the state used in the paper of Einstein, Podolsky, and Rosen, where the difference of the two position observables is supposed to be sharp, and accordingly the conjugate difference of momenta is infinite. This also implies that all measurement outcomes seen by Alice or Bob for Q_1, Q_2, P_1, P_2 are infinite with probability one. A detailed study is given in Ref. 35. It is all not as strange as it may seem, as one can see if one replaces the EPR state by a highly (but not infinitely) squeezed two-mode Gaussian state. It is then clear that all individual measurements Q_1, Q_2, P_1, P_2 have very broad distributions, and in the limit the probability for any finite interval goes to zero.

The proof will use the Kantorovich dual characterization (22) of Wasserstein metrics, and thereby excludes the case of one infinite exponent. However, both sides of the inequality are continuous at $\alpha \to \infty$, β fixed, so it actually suffices to consider $\alpha, \beta < \infty$, which we will do from now on.

The proof of the Theorem is by reduction to the covariant case, i.e., Proposition 15 combined with Proposition 12. The following Proposition summarizes what we need.

Proposition 17. *Let M be a phase space observable and $1 \leq \alpha, \beta < \infty$. Suppose that $D_\alpha(M^Q, E^Q)$ and $D_\beta(M^P, E^P)$ are both finite. Then there is a covariant observable $\overline{M}$ such that*

$$D_\alpha(\overline{M}^Q, E^Q) \leq D_\alpha(M^Q, E^Q) \quad \text{and} \quad D_\beta(\overline{M}^P, E^P) \leq D_\beta(M^P, E^P).$$

The analogous statement holds for calibration measures Δ_α^c instead of metric distances D_α.

The basic technique for the proof is averaging over larger and larger sets in phase space, and a compactness argument, that asserts that such an averaging process will have a limit. The most convenient general result based on just this idea is the Markov-Kakutani Fixed Point Theorem[36] (Theorem V.10.6). It says that a family of commuting continuous affine isomorphisms of a compact convex set must have a common fixed point. In our case the set in question will be the set of observables with given finite measurement uncertainties, and the transformations are the phase space translations $M \mapsto M^{(q,p)}$ defined as

$$M^{(q,p)}(Z) = W(q,p)^* M(Z + (q,p)) W(q,p) \tag{49}$$

for any measurable set Z in phase space. Note that this combines a Weyl translation with a translation of the argument in such a way that the common fixed points of these translations are precisely the covariant observables.

In order to satisfy the premises of the Markov-Kakutani Theorem, we have to define a topology for which the phase space translations are continuous, and for which the sets of observables with fixed finite uncertainties are compact. As often in compactness arguments this is the only subtle point, and we will be didactically explicit about it. The topology will be the "weak" topology, i.e., the "initial topology" (Ref. 37, Sec. I.§2.3) which makes the functionals

$$M \mapsto u_M(\rho, f) = \int f(q, p) \, dM_\rho(q, p) \quad \text{for } \rho \in \mathfrak{T}(\mathcal{H}), \ f \in \mathcal{C}_0(\mathbb{R}^2), \tag{50}$$

continuous. That is, the neighbourhoods are specified by requiring a finite number of these functionals to lie in an open set. Let I denote the set of pairs (ρ, f) of a density operator ρ on $\mathcal{H}$ and a function $f \in \mathcal{C}_0(\mathbb{R}^2)$ with $0 \leq f \leq 1$. Then for each such pair and every observable M we have $u_M(\rho, f) \in [0,1]$, which we consider as the (ρ, f)-coordinate of a point $M^\square$ in the cube $[0,1]^I$. It is clear that $M^\square$ determines M uniquely: the functional $\rho \mapsto u_M(\rho, f)$ is affine (i.e., respects convex combinations), so there is a unique operator $M'(f)$, with $u_M(\rho, f) = \operatorname{tr} \rho M'(f)$. Since $f \mapsto M'(f)$ is also affine, we can reconstruct the measure M from it so that $M'(f) = \int f(x) \, dM(x)$. We can therefore look at the observables as a subset of $[0,1]^I$. By definition, the weak topology on the set of observables is the one inherited from the product topology on the cube. By Tychonov's Theorem this is a compact set. Hence this theorem, which embodies the Axiom of Choice, will be the source for all compactness statements about observables in the sequel.

From the proof that $M \mapsto M^\square$ is injective it is clear that most points in $[0,1]^I$ do not correspond to observables. This suggests to single out the subset $[0,1]^I_{obs} \subset [0,1]^I$ of points which are affine in both ρ and f. Note that an affinity condition like $\lambda u_M(\rho_1, f) + (1 - \lambda) u_M(\rho_2, f) - u_M(\lambda \rho_1 + (1 - \lambda)\rho_2, f) = 0$ involves only three coordinates at a time. Therefore, the left-hand side of this equation is continuous, and the subset on which it is true is closed as the inverse image of $\{0\}$ under a continuous function. Since the arbitrary intersection of closed sets is closed, we conclude that $[0,1]^I_{obs}$ is compact, as a closed subset of a compact set. A similar argument shows that the Weyl translations are continuous on $[0,1]^I$. Indeed, to make any finite number of coordinates of a Weyl-translate $M^{(q,p)}$ lie in a specified open set, we only need to shift every ρ and f in this neighborhood description to find an appropriate condition on M.

However, $[0,1]^I_{obs}$ is not exactly the set of observables, because it also contains the zero element. What we get from an arbitrary point $M^\square \in [0,1]^I_{obs}$ is an operator valued measure M, which however need not be normalized. The subset of normalized observables, i.e., those which formally satisfy $u_M(\rho, 1) = 1$ is *not* closed, simply because $1 \notin \mathcal{C}_0(\mathbb{R}^2)$. One can define the normalization operator

for every $M^\square \in [0, 1]^I_{obs}$ as

$$M(1) = \sup_{f \leq 1} \int f(q, p) \, dM(q, p),\tag{51}$$

since the net of functions $f \in \mathcal{C}_0(\mathbb{R}^2)$ is directed, and so in the weak operator topology the limit of any increasing net in $\mathcal{C}_0(\mathbb{R}^2)$, which pointwise goes to 1 is the same. However, this limit is not a continuous function in the weak topology on observables, and may well be strictly smaller than $\mathbb{1}$. In fact, it is easy to construct sequences of normalized observables which converge to zero: it is enough to shift any observable to infinity, i.e., to take $N_{(q, p)}(Z) = M(Z + (q, p))$ without the Weyl operators used in (49). The region where the probability measure $\text{tr} \, \rho M(\cdot)$ is mostly concentrated will thus be shifted away from the region where a function $f \in \mathcal{C}_0(\mathbb{R}^2)$ is appreciably different from zero, with the consequence that $u_{N_{(q,p)}}(\rho, f) \to 0$ for all ρ and f.

This normalization problem can be shifted, but not resolved, by allowing instead of $\mathcal{C}_0(\mathbb{R}^2)$ a larger algebra $\mathcal{A}$ of continuous functions, such as those going to a constant at infinity (thus $1 \in \mathcal{A}$), or even all bounded continuous functions. The problem is then that an observable defined in terms of bilinear functionals u_M with $f \in \mathcal{A}$ define measures not on the phase space $\mathbb{R}^2$, but on a compactification of $\mathbb{R}^2$, which can be understood as the set of pure states of $\mathcal{A}$. In the examples mentioned these are the one point compactification and the Stone-Čech-compactification, respectively. So we have the choice of (a) using $\mathcal{C}_0(\mathbb{R}^2)$, for which the set of normalized observables is not compact or (b) using some algebra $\mathcal{A} \supset \mathcal{C}_0(\mathbb{R}^2)$, for which we may get measures with a positive weight at infinity. The connection of these two points of view is clarified by considering a sequence of observables which converges to zero in the sense of the previous paragraph. The missing normalization of the limit then simply shows up as a positive contribution from the compactification points. Thus we get a normalized observable, but the probability to find a result on ordinary (uncompactified) phase space is zero. The key point of our proof will thus be to show that this phenomenon cannot happen, provided that both uncertainties are fixed to be finite.

The principles used for calibration and metric uncertainty are rather similar, so we largely treat these cases in parallel. Throughout, we keep the exponents $1 \leq \alpha, \beta < \infty$ fixed. Moreover, we fix some uncertainty levels Δ_Q and Δ_P, and in the calibration case some parameters $\varepsilon_Q, \varepsilon_P > 0$. We then consider the sets $\mathcal{M}$ and $\mathcal{M}_c$ of (not necessarily normalized) positive operator valued measures on $\mathbb{R}^2$ defined by the membership conditions

$$M \in \mathcal{M} \Leftrightarrow \quad D_\alpha(M^Q, E^Q) \leq \Delta_Q \quad \text{and} \quad D_\beta(M^P, E^P) \leq \Delta_P,\tag{52}$$

$$M \in \mathcal{M}_c \Leftrightarrow \quad \Delta_\alpha^{\varepsilon_Q}(M^Q, E^Q) \leq \Delta_Q \quad \text{and} \quad \Delta_\beta^{\varepsilon_P}(M^P, E^P) \leq \Delta_P.\tag{53}$$

Moreover, we denote by $\mathcal{N} \subset \mathcal{M}$ and $\mathcal{N}_c \subset \mathcal{M}_c$ the respective subsets of normalized observables. Our aim is to show that these are weakly compact, by first showing that $\mathcal{M}$ and $\mathcal{M}_c$ are compact and then that the normalized subsets are closed under weak limits.

Proposition 18. The sets $\mathcal{M}$ and $\mathcal{M}_c$ are weakly compact and convex.

Proof. The techniques for the two cases are similar, and are based on a description of the respective sets as the sets of (not necessarily normalized) observables satisfying some set of weakly continuous linear constraints derived from (52) resp. (53). We begin with $\mathcal{M}$, using the Kantorovich dual description of D_α from the equality in (22). Including the supremum (26) over states, and using Lemma 5 the inequality $D_\alpha(M_\rho^Q, E_\rho^Q) \leq \Delta_Q$ becomes equivalent to the condition that for all ρ and all $\Psi, \Phi \in \mathcal{C}_0(\mathbb{R})$ satisfying $\Phi(y) - \Psi(x) \leq D(x, y)^\alpha$, we have

$$\int \Phi(y) \, dM_\rho^Q(y) - \int \Psi(x) \, dE_\rho^Q(x) \leq \Delta_Q^\alpha .\tag{54}$$

Indeed the supremum over Φ, Ψ, ρ of the left-hand side is just $D_\alpha(M^Q, E^Q)^\alpha$. We can further rewrite this as

$$\int \Phi(y)\chi(p) \, dM_\rho(x, p) \leq \Delta_Q^\alpha + \int \Psi(x) \, dE_\rho^Q(x),\tag{55}$$

where $\chi \in \mathcal{C}_0(\mathbb{R})$ and $0 \leq \chi \leq 1$. The validity of this inequality for all χ, Φ, Ψ with the specified conditions is still equivalent to (52). Moreover, $\Phi(y)\chi(p) \in \mathcal{C}_0(\mathbb{R}^2)$, so that left-hand side depends on M continuously with respect to the weak topology. Of course, the momentum part is treated in the same way, together proving compactness of $\mathcal{M}$. Convexity is obvious, because each of these constraints is linear.

For the calibration case let us write out the definition of $\mathcal{M}_c$. The conditions are

$$\text{tr}\Big(\rho(Q - q)^\alpha\Big) \leq \varepsilon_Q^\alpha \quad \Rightarrow \int |q' - q|^\alpha \, dM_\rho(q', p') \leq \Delta_Q^\alpha, \tag{56}$$

$$\text{tr}\Big(\rho(P - p)^\beta\Big) \leq \varepsilon_P^\beta \quad \Rightarrow \int |p' - p|^\beta \, dM_\rho(q', p') \leq \Delta_P^\beta . \tag{57}$$

Here p, q, ρ are arbitrary, and the left-hand side of these implications (which do not contain M) only serve to select a subset of parameters for which the right-hand side is to hold. Now the integrals on the right-hand side do involve unbounded functions not in $\mathcal{C}_0(\mathbb{R}^2)$, so are not directly linear conditions on functionals of the form (50). However, the condition in (56) can equivalently be described as

$$\int f(q', p') \, dM_\rho(q', p') \leq \Delta_Q^\alpha \quad \text{for all } f \in \mathcal{C}_0(\mathbb{R}^2) \text{ with } f(q', p') \leq |q' - q|^\alpha. \tag{58}$$

With a similar rewriting of the momentum conditions we get inequalities $u_M(\rho, f) \leq \Delta_Q^\alpha$ on weakly continuous functionals, so that the set $\mathcal{M}_c$ is indeed weakly compact. $\qquad\square$

We now have to show that the respective normalized sets are closed. The basic idea is to use the fact that there is some unbounded function, which has a uniform finite upper bound on the set of measures under consideration. Therefore, probability cannot "sneak off to infinity." This idea (in the case of scalar measures) is made precise in the following Lemma.

Lemma 19. Let $(\mu_i)_{i \in I}$ denote a weakly convergent net of probability measures on $\mathbb{R}^2$, characterized as positive functionals on $\mathcal{C}_0(\mathbb{R}^2)$, which are normalized in the sense that $\sup_{f \leq 1} \int f(x) \, d\mu_i(x) = 1$ for all i. Let $h : \mathbb{R}^2 \to \mathbb{R}_+$ be a continuous function diverging at infinity, so that $(1 + h)^{-1} \in \mathcal{C}_0(\mathbb{R}^2)$, and assume that the expectations of h are uniformly bounded in the precise sense that there is a constant C, independent of i, such that

$$f \in \mathcal{C}_0(\mathbb{R}^2) \; \& \; f \leq h \; \Rightarrow \forall_{i \in I} \int f(x) d\mu_i(x) \leq C.$$

Then the weak limit $\mu = \lim_i \mu_i$ is also normalized.

Proof. Clearly, the functions $(1 + \lambda h)^{-1}$ go to 1 as $\lambda \to 0$. Moreover,

$$1 - \int \frac{d\mu(x)}{1 + \lambda h(x)} = 1 - \lim_i \int \frac{d\mu_i(x)}{1 + \lambda h(x)} = \lim_i \sup_{f \leq 1} \int \Big(f(x) - \frac{1}{1 + \lambda h(x)}\Big) d\mu_i(x)$$

$$= \lim_i \sup_{f \leq 1} \left\{ \lambda \int \frac{f(x)h(x)}{1 + \lambda h(x)} \, d\mu_i(x) + \int \frac{f(x) - 1}{1 + \lambda h(x)} d\mu_i(x) \right\}$$

$$\leq \lambda C + 0. \tag{59}$$

Here the supremum is over continuous functions f with compact support, so that $fh \in \mathcal{C}_0(\mathbb{R}^2)$. Hence, as $\lambda \to 0$, we get $\int (1 + \lambda h)^{-1} d\mu \to 1$, and μ is normalized. $\qquad\square$

The operator valued version follows in a straightforward way. The normalization operator of a general positive operator valued measure on $\mathbb{R}^2$ is, by definition, the operator $M(\mathbb{R}^2)$ such that, for all density operators ρ

$$\text{tr}\,\rho M(\mathbb{R}^2) = \sup_{f \leq 1} \int f(q, p) \, dM_\rho(q, p) \,, \tag{60}$$

where the limit is over the increasing net of functions $f \in \mathcal{C}_0(\mathbb{R}^2)$ with $f \leq 1$. Observables are the normalized operator measures, i.e., those with $M(\mathbb{R}^2) = \mathbb{1}$. Then we can state the following operator valued version of the previous Lemma:

Corollary 20. Let $(M_i)_{i \in I}$ denote a weakly convergent net of observables on $\mathbb{R}^2$, and assume that there is a density operator ρ without eigenvalue zero, and a continuous function $h : \mathbb{R}^2 \to \mathbb{R}_+$ diverging at infinity, such that $\int h(q, p)\, dM_{i, \rho}(q, p) \leq C < \infty$ for a constant independent of i. Then the weak limit of the sequence is also normalized.

Indeed, we can just apply the previous Lemma to conclude that $\operatorname{tr} \rho(\mathbb{1} - M(\mathbb{R}^2)) = 0$, which implies $M(\mathbb{R}^2) = \mathbb{1}$ because ρ has dense range. We note that the same argument holds if the condition is met not for a single ρ but for a family of states ρ_k with bounds C_k possibly depending on k, provided that the union of the ranges of the ρ_k is dense.

Proposition 21. The sets $\mathcal{N}$ and $\mathcal{N}_c$ of observables, defined after Eqs. ([52](#)) and ([53](#)) are weakly compact.

Proof. In both cases, we will apply the corollary with the same function $h(q, p) = |q|^\alpha + |p|^\beta$. Now consider ρ to be a Gibbs state of the harmonic oscillator. Its preparation uncertainties $D_\alpha(E_\rho^Q, \delta_0)$ and $D_\beta(E_\rho^P, \delta_0)$ are finite for all $\alpha, \beta < \infty$. Then for any measure in $\mathcal{N}$ the triangle inequality for the metric implies $D_\alpha(M_\rho^Q, \delta_0) \leq D_\alpha(M_\rho^Q, E_\rho^Q) + D_\alpha(E_\rho^Q, \delta_0) \leq \Delta_Q + D_\alpha(E_\rho^Q, \delta_0)$. Therefore, on $\mathcal{N}$ we have the uniform bound

$$\int h(q, p)\, dM_\rho(q, p) \leq C = (\Delta_Q + D_\alpha(E_\rho^Q, \delta_0))^\alpha + (\Delta_P + D_\beta(E_\rho^P, \delta_0))^\beta, \tag{61}$$

showing that the limit of any weakly convergent sequence from $\mathcal{N}$ will be normalized according to the corollary. It is also in $\mathcal{M}$ due to Proposition 18, hence in $\mathcal{N}$. It follows that $\mathcal{N}$ is a closed subset of the compact set $\mathcal{M}$ and hence compact.

In the calibration case we have to do some additional work, since we have assumptions only about either position calibrating states which are ε_Q-concentrated, or momentum calibrating states which are sharp in momentum. However, from such knowledge we can also infer something about averages of the state over some translations. So let ρ_Q be a Gaussian position calibrating state, say, with $D_\alpha(E_{\rho_Q}, \delta_0) \leq \varepsilon_Q$, so that we can conclude $D_\alpha(M_{\rho_Q}^Q, \delta_0) \leq \Delta_Q$. Consider the phase space translates $\rho_Q(q, p)$ of these states, which satisfy the calibration condition at the point q, and hence

$$D_\alpha(M_{\rho_Q(q,p)}^Q, \delta_q)^\alpha \leq \Delta_Q^\alpha \tag{62}$$

Now consider some probability density f on phase space and the state $\rho = f * \rho_Q = \int f(q, p)\rho_Q(q, p)dqdp$. Then by joint convexity of D_α in its arguments we have $D_\alpha(M_\rho^Q, f^Q)^\alpha \leq \Delta_Q^\alpha$, where f^Q is the position marginal of f. Hence the calibration condition forces $D_\alpha(M_\rho^Q, \delta_0)^\alpha \leq [\Delta^Q + D_\alpha(f^Q, \delta_0)]^\alpha$ uniformly with respect to $M \in \mathcal{N}_c$. A similar relation follows for the averages $g * \rho_P$ of momentum calibrating state. What we therefore need to draw the desired conclusion are the following: A position calibrating state ρ_Q as described, and a momentum calibrating state ρ_P, together with some densities f, g in phase space such that $\rho = f * \rho_Q = g * \rho_P$, and this state has no zero eigenvalues. If we take all these objects Gaussian, they are described completely by their covariance matrices and the "$*$" operation corresponds to addition of covariance matrices. Therefore we can just choose appropriate covariance matrices for ρ_Q and ρ_P, and choose f as the Wigner function of ρ_P and g as the Wigner function of ρ_Q. The covariance matrix of ρ is then the sum of those for ρ_Q and ρ_P, and clearly does not belong to a pure state. Consequently, ρ corresponds to an oscillator state with strictly positive temperature, and hence has no zero eigenvalues. From the estimates given, it is clear that for this ρ a bound of the form ([61](#)) holds, so $\mathcal{N}_c$ is also weakly compact. ☐

Summary of proof: Applying the Markov-Kakutani fixed point theorem to the transformations $M \mapsto M^{(q, p)}$ acting on the convex compact sets $\mathcal{N}$ and $\mathcal{N}_c$, respectively, proves Lemma 17. Combining this with the results on the covariant case (Proposition 15) gives the theorem.

V. DISCUSSION OF THE CONSTANTS

A. Overview

In this section, we give a brief discussion of the constants $c_{\alpha\beta} = c_{\beta\alpha}$ which appear in both the preparation and the measurement uncertainty relations. Figure 1 gives the basic behaviour. The methods for arriving at these plots will be described below.

Since for a probability measure the α-norms increase monotonically, we have that $D_\alpha(\mu, \delta_y)$ is increasing in α. Hence the constants $c_{\alpha\beta}$ are increasing in α and β. For every pair of finite values we can use a Gaussian trial state, for which all moments are finite. Therefore, $c_{\alpha\beta} < \infty$. It is interesting to discuss also the limit in which one of the exponents diverges. For $\beta \to \infty$, the β-norm goes to the ∞-norm, i.e., the supremum norm. This is only finite (say L) for probability distributions with bounded support, namely, the interval $[-L, L]$. The limit $c_{\alpha, \infty} = \lim_{\beta\to\infty} c_{\alpha\beta}$ thus makes a statement how small the α-norm of a quantum position distribution can be when the momentum is confined to the interval $[-1, 1]$. As a family of trial states with finite α-norm we can take the smooth functions on the interval which vanish with all their derivatives at the boundary. Hence $c_{\alpha\infty} < \infty$ for all $\alpha < \infty$. This case is of interest when particles are prepared by passing through a slit: This will strictly bound the initial position distribution, and hence implies a lower bound on the spread of the momentum distribution, which for free particles is essentially the same as the ballistically scaled position distribution at large times as detected by a far away screen. If the profile of the beam in the slit is uniform, like a piece of a plane wave, the Fourier transform will be of the form $\sin(x)/x$, for which even the first moment diverges. Hence the uncertainty relations in this case describe how small the lateral divergence of a beam can be made and how to choose the optimal beam profile for that. The free parameter α allows the optimization to concentrate either on the center or on the tails of the distribution. Of course, taking both exponents to be infinite is asking too much: There are no Hilbert space vectors which have strictly bounded support in both position and momentum. Hence $\lim_{\alpha\to\infty} c_{\alpha\infty} = \infty$.

B. Hirschman's lower bound

A good lower bound on $c_{\alpha\beta}$ comes from the work of Hirschman:[6] He derived uncertainty relations with general exponents from the entropic uncertainty relations $H(E_\rho^Q) + H(E_\rho^P) \geq \log(e\pi)$, where $H(\rho) = -\int \rho(x) \log \rho(x) \, dx$ denotes the Shannon entropy of a probability distribution ρ with respect to Lebesgue measure. At the time of Hirschman's work, the best constant in this in equality was only conjectured, and proven only later in Ref. 38. We recapitulate Hirschman's nice argument, if only to free it of the unnecessary restriction $\alpha = \beta$ which he made. This was lifted in Ref. 31, although these authors were more interested in further generalizations than in finding tight constants. The basic idea of Hirschman is to use variational calculus to maximize the entropy among all probability distributions with given αth moment $M_\alpha = \int |x|^\alpha \rho(x) \, dx$. This gives a probability density proportional to $\exp(-\zeta |x|^\alpha)$, where the Lagrange multiplier ζ can be determined explicitly from M_α. It turns out that for the maximizing state the entropy is of the form $(1/\alpha) \log M_\alpha = \log D_\alpha(\rho, \delta_0)$ plus an additive constant $A(\alpha)$ independent of M_α. Since for an arbitrary distribution with this moment the entropy must be lower, we get the inequality

$$H(\rho) \leq \log D_\alpha(\rho, \delta_0) - A(\alpha), \tag{63}$$

$$A(\alpha) = \frac{\log(e\alpha)}{\alpha} + \log\left(2\Gamma(1 - \frac{1}{\alpha})\right). \tag{64}$$

Since the entropy does not change under translation, we may also replace the point zero in (63) by any other one, like the one minimizing (13). That is, we may replace D_α by the α-spread. Combining the entropic uncertainty relation with (63) then gives

$$\log(\Delta_\alpha(E_\rho^Q)\Delta_\beta(E_\rho^P)) \geq H(E_\rho^Q) + H(E_\rho^P) + A(\alpha) + A(\beta)$$

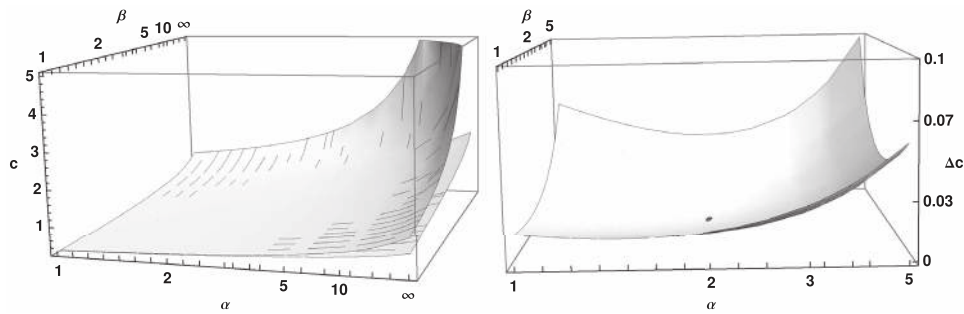$$\geq \log(e\pi) + A(\alpha) + A(\beta) = \log c_{\alpha\beta}^H \tag{65}$$

FIG. 2. Left panel: Hirschman's lower bound in comparison with the exact bound. Scaling of axes as in Fig. 1. Right panel: Difference of the two functions around $\alpha = \beta = 2$ (marked), where the bound is exact.

with the constants

$$c_{\alpha\beta}^H = \frac{\pi \exp(1 - \frac{1}{\alpha} - \frac{1}{\beta}) \alpha^{1/\alpha} \beta^{1/\beta}}{4\Gamma(1 + \frac{1}{\alpha})\Gamma(1 + \frac{1}{\beta})} \ . \tag{66}$$

As Fig. 2 shows, the Hirschman bound is quite good and exact only at $\alpha = \beta = 2$. Indeed, for the bound to be tight the probability densities would have to be $\rho_Q(x) \propto \exp(-\zeta |x|^\alpha)$ and $\rho_P(p) \propto \exp(-\zeta |p|^\beta)$. But this is not compatible with a Fourier pair, except when both exponents are 2. Another feature which is missed by Hirschman's bound is the divergence when both exponents become infinite. Indeed, from the point of view of entropic uncertainty, there is no obstruction against both momentum and position being compactly supported.

The Hirschman techniques gives us no handle on estimating the first excited state $g'_{\alpha\beta}$.

C. Exactly solvable cases

Let us now turn again to the optimal bounds. The ground state problem for $H(\alpha, \beta)$ can be solved in closed form (up to the solution of an explicitly given transcendental equation) only for a few special values. When $\beta = 2$ it is a standard Schrödinger operator for some anharmonic oscillator, which is harmonic for $\alpha = 2$, leading to the well-known value $c_{22} = \frac{1}{2}$. For $\alpha = \infty$ we get a particle in the box $[-1, 1]$, for which the ground state is $\psi(x) = \cos(\pi x/2)$. This leads to $c_{2,\infty} = \pi/2$. At the other end we have the potential $|Q|$, which we can consider as a linear potential on the half line with Neumann boundary conditions ($\psi'(0) = 0$). This is solved by $\psi(x) = \text{Ai}(x - \lambda)$, where Ai is the Airy function, and $-\lambda \approx 1.0188$ is the first zero of the derivative Ai′, which is also the eigenvalue. Hence $c_{12} \approx 0.3958$. Finally, in the case $\alpha = 2m$, $m \in \mathbb{N}$, $\beta = \infty$, H is a differential operator on the interval $[-1, 1]$ with constant coefficients. Then $\langle \psi, H\psi \rangle$ is finite if ψ is in the domain of P^m considered as an operator on the whole line. This entails that the derivatives up to order $m - 1$ are continuous, and hence vanish at the boundary. Unfortunately, the equations characterizing the linear combinations of exponential functions satisfying the boundary conditions are transcendental with complexity increasing rapidly with m. For $\alpha = 4$, we have to solve the equation $\tan \gamma = -\tanh \gamma$ for the eigenvalue γ^4, and $c_{4\infty} = \gamma \approx 2.365$. For $\alpha = 6$, the relevant solution of the fairly complicated equation is exactly π, and $c_{6\infty} = \pi$. Similarly, and with reasonable effort, one can get $c_{8\infty}$ and $c_{10,\infty}$.

In all cases described here, the excited states, and in particular the first, $g'_{\alpha\beta}$ can be obtained in the same way. Table I summarizes the results.

D. Expansion in oscillator basis

For general exponents a numerical approach which works well for small α, β and for $\alpha \approx \beta$ is to compute the matrix elements of $H(\alpha, \beta)$ in the harmonic oscillator basis, truncated at some level n, and to compute the ground state of the resulting matrix. Since already the coherent bound (corresponding to $n = 0$) is fairly good, even small n gives a fairly good approximation. The

TABLE I. Exactly solvable values of the optimal lower bound.

α	β	$c_{\alpha\beta}$	$c'_{\alpha\beta}$
1	2	0.396	1.376
2	2	$1/2 = 0.500$	$3/2 = 1.500$
2	∞	$\pi/2 = 1.571$	$\pi = 3.142$
4	∞	2.365	3.927
6	∞	$\pi = 3.142$	4.714
8	∞	3.909	5.498
10	∞	4.672	6.279
∞	∞	∞	∞

computation of the matrix elements can be done exactly (in terms of Γ-functions), so the numerical error is practically only on the truncation. The results are shown in Fig. 3.

E. One index infinite

It is apparent from Fig. 3 that for high exponents the approximation in terms of oscillator eigenfunctions becomes unreliable. The case of one infinite exponent is again easier to handle, because instead of a high exponent one just has to implement a support condition. It turns out that for $\beta = \infty$ and all α a good first approximation is the wave function

$$\hat{\Psi}(p) \propto (1 - p^2)_+^\alpha, \tag{67}$$

where x_+ denotes the positive part of $x \in \mathbb{R}$ (i.e., $x_+ = x$ for $x \geq 0$ and $x_+ = 0$ for $x \leq 0$). The αth moment of the position distribution can be evaluated explicitly giving the bound

$$c_{\alpha\infty} \leq \left(\frac{\Gamma\left(\frac{\alpha}{2} + 1\right) \Gamma\left(\alpha + \frac{3}{2}\right)}{\Gamma\left(\frac{\alpha+3}{2}\right)} \right)^{\frac{1}{\alpha}} = \frac{\alpha}{e} + \frac{\ln(4\pi\alpha)}{2e} + \mathbf{o}(1), \quad \text{as } \alpha \to \infty, \tag{68}$$

where the second expression is the Stirling approximation.

One can improve this in a similar way as for the Gaussian trial function, by multiplying (67) with polynomials in p, i.e., by expressing the Hamiltonian in terms of associated Legendre functions. This confirms the close approximation shown in Fig. 4 also for the non-integer values, with an error decreasing exponentially.
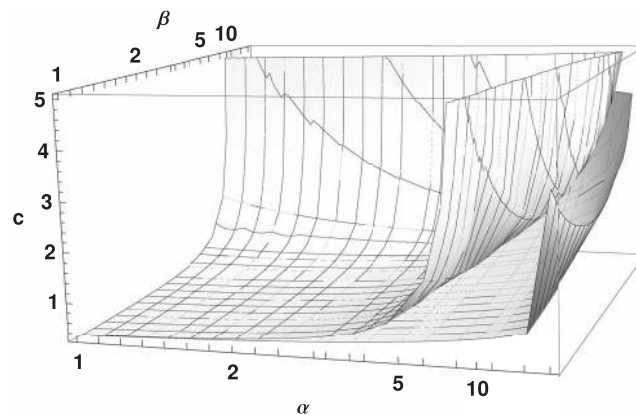

FIG. 3. Comparison of the coherent state uncertainty product (top) and the minimized product using oscillator eigenfunctions up to $n = 200$ (bottom). Axes scaling as in Fig. 1.
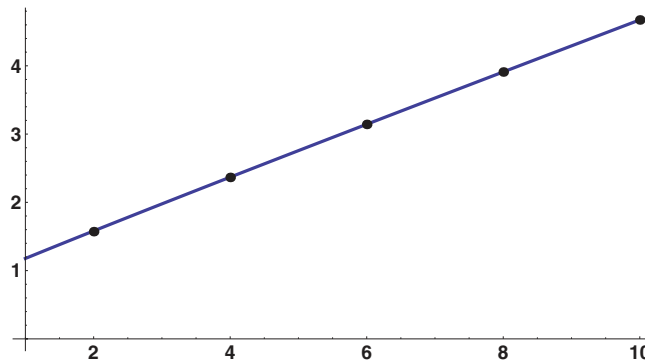
FIG. 4. The upper bound (68). The dots represent the known exact values of $c_{2n,\infty}$ cited above.

VI. EXTENSIONS AND GENERALIZATIONS

A. Other observables

It is clear that the basic definitions of errors can be applied to arbitrary observables. An appropriate metric has to be chosen on the outcome space of each observable. Then, whenever two observables A, B are not jointly measurable, there will be a measurement uncertainty relation, which expresses quantitatively that $D_\alpha(M^A, E^A)$ and $D_\beta(M^B, E^B)$ cannot both be small. For the analogous statement of calibration errors it is needed that A and B are projection valued. As in the case of preparation uncertainty relations there may be many ways of expressing mathematically that ΔA and ΔB "cannot both be small."

The product form $\Delta A \, \Delta B \geq c$ is a rather untypical expression of this sort, which is specific to canonical pairs (Q, P) and their dilation symmetry $Q \mapsto \lambda Q; P \mapsto P/\lambda$. One could also say: The product form is fixed by dimensional analysis. For general A,B one should think of uncertainty trade-offs in terms of an "uncertainty diagram" describing the set of pairs $(\Delta A, \Delta B)$ realizable by appropriate choice of preparation or approximate joint measurement (see Fig. 5 for an illustrative example). An "uncertainty relation" would be any inequality that excludes the origin $\Delta A = \Delta B = 0$ and some region around it. Of course, contrary to the entire textbook literature the Robertson form $\Delta A \Delta B \geq \frac{1}{2}|\langle[A, B]\rangle|$ (like Schrödinger's improvement[39]) is *not* an uncertainty relation in this sense, due to the state dependence of the right-hand side. In fact the *only* cases in which the best constant in an "uncertainty relation" of product form is positive are canonical pairs. In contrast, a relation of the form $(\Delta A)^2 + (\Delta B)^2 \geq c^2$ can always be used to make a non-trivial statement, even if this does not capture the full story contained in the uncertainty diagram.
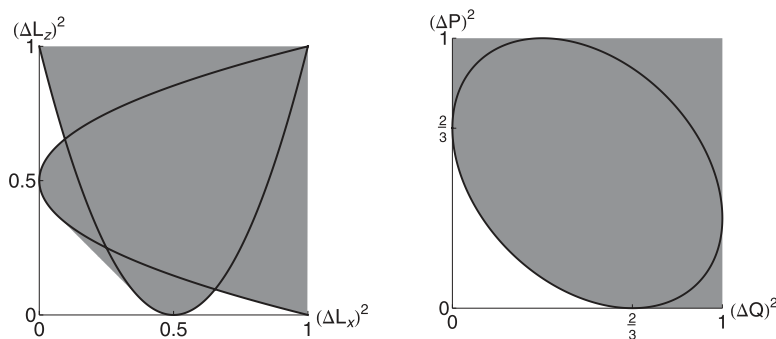

FIG. 5. Left panel: Preparation uncertainty diagram for two angular momentum components of a Spin-1 system. Boundary lines are parabolas indicated and, partly, their convex hull. Results are from Ref. 41. Right panel: Uncertainty diagram for discrete canonical variables and discrete metric. It simultaneously represents the uncertainties for preparation, as well as measurement using either the metric criterion or the calibration criterion. The boundary line is part of the ellipse indicated. The diagram is drawn for dimension $d = 3$.

For any pair of observables A, B, and choices of metrics and exponents, we now have two uncertainty diagrams: one for preparation uncertainty and one for measurement uncertainty. It is a very special feature of the canonical case studied in this paper, perhaps due to the very high symmetry, that the two diagrams coincide. In general they will be different. Indeed (for sharp observables), the origin is included in the preparation uncertainty diagram if and only if A and B have at least *one* common eigenvector, whereas the measurement uncertainty diagram contains the origin if and only if the observables commute, and hence have *a basis* of common eigenvectors. An example in the opposite direction is given by a pair of jointly measurable unsharp observables for which no output distribution is ever concentrated on a point.

This would leave the logical possibility that for sharp observables the preparation uncertainty diagram is always included in the measurement diagram, but much too little is known about either kind of uncertainty to even offer this as a conjecture.

B. General phase spaces, including finite ones

The methods in this paper do extend to discrete canonical pairs, i.e., to pairs of unitaries U, V which commute up to a root of unity ($UV = \exp(2\pi i k/d)VU$). The observables in question are then the spectral measures of U and V. The irreducible representations of this relation (the analogue of the Schrödinger representation studied in this paper) are d dimensional, with "position" U represented as a multiplication operator and "momentum" V the cyclic shift by k steps. Further generalizations allow any locally compact Abelian group X to replace the cyclic group X in this example, with the position observable on $\mathcal{L}^2(X)$ and momentum generated by the shifts, corresponding by a Fourier transform to an observable on the dual group $\widehat{X}$. A joint measurement of these thus has the outcome space $X \times \widehat{X}$, also called phase space. The case $X \cong \widehat{X} \cong \mathbb{R}$ leads back to standard phase space, $X \cong \widehat{X} \cong \mathbb{Z}_d$ is the cyclic case. However, this class also contains the "Fourier series" case $X = \{e^{it} | t \in (-\pi, \pi]\}$, $\widehat{X} = \mathbb{Z}$, and arbitrary products of all these examples, like the phase spaces for quantum systems with many canonical degrees of freedom.

Then the methods of this paper apply, with the following modifications:

- One has to choose a translation invariant metric on each of the spaces X and $\widehat{X}$. On non-compact groups X, $\widehat{X}$ it should have compact level sets, and hence diverge at infinity.
- The harmonic analysis[17] sketched in Remark 2 carries over.[40] In particular, all covariant phase space measurements are parameterized by density operators, and their marginals are formed by convolution as in Eq. (6).
- The properties of Wasserstein metrics under convolution were already considered at the required level of generality in Sec. III.
- The averaging argument carries over, with the only modification that the compactness discussion becomes superfluous in the finite case.
- Hence our main result holds in the form that for any X, any choice of metrics, and any α, β the uncertainty diagrams for (a) preparation uncertainty, (b) measurement uncertainty according to calibration criteria, and (c) measurement uncertainty according to Wasserstein metrics coincide.

To be precise we have shown only that the monotone closures of these diagrams coincide, i.e., the diagrams in which we only care how small uncertainties can be, so that with every point also the positive quadrant above it is included, and white spaces as those near the axes in Fig. 5 are filled in. If we restrict to covariant measurements the diagrams would coincide even without the monotone closure, since the measurement uncertainties are just equal to suitable preparation uncertainties. However, the averaging argument gives only that for every pair of measurement uncertainties there is a pair of, in general, smaller preparation uncertainties, so the monotone closure is needed.

Rather than displaying a zoo of uncertainty diagrams we consider here just the case of a finite cyclic group $X = \mathbb{Z}_d$. For finite outcome sets it is often natural to choose a metric which makes the uncertainty criteria independent of a relabelling of the outcomes, like entropic uncertainty relations. This forces the discrete metric $D(x, y) = (1 - \delta_{xy})$. Since then $D(x, y)^\alpha$ is independent of α the distance functions $D_\alpha(\mu, \nu)$ all express the same quantity, which turns out to be essentially the

variation norm:

$$D_\alpha(\mu, \nu) = \left(\frac{1}{2} \|\mu - \nu\|_1 \right)^{1/\alpha}, \tag{69}$$

$$\Delta_\alpha(\mu) = 1 - \max_x \mu(\{x\}). \tag{70}$$

The proof of (69) is easiest by using the dual characterization of D_1 as a supremum over functions with Lipshitz constant 1, and noting that this is equal to the supremum over all functions $f: X \to [0, 1]$. Consider now a density operator ρ on $\ell^2(X)$. Its position distribution is given by the expectations of the projections $|x\rangle\langle x|$, and by translation invariance it suffices to consider one of them, say $\psi^Q = |0\rangle$. Similarly, the momentum probabilities are given by the expectation of $|\psi^P\rangle\langle\psi^P|$ with $\langle\psi^P|x\rangle = 1/\sqrt{d}$ and its momentum translates. Therefore the uncertainty diagram is

$$\left\{ (\Delta_1(E_\rho^Q), \Delta_1(E_\rho^P)) \right\} = \left\{ (1 - \langle\psi^Q|\rho|\psi^Q\rangle,\ 1 - \langle\psi^P|\rho|\psi^P\rangle) \right\}, \tag{71}$$

where ρ runs over all density operators. Clearly this depends only on the restriction of ρ to the two dimensional subspace generated by ψ^Q and ψ^P. When $d = 2$, this is the whole space and the diagram is a section of the Bloch sphere in some slanted coordinates, i.e., an ellipse. For higher dimensions we get, in addition the point $(1, 1)$ and all segments connecting the ellipse to this point. The monotone closure is always the same. It is easy to see that the ellipse is centered at the point $(\frac{1}{2}, \frac{1}{2})$, and touches the axes at $1 - 1/d$. This completely fixes the diagram (right panel in Fig. 5).

We have covered here only the uncertainty relations which come out of our analysis practically without additional work. Of course, there are many other pairs of observables one would be interested in. For some studies in this direction, we recommend to refer Refs. 42–44

Finally, we note that the special case of the qubit observables has extensively been studied in a separate paper,[45] where additive error trade-off relations are proven that can be tested by the same experiments performed to test an inequality due to Ozawa.

C. More state dependence?

It is clear that the inequality $D_\alpha(M_\rho^Q, E_\rho^Q)D_\beta(M_\rho^P, E_\rho^P) \geq c$, if it were true for all ρ, would be a much stronger theorem than ours, which claims only a relation for the suprema of each of the factors. However, such a relation trivially fails, for example, by choosing for M an ideal position measurement plus the random generation of a momentum output drawn according to E_ρ^P. Rather than touting this as a refutation of Heisenberg's paper, one can look for true relations which are intermediate between the state dependent one and the double supremum considered in this paper. A natural candidate is

$$\inf_M \sup_\rho D_\alpha(M_\rho^Q, E_\rho^Q)D_\beta(M_\rho^P, E_\rho^P) \leq \inf_M \sup_{\rho,\sigma} D_\alpha(M_\rho^Q, E_\rho^Q)D_\beta(M_\sigma^P, E_\sigma^P) = c_{\alpha\beta}. \tag{72}$$

Note that by the argument given above, switching inf and sup on the left-hand side would again trivially produce zero.

The coupled supremum is difficult to compute. Reeb[46] has evaluated at least a restricted version of it, namely, for $\alpha = \beta = 2$ and both M and ρ Gaussian. Indeed this can be done in a straightforward way using Example 6. Let us take ρ as centered and with spreads r_Q and r_P. Take s_Q, s_P be the corresponding ones for the likewise Gaussian state σ defining M. Then we can use the explicit form of the Wasserstein-2 metric to get

$$D_2(M_\rho^Q, E_\rho^Q) = \sqrt{r_Q^2 + s_Q^2} - r_Q = r_Q\left(\sqrt{1 + (s_Q/r_Q)^2} - 1\right).$$

By concavity of the square root, this is bounded above by $s_Q^2/(2r_Q)$, and because $r_Q r_P \geq (1/2)\hbar = s_Q s_P$, the uncertainty product is bounded by $(1/4)(\hbar/2)$. This bound is not tight. Since $f(x) = \sqrt{1 + x^2} - x$ is decreasing, the maximum is taken on the minimal uncertainty r_Q, r_P, i.e., the maximum over all Gaussian inputs is the maximum of $f(x)f(1/x)$, which is attained at $x = 1$. This

translates into $r_Q = s_Q$ and $r_P = s_P$, so the maximum over all Gaussian inputs is

$$\sup_{\rho \text{ Gaussian}} D_2(M_\rho^Q, E_\rho^Q) D_2(M_\rho^P, E_\rho^P) = (\sqrt{2} - 1)^2 \frac{\hbar}{2} \approx .17 \frac{\hbar}{2}. \tag{73}$$

Thus it appears that there might be a proper gap in (72), but the evidence is rather indirect. One should also point out that while the double sup version has a straightforward interpretation coupling two figures of merit, it is not so clear what the coupled sup would be telling us.

D. Finite operating ranges

The figure of merit obtained by taking the worst case over *all* input states is very demanding indeed. In practice, for assessing the performance of a microscope we would not worry about the resolution on objects light years away. Therefore it is reasonable to restrict the supremum to states localized in some finite *operating range*. Shrinking this range to zero would bring us essentially back to the state dependent approach. For a good microscope the operating range should at least be large compared to the resolution. What we considered in this paper is the idealization in which this ratio goes to infinity.

We do plan to make this explicit, and set up uncertainty relations also with finite operating ranges, which in the limit converge to the ones given in this paper. In fact, for the squared noise operator approach this has already been considered by Appleby.[47]

E. Entropic versions

Shortly before this paper was completed, a related paper[43] on entropic state independent noise-disturbance uncertainty relation appeared, providing a kind of entropic version of the idea of calibration. In this paper, the noise $N(\mathcal{M}, A)$ in an approximate measurement $\mathcal{M}$ of a discrete sharp nondegenerate finite-level observable A is quantified (in the form of entropy) by how well it is possible to guess from the measurement outcome distributions the input eigenstate ρ_k from a uniform distributions of such inputs. Similarly, the unavoidable disturbance $D(\mathcal{M}, B)$ quantifies (in the form of entropy) the extent to which the action of $\mathcal{M}$ necessarily reduces the information about which eigenstate σ_l of the observable B, another discrete sharp nondegenerate observable, was initially chosen among a uniform distribution of them. Using the Maassen-Uffink entropic uncertainty relation for preparations,[48] the entropic noise-disturbance trade-off relation then takes the additive form

$$N(\mathcal{M}, A) + D(\mathcal{M}, B) \geq -\log c,$$

where $c = \max_{k,l} \operatorname{tr} \rho_k \sigma_l$ is the same constant as in the preparation relation. It seems to be an open question if there is a measurement which saturates the inequality. It remains to be seen how this approach extends beyond the finite-dimensional case.

VII. CONCLUSION AND OUTLOOK

We have formulated and proved a family of measurement uncertainty relations for canonical pairs of observables. This gives one possible rigorous interpretation of Heisenberg's 1927 statements.

The particular case of canonical variables is special, due to the phase space symmetry of the problem. This leads to the complete equivalence of the possible values of $(\Delta Q, \Delta P)$ between preparation and measurement uncertainty, even when the exponents α, β are varied. In order to establish this we had to generalize standard preparation uncertainty relation to general power means as well, and gave a characterization of the optimal constants in terms of a ground state problem to be solved numerically.

ACKNOWLEDGMENTS

[1] W. Heisenberg, "Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik," Z. Phys. **43**, 172–198 (1927).

[2] E. Kennard, "Zur Quantenmechanik einfacher Bewegungstypen," Z. Phys. **44**, 326–352 (1927).

[3] H. Robertson, "The uncertainty principle," Phys. Rev. **34**, 163–164 (1929).

[4] H. Weyl, *Gruppentheorie und Quantenmechanik* (Hirzel, Leipzig, 1928).

[5] P. Busch, T. Heinonen, and P. Lahti, "Heisenberg's uncertainty principle," Phys. Rep. **452**, 155–176 (2007).

[6] I. I. Hirschman, Jr., "A note on entropy," Am. J. Math. **79**, 152–156 (1957).

[7] M. Ozawa, "Physical content of Heisenberg's uncertainty relation: Limitation and reformulation," Phys. Lett. A **318**, 21–29 (2003).

[8] M. Ozawa, "Uncertainty relations for noise and disturbance in generalized quantum measurements," Ann. Phys. **311**, 350–416 (2004).

[9] M. Ozawa, "Universal uncertainty principle in the measurement operator formalism," J. Opt. B **7**, S672–S681 (2005).

[10] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, "Experimental demonstration of a universally valid error-disturbance uncertainty relation in spin measurements," Nat. Phys. **8**, 185–189 (2012).

[11] L. Rozema, A. Darabi, D. Mahler, A. Hayat, Y. Soudagar, and A. Steinberg, "Violation of Heisenberg's measurement-disturbance relationship by weak measurements," Phys. Rev. Lett. **109**, 100404 (2012).

[12] P. Busch, P. Lahti, and R. Werner, "Quantum root-mean-square error and measurement uncertainty relations," Rev. Mod. Phys. (unpublished); preprint arXiv:1312.4393 (2013).

[13] P. Busch, P. Lahti, and R. Werner, "Proof of Heisenberg's error-disturbance relation," Phys. Rev. Lett. **111**, 160405 (2013).

[14] R. F. Werner, "The uncertainty relation for joint measurement of position and momentum," Quantum Inf. Comput. **4**, 546–562 (2004).

[15] E. Davies, *Quantum Theory of Open Systems* (Academic Press, 1976).

[16] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).

[17] R. F. Werner, "Quantum harmonic analysis on phase space," J. Math. Phys. **25**, 1404–1411 (1984).

[18] K. Husimi, "Some formal properties of the density matrix," Proc. Phys.-Math. Soc. Jpn. **22**, 264 (1940).

[19] G. Cassinelli, E. De Vito, and A. Toigo, "Positive operator valued measures covariant with respect to an irreducible representation," J. Math. Phys. **44**, 4768–4775 (2003).

[20] J. Kiukas, P. Lahti, and K. Ylinen, "Phase space quantization and the operator moment problem," J. Math. Phys. **47**, 072104 (2006).

[21] C. Carmeli, T. Heinonen, and A. Toigo, "Position and momentum observables on $\mathbb{R}$ and on $\mathbb{R}^3$," J. Math. Phys. **45**, 2526–2539 (2004).

[22] C. Carmeli, T. Heinonen, and A. Toigo, "On the coexistence of position and momentum observables," J. Phys. A **38**, 5253–5266 (2005).

[23] B. Simon, "The classical moment problem as a self-adjoint finite difference operator," Adv. Math. **137**, 82–203 (1998).

[24] C. Villani, *Optimal Transport: Old and New* (Springer, 2009).

[25] T. Champion, L. D. Pascale, and P. Juutinen, "The ∞-Wasserstein distance: Local solutions and existence of optimal transport maps," SIAM J. Math. Anal. **40**, 1–20 (2008).

[26] P. Jylhä, "The l^∞ optimal transport: Infinite cyclical monotonicity and the existence of optimal transport maps," Calculus Var. Partial Differ. Equations February, 1–24 (2014).

[27] H. Wiseman, "Extending Heisenberg's measurement-disturbance relation to the twin-slit case," Found. Phys. **28**, 1619–1631 (1998).

[28] E. A. Carlen and W. Gangbo, "Constrained steepest descent in the 2-Wasserstein metric," Ann. Math. (2) **157**, 807–846 (2003).

[29] J. Kiukas and P. Lahti, "A note on the measurement of phase space observables with an eight-port homodyne detector," J. Mod. Opt. **55**, 1891–1898 (2008).

[30] P. Busch and D. B. Pearson, "Universal joint-measurement uncertainty relation for error bars," J. Math. Phys. **48**, 082103 (2007).

[31] M. G. Cowling and J. F. Price, "Bandwidth versus time concentration: The Heisenberg-Pauli-Weyl inequality," SIAM J. Math. Anal. **15**, 151–165 (1984).

[32] G. B. Folland and A. Sitaram, "The uncertainty principle: A mathematical survey," J. Fourier Anal. Appl. **3**, 207–238 (1997).

[33] B. Simon, *Trace Ideals and Their Applications*, London Mathematical Society Lecture Note Series Vol. 35 (Cambridge University Press, Cambridge, 1979), pp. viii+134.

[34] E. B. Davies, *One-Parameter Semigroups*, London Mathematical Society Monographs Vol. 15 (Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1980), pp. viii+230.

[35] M. Keyl, D. Schlingemann, and R. F. Werner, "Infinitely entangled states," Quantum Inf. Comput. **3**, 281–306 (2003).

[36] N. Dunford and J. T. Schwartz, *Linear Operators. Part I*, Wiley Classics Library (John Wiley & Sons Inc., New York, 1988), pp. xiv+858 [General Theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication].

[37] N. Bourbaki, *General Topology, Part I* (Herrmann, Paris, 1966).

[38] W. Beckner, "Inequalities in Fourier analysis," Ann. Math. (2) **102**, 159–182 (1975).

[39] E. Schrödinger, "Zum Heisenbergschen Unschärfeprinzip," Berliner Berichte **19**, 296–303 (1930).

[40] J. Schultz, J. Kiukas, and R. F. Werner, "Quantum harmonic analysis on locally compact Abelian groups," (unpublished).

[41] O. Sachse, "Unschärferelation für diskrete Observable," B.Sc. thesis, Hannover, 2013.

[42] C. Carmeli, T. Heinosaari, and A. Toigo, "Sequential measurements of conjugate observables," J. Phys. A **44**, 285304 (2011).

[43] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde, "Noise and disturbance in quantum measurements: An information-theoretic approach," Phys. Rev. Lett. **112**, 050401 (2014).

[44] P. Coles and F. Furrer, "Entropic formulation of Heisenberg's error-disturbance relation," preprint arXiv:1311.7637 (2013).

[45] P. Busch, P. Lahti, and R. Werner, "Heisenberg uncertainty for qubit measurements," Phys. Rev. A **89**, 012129 (2014).

[46] D. Reeb, private communication, IMS Singapore (Summer 2013).

[47] D. M. Appleby, "Error principle," Int. J. Mod. Theor. Phys. **37**, 2557–2572 (1998).

[48] H. Maassen and J. B. M. Uffink, "Generalized entropic uncertainty relations," Phys. Rev. Lett. **60**, 1103–1106 (1988).