

Heisenberg-style bounds for arbitrary estimates of shift parameters including prior information

Michael J W Hall¹ and Howard M Wiseman

Centre for Quantum Computation and Communication Technology (Australian Research Council), Centre for Quantum Dynamics, Griffith University, Brisbane, QLD 4111, Australia

E-mail: michael.hall@griffith.edu.au

New Journal of Physics **14** (2012) 033040 (22pp)

Received 22 January 2012

Published 28 March 2012

Online at <http://www.njp.org/>

doi:10.1088/1367-2630/14/3/033040

Abstract. A rigorous lower bound is obtained for the average resolution of any estimate of a shift parameter, such as an optical phase shift or a spatial translation. The bound has the asymptotic form $k_I/\langle 2|G| \rangle$ where G is the generator of the shift (with an arbitrary discrete or continuous spectrum), and hence establishes a universally applicable bound of the same form as the usual Heisenberg limit. The scaling constant k_I depends on prior information about the shift parameter. For example, in phase sensing regimes, where the phase shift is confined to some small interval of length L , the relative resolution $\delta\hat{\Phi}/L$ has the strict lower bound $(2\pi e^3)^{-1/2}/\langle 2m|G_1| + 1 \rangle$, where m is the number of probes, each with generator G_1 , and entangling joint measurements are permitted. Generalizations using other resource measures and including noise are briefly discussed. The results rely on the derivation of general entropic uncertainty relations for continuous observables, which are of interest in their own right.

¹ Author to whom any correspondence should be addressed.

Contents

1. Introduction	2
2. Estimation bounds from entropic uncertainty relations	5
2.1. General estimation schemes	5
2.2. Exploiting entropic uncertainty relations	6
2.3. Dependence on prior information	6
3. Examples: discrete shift generators	9
3.1. Phase shift estimation	9
3.2. Time estimation for discrete Hamiltonians	10
4. Examples: continuous shift generators	11
4.1. Time estimation for continuous Hamiltonians	11
4.2. Spatial displacement estimation	12
5. Including the effects of noise	12
6. Discussion	13
Acknowledgments	15
Appendix A. Entropic uncertainty relations involving continuous positive-operator-valued measures	15
Appendix B. Entropy bounds for fixed $\langle G - g \rangle$	20
References	21

1. Introduction

In many measurement scenarios, an environmental variable acts to translate or shift a property such as the optical phase or position of a probe state. Accurate estimation of the shift parameter allows a correspondingly accurate measurement of the environmental variable. For example, interferometric measurements of quantities such as temperature, strain and gravitational wave amplitudes rely on estimation of an optical phase shift. An important aim of quantum metrology is to determine the fundamental bounds on the resolution of such estimates [1, 2].

One useful tool in this respect is the quantum Cramer–Rao inequality [3, 4], which can be used to obtain bounds on the resolution of a shift parameter in terms of the variance of the operator that generates the shift [2–6]. For example, the root mean square error of any unbiased estimate \hat{X} of a shift parameter X , for any fixed value $X = x_0$, satisfies [5, 6]

$$\delta \hat{X}_{x_0} = \langle (\hat{X} - x_0)^2 \rangle^{1/2} \geq 1/(2\sqrt{m}\Delta G), \quad (1)$$

where G is the shift generator, ΔG is the uncertainty of G for the probe state, and m is the number of (independently measured) copies of the probe state.

One can also use the Cramer–Rao inequality to derive resolution bounds in term of other quantities, such as the maximum and minimum eigenvalues of G when they exist [7, 8], or the mixedness of the probe state [9]. While such bounds are universally valid, scientists often wish to obtain bounds in terms of other resources, such as the average energy or mean photon number of the probe state. But, for example, if $G = N$ is the photon number of a single mode field, then ΔN can be arbitrarily large or small relative to the average photon number $\langle N \rangle$.

A simple example which proves this point is a probe state with number distribution $p_0 = 1 - w$, $p_n = w$ for some $n \neq 0$, and vanishing otherwise. Then $\langle N \rangle / \Delta N = [(1 - w)/w]^{1/2}$ ranges over $(0, \infty)$ as w ranges over $(0, 1)$.² Thus, the lower bound (1) does not limit phase resolution in terms of average photon number. This illustrates that the quantum Cramer–Rao inequality cannot always be used to derive bounds in terms of the resources of interest required to achieve a given resolution, in particular when the eigenvalue range is unbounded.

An alternative tool for bounding resolution is the Heisenberg limit [10]. While less developed than the quantum Cramer–Rao inequality, it may be heuristically characterized as an asymptotic lower bound on measurement resolution that scales inversely with the number of resources available and is achievable, up to a constant factor [11–17]. These references use various means to quantify measurement resolution and number of resources, but at this point it suffices to quote the earliest result (obtained numerically) [11]: the root mean square error of a canonical phase estimate on a single mode field, for any applied phase shift ϕ_0 , can asymptotically scale no better than

$$\delta \hat{\Phi}_{\phi_0} \gtrsim 1.38 / \langle N \rangle.$$

This scaling with the average photon number $\langle N \rangle$, rather than with ΔN as in equation (1), has the advantage of providing a necessary condition on the energy resources required for a given phase resolution.

Recently, progress has been made in generalizing the Heisenberg limit to obtain *non*-asymptotic resolution bounds for *arbitrary* estimates of shift parameters. These may be called Heisenberg-style bounds. Firstly, for a shift generator G with a discrete spectrum and finite lowest eigenvalue g_{\min} , Giovannetti *et al* [18] have bounded the root mean square error of any estimate, when averaged over two fixed values of the shift parameter, by $0.076 / \langle G - g_{\min} \rangle$ —providing that the fixed values and errors satisfy a particular constraint. Secondly, when G is further restricted to integer eigenvalues, Hall *et al* have obtained the constraint-free bound of $0.559 / \langle G - g_{\min} + 1 \rangle$, for the root mean square deviation of any estimate, when uniformly averaged over *all* values of the shift parameter [19].

Despite this progress, none of the above results characterize the overall performance of a given estimate in the important case that prior information is available about the value of the shift parameter. For example, in a *phase sensing* regime [20], such as gravitational wave detection, the value of an applied phase shift is *a priori* known to lie within some small interval about zero. Hence, only the performance of the estimate over the interval is of interest—it is irrelevant how well or how badly the estimate may perform outside this interval. Indeed, in a stimulating paper, Rivas and Luis [21] recently proposed a phase estimation scheme that improves on the scalings of the above Heisenberg bounds for small phase shifts.

It is therefore of great interest to determine fundamental bounds when prior information is available to be exploited. For example, the quantum van Trees inequality generalizes the quantum Cramer–Rao inequality, to bound the mean square error averaged over the prior probability density $q(x)$ of the shift parameter [6, 22], leading to ΔG in equation (1) being correspondingly replaced by $[\text{Var } G + F_q]^{1/2}$, where $F_q = \int q(x) (d \ln q(x) / dx)^2$ is the Fisher information of $q(x)$. Note however that, for estimates confined to some bounded interval with

² Moreover, if one chooses $w = n^{-3/2}$, then the value of ΔN diverges with increasing n , while $\langle N \rangle$ becomes arbitrarily small.

$q(x)$ uniform over this interval, F_q vanishes and no improvement is obtained over the lower bound in equation (1).

Tsang [23] very recently obtained the first results exploiting prior information in the context of Heisenberg-style bounds. For example, he showed that, for discrete shift generators and a prior density $q(x)$ uniform over any interval of length L , the root mean square error of any estimate is bounded by a function which asymptotically approaches $0.154/\langle G - g_{\min} \rangle$, under a constraint that L is sufficiently large (see also [24]).

The central result of the present paper is a general Heisenberg-style lower bound that takes arbitrary prior information into account. It is constraint-free and does not require the generator to be discrete or to have a finite lowest eigenvalue. In particular, we show that the average mean square deviation, $(\delta \hat{X})^2 = \langle (\hat{X} - X)^2 \rangle$, of any estimate \hat{X} of some shift parameter X , over any prior distribution of X , satisfies

$$\delta \hat{X} \geq \frac{k_I}{c + 2\Delta^{-1} \langle |G - g| \rangle}. \quad (2)$$

Here k_I depends on the prior information available about the shift parameter, G is the shift generator, and g is an arbitrary eigenvalue of G . For continuous spectra, $c = 0$ and $\Delta = 1$, while for discrete spectra $c = 1$ and Δ is the minimal spectral gap of G . Further, if the spectrum of G has a smallest value, g_{\min} , then the factor of 2 in the denominator can be removed for the choice $g = g_{\min}$.

The scaling constant k_I in equation (2) depends on the prior distribution of the shift parameter, and can in principle be arbitrarily small for a sufficiently narrow prior distribution. In particular, for any generator with a discrete spectrum, if the shift parameter X is known *a priori* to be randomly distributed over an interval of length L , it will be shown that $k_I \geq L(2\pi e^3)^{-1/2}$. Thus there is a fundamental ‘relative resolution’ bound for how well the shift can be resolved, relative to the size of the interval to which it is confined. Moreover, with the additional assumption that the state is a number m of identical copies of some probe state, each with discrete generator G , it will be shown that inequality (2) implies that

$$\frac{\delta \hat{X}}{L} \geq \frac{(2\pi e^3)^{-1/2}}{1 + 2m\Delta^{-1} \langle |G - g| \rangle}, \quad (3)$$

with a similar bound conjectured when G has a continuous spectrum. Note that this inequality allows for arbitrary measurements over the m probes, even entangling joint measurements, unlike bounds obtained from the Cramer–Rao inequality, which assumes a fixed measurement. The scaling of equation (3) with m^{-1} contrasts to the $m^{-1/2}$ scaling in equation (1). While neither bound is necessarily achievable for a given probe state and a given m , the scaling as $m^{-1/2}$ in equation (1) is as expected from elementary statistics. Hence we do not expect that our bounds would be tight for a fixed probe state and large m .

Similarly to equation (2), if the spectrum of G has a smallest value g_{\min} then the factor of 2 in the denominator of equation (3) can be removed for $g = g_{\min}$. For example, for phase shifts generated by the photon number operator N of a single mode field, the relative resolution of any estimate over an interval of length L is bounded below by $(2\pi e^3)^{-1/2}/\langle 1 + mN \rangle$, generalizing the result of Hall *et al* [19], which was limited to the case of a completely random phase shift, with $L = 2\pi$.

The results of the paper are obtained via the derivation of suitable entropic uncertainty relations for the shift generator G and the error $\hat{X} - X$ (section 2 and [appendices](#)). Examples are

given for estimates of optical phase, time, and spatial displacements, including generalizations to alternative resource measures such as the support of the energy distribution of the probe state (sections 3 and 4). A further generalization of inequalities (2) and (3) is given which quantifies the effects of noise (section 5), followed by a brief discussion (section 6).

2. Estimation bounds from entropic uncertainty relations

2.1. General estimation schemes

Consider a general shift parameter estimation scheme, in which a probe state ρ_0 undergoes a shift generated by some operator G , to the state $\rho_x = e^{-iGx} \rho_0 e^{iGx}$. A measurement of observable M on the probe then outputs some estimated value \hat{x} of the actual shift value x , and is described by some positive-operator-valued measure (POVM) $\mathcal{M} = \{M_{\hat{x}}\}$ [2–4]. A standard notation will be used in which random variables and operators appear in upper case, and (eigen)values of these quantities appear in lower case. Thus, the shift parameter will be denoted by X , and its estimate by \hat{X} .

If the prior probability density of the shift parameter is denoted by $q(x)$, then the probability density of the error or deviation, $Y = \hat{X} - X$, of the estimated value from the true value, is given by

$$p_Y(y) = \int dx q(x) p_{\hat{X}}(x+y|x) = \int dx q(x) \text{tr}[\rho_x M_{x+y}] = \text{tr}[\rho_0 \overline{M}_y], \quad (4)$$

where the POVM $\overline{M} \equiv \{\overline{M}_y\}$ is defined by

$$\overline{M}_y := \int dx q(x) e^{iGx} M_{x+y} e^{-iGx}. \quad (5)$$

A ‘good’ estimate of the shift parameter will be one for which the error is small on average, i.e. for which $p_Y(y)$ is highly peaked about $y = 0$. This may be quantified by the mean square deviation

$$(\delta \hat{X})^2 := \langle (\hat{X} - X)^2 \rangle = \langle Y^2 \rangle = \int dy y^2 p_Y(y). \quad (6)$$

Note that for periodic shift parameters, such as phase, the integration may be taken over an interval centered on $y = 0$ [19]. The quantity $\delta \hat{X}_{x_0}$ in equation (1) corresponds to the case $q(x) = \delta(x - x_0)$.

Equations (4)–(6) generalize the case of phase estimates considered in Hall *et al* [19], where the prior density was restricted to be uniform, i.e. $q(x) = 1/2\pi$. Under this restriction the POVM \overline{M} is covariant, with $\overline{M}_{x+y} = e^{-iGy} \overline{M}_x (e)^{iGy}$, allowing a connection to be made between $\delta \hat{X}$ and G via an entropic uncertainty relation for canonically conjugate number and phase operators [19]. However, that method fails whenever $q(x)$ is non-uniform, as \overline{M} is no longer covariant. Note also that $q(x)$ is necessarily non-uniform for non-periodic shift parameters.

It turns out that the key to generalizing the approach of Hall *et al* is the extension of existing entropic uncertainty relations for arbitrary discrete observables [25–27], to the case of continuous POVMs, such as \overline{M} in equation (5) above. The necessary extensions are derived in appendix A. It will now be shown how these lead to Heisenberg-style lower bounds for $\delta \hat{X}$, as per inequalities (2) and (3). Examples and generalizations are given in sections 3–5.

2.2. Exploiting entropic uncertainty relations

Suppose that one has an entropic uncertainty relation for the observables G and \overline{M} (the observable corresponding to the POVM \overline{M}) of the form

$$H(G) + H(\overline{M}) \geq \ln K_I$$

for some constant K_I . Here $H(A)$ denotes the Shannon entropy of the measurement distribution of observable A , for a probe in state ρ_0 . Several such uncertainty relations are given in appendix A. In general, K_I depends on the prior information encoded in the prior density $q(x)$, and its form is discussed in section 2.3 below.

From equation (4) the statistics of \overline{M} and Y are identical, and thus the above uncertainty relation can be rewritten as

$$H(\hat{X} - X) = H(Y) \geq \ln K_I - H(G). \quad (7)$$

Furthermore, consider the variational quantity $J = H(Y) + \alpha \langle 1 \rangle + \beta \langle Y^2 \rangle$, where α and β are Lagrange multipliers fixing the normalization of $p_Y(y)$ and the value of $\langle Y^2 \rangle$ respectively. The variational equation $\delta J / \delta p_Y = 0$ leads directly (Chapter 12 of [28]) to the upper bound $H(Y) \leq \frac{1}{2} \ln [2\pi e \langle Y^2 \rangle]$, saturated by the Gaussian distribution $p_Y(y) = (2\pi \langle Y^2 \rangle)^{-1/2} \exp[-y^2/2\langle Y^2 \rangle]$. Combining this bound with equations (6) and (7) yields the lower bound

$$\delta \hat{X} \geq \frac{K_I}{(2\pi e)^{1/2}} e^{-H(G)} \quad (8)$$

for the root mean square deviation of the estimate.

Inequalities (7) and (8) provide information-theoretic bounds on the performance of the estimate, in terms of the entropy of the shift generator for the probe state. This is already useful in contexts where entropy itself can be considered as a resource. Inequality (8) is also useful for determining alternative bounds on resolution, under various constraints on G , as will be discussed in sections 3 and 4.

Finally, as shown in appendix B, if g is an arbitrary eigenvalue of G , then the entropy of G is bounded above by

$$H(G) \leq 1 + \ln [c + 2\Delta^{-1} \langle |G - g| \rangle], \quad (9)$$

where the factor of 2 can be dropped if G has a minimum eigenvalue g_{\min} and $g = g_{\min}$. For discrete generators, $\Delta = \min_{g' \neq g''} |g' - g''|$ is the minimum spectral gap of G , and $c = 1$. For continuous generators, $\Delta = 1$ and $c = 0$. Substitution of equation (9) into equation (8) immediately yields inequality (2), with

$$k_I = (2\pi e^3)^{-1/2} K_I. \quad (10)$$

Before discussing specific examples and generalizations of the generic Heisenberg bound (2), the dependence of the constant K_I on the prior information encoded in $q(x)$ will be examined, yielding a derivation of the relative resolution bound (3) for discrete generators, and an analogous conjectured bound for continuous generators.

2.3. Dependence on prior information

2.3.1. Discrete generators. For a shift generator G with a discrete spectrum, a suitable scaling constant K_I follows from inequality (A.2) of appendix A as

$$\begin{aligned} K_I^{-1} &= \sup_{g,y,\psi} \langle \psi | \Gamma_g \overline{M}_y \Gamma_g | \psi \rangle = \sup_{g,y,\psi} \langle \psi_g | \overline{M}_y | \psi_g \rangle \langle \psi | \Gamma_g | \psi \rangle \\ &= \sup_{g,y} \langle g | \overline{M}_y | g \rangle, \end{aligned}$$

where Γ_g denotes the projector on to the unit eigenspace of eigenvalue g , $|\psi_g\rangle := \Gamma_g|\psi\rangle / \langle\psi|\Gamma_g|\psi\rangle^{1/2}$, and $|g\rangle$ denotes any normalized eigenstate of G . The last equality holds since (i) $|\psi_g\rangle$ is always equal to some $|g\rangle$ by construction, and (ii) the factor $\langle\psi|\Gamma_g|\psi\rangle$ is maximized when $|\psi\rangle = |\psi_g\rangle$. Hence, using equation (5) for \overline{M}_y , one has

$$K_I^{-1} = \sup_{g,y} \int dx q(x) \langle g|M_{x+y}|g\rangle = \sup_{g,y} \int dx q(x) p_g(x+y), \quad (11)$$

where $p_g(x)$ denotes the measurement distribution when the probe state is replaced by eigenstate $|g\rangle$ of G .

Thus, the value of K_I is determined by the maximum possible value of the convolution of the prior probability density $q(x)$ with the measurement distributions $\{p_g\}$. Moreover, noting that

$$\int dx q(x) p_g(x+y) \leq q_{\max} \int dx p_g(x+y) = q_{\max},$$

where q_{\max} denotes the maximum value of the prior probability density $q(x)$, one has

$$K_I \geq 1/q_{\max}. \quad (12)$$

This constraint on K_I leads to the universal relative resolution bound (3) for discrete generators. In particular, an estimate based on (a possibly entangling joint measurement on) m copies of a probe state corresponds to replacing ρ_0 by $\otimes^m \rho_0$ and G by $G_T = G_1 + G_2 + \dots + G_m$, where G_j refers to G for the j th copy. Note that mg is an eigenvalue of G_T , $|G_1 + \dots + G_m - mg| \leq |G_1 - g| + \dots + |G_m - g|$, and $\Delta_T = \Delta$. Hence, choosing $q(x)$ to be uniform over an interval of length L and vanishing elsewhere, $q_{\max} = 1/L$ and inequality (3) follows from relations (2), (10) and (12).

The lower bound in equation (12) can be approached, in principle, if the measurement distribution $p_g(x)$ is sufficiently peaked around some value x_g , for some eigenstate $|g\rangle$ of G . In particular, this allows y to be chosen in equation (11) such that $p_g(x+y)$ is peaked around the maximum value of $q(x)$. Thus, the more the estimate is concentrated around some value, given a system prepared in some eigenstate $|g\rangle$ of G , the closer the constant K_I will be to $1/q_{\max}$, allowing the possibility of approaching the lower bound in equation (3). This possibility is further discussed in section 3.1.

Note finally that the bound in equation (12) is only useful when the prior probability density $q(x)$ is not infinitely peaked. Although q_{\max} will be finite for any physical prior distribution, it is of interest to find stronger bounds not subject to this limitation. For example, note that any probe state corresponding to an eigenstate $|g\rangle$ of G is invariant under shifts generated by G , implying no corresponding estimate can improve on prior knowledge about the shift parameter. It is therefore natural to define an estimate to be ‘ignorance respecting’ if the measurement distribution for any eigenstate of G , $p_g(x)$, is not any better concentrated than the prior probability density $q(x)$, in the standard sense that p_g majorizes q [29]. This implies in particular that $\int dx F(p_g(x)) \leq \int dx F(q(x))$ for any continuous convex function F [29]. Choosing $F(z) = z^2$, and writing $C_r := \int dx r(x)^2$ for probability density r , it follows via equation (11) and the Schwarz inequality that

$$K_I^{-1} = \sup_{g,y} \int dx q(x) p_g(x+y) \leq \sup_g [C_q C_{p_g}]^{1/2} \leq C_q$$

for ignorance-respecting estimates, i.e.

$$K_I \geq \left[\int dx q(x)^2 \right]^{-1}. \quad (13)$$

This is stronger than the lower bound (12), and can be non-trivial even when $q_{\max} = \infty$.

2.3.2. Continuous generators. Similar results hold for a shift generator G with a continuous spectrum. In particular, from equation (A.6) of appendix A, identifying G with Y , a suitable scaling constant K_I follows as

$$K_I^{-1} = \sup_{g,y} \langle g | \overline{M}_y | g \rangle = \sup_{g,y} \int dx q(x) \langle g | M_{x+y} | g \rangle. \quad (14)$$

Here $|g\rangle$ ranges over all (typically degenerate) unit eigenkets appearing in any spectral decomposition of Γ_g (i.e. $|g\rangle \equiv |g, d\rangle$ for some d in an orthogonal expansion $\Gamma_g = \sum_d |g, d\rangle \langle g, d|$, where d is an arbitrary degeneracy index).

It has not been possible at this time to prove a general relative resolution bound for continuous generators, analogous to (3). In particular, for continuous generators the ket $|g\rangle$ is not normalizable, so that $\langle g | M_{x+y} | g \rangle$ in equation (14) does not correspond to some measurement probability density $p_g(x)$. However, it is conjectured that

$$\frac{\delta \hat{X}}{L} \geq \frac{\sqrt{\pi/2e}}{2L \langle |G - g| \rangle} \quad (\text{conjecture}), \quad (15)$$

for such generators, if the prior distribution is uniform over an interval of length L . Here, as always, the factor of 2 in the denominator can be dropped for a bounded spectrum for the choice $g = g_{\min}$. The support for this conjecture arises from a correspondence between the cases of no prior information and covariant estimates, as will now be detailed.

First, if the POVM $\{M_{\hat{x}}\}$ is covariant, then from equation (5) it follows that $\overline{M}_y = \int dx q(x) M_y = M_y$, for any prior distribution $q(x)$. Hence, covariant estimates cannot make use of any prior information. Conversely, any estimate that does make use of prior information must be non-covariant.

Second, for continuous generators, $H(G) + H(M) \geq \ln \pi e$ for any covariant POVM \mathcal{M} from equation (A.11), where the bound is saturated when M is canonically conjugate to G . Hence one may take $K_I = \pi e$ for covariant estimates. Moreover, for m probe states, any covariant estimate of X satisfies the same entropic uncertainty relation with G replaced by $G_T = G_1 + \dots + G_m$. Hence, using equations (2), (10), and $|G_1 + \dots + G_m - mg| \leq |G_1 - g| + \dots + |G_m - g|$, one has the rigorous bound

$$\delta \hat{X}_{\text{cov}} \geq \frac{\sqrt{\pi/(2e)}}{2m \langle |G - g| \rangle}, \quad (16)$$

for covariant estimates, where the expectation value is with respect to ϱ_0 . As usual, the denominator can be replaced by $m \langle G - g_{\min} \rangle$ if G is bounded below. This bound includes covariant estimates based on entangling joint measurements; a class of covariant estimates based on independent measurements is also briefly discussed in appendix A.

Finally, since inequality (16) corresponds to no use of prior information, it can be interpreted in a limiting sense as a resolution bound relative to a prior distribution $q(x)$ which

is uniform over the whole real line. This corresponds to $L \rightarrow \infty$ in equation (15), where more generally the conjecture claims the same relative bound holds for prior distributions uniform over any finite interval of length L .

3. Examples: discrete shift generators

3.1. Phase shift estimation

For the case where the spectrum of the generator G is a subset of the integers one has $e^{iG\phi} = e^{iG(\phi+2\pi)}$, and hence the corresponding shift parameter Φ may be treated as a phase parameter, taking values on the unit circle. It follows from equations (2), (10) and (12) that

$$\delta\hat{\Phi} \geq \frac{(2\pi e^3)^{-1/2} K_I}{\langle 2|G - g| + 1 \rangle} \geq \frac{(2\pi/e^3)^{1/2}}{2\pi q_{\max} \langle 2|G - g| + 1 \rangle}, \quad (17)$$

where $q_{\max} \geq 1/2\pi$ is the maximum value of the prior probability density $q(\phi)$ for Φ . Moreover, if it is known *a priori* that the phase shift is confined to an interval of length L , then $q(x) = L^{-1}$ over the interval and vanishes elsewhere, and hence from equation (3) for m identical systems the relative resolution is bounded by

$$\frac{\delta\hat{\Phi}}{L} \geq \frac{(2\pi e^3)^{-1/2}}{1 + 2m \langle |G - g| \rangle}. \quad (18)$$

As previously, the term $2|G - g|$ in the denominators of equations (17) and (18) may be replaced by $G - g_{\min}$ when the spectrum is bounded below. Hence for a single mode field, with number operator $N \geq 0$, one has $\delta\hat{\Phi} \geq (2\pi/e^3)^{1/2}/\langle N + 1 \rangle$ in the case of no prior information, with $q(\phi) = 1/2\pi = q_{\max}$. However, strong numerical evidence has been given that the numerator $(2\pi/e^3)^{1/2} \approx 0.559$ for this case can be replaced by a best possible value of ≈ 1.376 , which is asymptotically achievable for large $\langle N \rangle$ on a suitable probe state via the canonical phase estimate $M_\phi = (2\pi)^{-1} \sum_{m,n} e^{-i(m-n)\phi} |m\rangle\langle n|$ [19, 30]. Hence it is conjectured that the bounds equations (17) and (18) are not tight, and that a similar replacement can be made in the case of arbitrary prior information.

To exploit any prior information, corresponding to a scaling constant $K_I < 2\pi$ in equation (17), is non-trivial. Note that a covariant estimate is not suitable, as for any such estimate one has $\langle g|M_\phi|g\rangle = (2\pi)^{-1}$ for all eigenstates $|g\rangle$ of G [4, 12], yielding $K_I = 2\pi$ from equation (11). Indeed, equation (11) implies that a necessary condition for exploiting prior information is that *the estimate must return a non-uniform distribution over $[0, 2\pi]$ when some number eigenstate is input as a probe state*. This is counterintuitive, since such eigenstates are invariant under phase shifts and hence cannot generate any useful phase information. However, it must be kept in mind that it is the actual probe state, ρ_0 , rather than a notional probe state, that is relevant for actually estimating the phase.

For example, in the recently proposed phase estimation scheme of Rivas and Luis, applicable to small phase shifts of a single mode field generated by the photon number operator N , the estimate is proportional to the result of a homodyne measurement on the probe state [21]. Hence, if a number eigenstate $|n\rangle$ was to be input as a probe state, such a homodyne measurement would generate the (clearly non-uniform) statistics proportional to the quadrature distribution $|\langle x|n\rangle|^2$. Thus, the Rivas and Luis scheme satisfies the above necessary condition for

exploiting prior information. However, as noted by Rivas and Luis following their equation (29), while their estimate has an arbitrarily low root mean square error $\delta\hat{\Phi}_0$ for a fixed phase shift value of zero, it can only further achieve an error $\delta\hat{\Phi}_\phi \approx \delta\hat{\Phi}_0 \ll L$, for each ϕ in an interval of length L about $\phi = 0$ (corresponding to $\Delta\phi \ll \delta\phi$ in the notation of [21]), if $\delta\hat{\Phi}_0 \gg 1/N_T$, where $N_T \equiv m \text{tr}[N_{Q_0}]$. Thus, unfortunately, noting that the averaged root mean square error $\delta\hat{\Phi}$ over the interval is $\approx \delta\hat{\Phi}_0$ in this case, this scheme does not approach the lower bound in equation (18) above, nor even the numerically optimal bound $\delta\hat{\Phi} \geq 1.376/\langle N + 1 \rangle$ when prior information is *not* exploited [19].

The above derivation establishes ultimate bounds on phase resolution. In particular, for any phase estimation scheme it is impossible to have a scaling better than $1/q_{\max}$, or an asymptotic scaling better than $1/\langle m|G| \rangle$. However, it remains a challenge for further work to determine how closely the above phase estimation bounds can be approached (up to some numerical factor), via a suitable measurement and probe state.

Finally, it is worth noting that while $\langle G \rangle$ is a transparent measure of resources when G is a photon number operator or similar, more generally the quantity $\langle |G - g| \rangle$ may not be. However, it is straightforward to generalize the method used in section 2 to obtain bounds for resolution in terms of other quantities. For example, in the context of phase measurements, consider the shift generator $G = \frac{1}{2}(N_A - N_B)$, where N_A and N_B are the number operators for respective single-mode fields input to a Mach–Zehnder interferometer [31]. If the total number of input photons is bounded by some fixed maximum value, i.e. $N_A + N_B \leq N_{\max}$, then G can only take $2N_{\max} + 1$ distinct values: $0, \pm 1/2, \pm 1, \dots, \pm N_{\max}/2$. Hence $H(G) \leq \ln(2N_{\max} + 1)$, and equations (8) and (12) yield the bounds

$$\delta\hat{\Theta} \geq \frac{(2\pi e)^{-1/2} K_I}{2N_{\max} + 1}, \quad \frac{\delta\hat{\Theta}}{L} \geq \frac{(2\pi e)^{-1/2}}{2mN_{\max} + 1}, \quad (19)$$

for the resolution of the corresponding shift parameter $\Theta \in [0, 4\pi]$, where m and L have the meanings introduced earlier.

3.2. Time estimation for discrete Hamiltonians

Time estimates correspond to the case $G = E/\hbar$, where E is a Hamiltonian operator with lowest eigenvalue ϵ_0 . Thus G generates the time shift operator $e^{-iEt/\hbar}$. For the case of a discrete spectrum, the Heisenberg bound (2) (recalling the factor of 2 can be removed for $g = \epsilon_0$), in combination with relations (10) and (12), yields

$$\delta\hat{T} \geq \frac{k_I}{1 + \langle E - \epsilon_0 \rangle / \Delta_E} \geq \frac{1}{q_{\max}} \frac{(2\pi e^3)^{-1/2}}{1 + \langle E - \epsilon_0 \rangle / \Delta_E}. \quad (20)$$

Here, Δ_E denotes the smallest energy gap between distinct eigenvalues of E (thus, the bound is only useful for $\Delta_E > 0$), and q_{\max} is the largest value of the prior density $q(t)$. Further, if the time shift is *a priori* uniformly distributed over an interval of length τ (which can be arbitrarily large if the system is not periodic), equation (3) yields the scaling bound

$$\frac{\delta\hat{T}}{\tau} \geq \frac{(2\pi e^3)^{-1/2}}{1 + m\langle E - \epsilon_0 \rangle / \Delta_E} \approx \frac{0.089}{1 + m\langle E - \epsilon_0 \rangle / \Delta_E}, \quad (21)$$

where m is the number of identically prepared copies.

When the energy differences $\epsilon_j - \epsilon_k$ are incommensurate, the system will be almost periodic and $q(t)$ must be defined over the whole real line. However, despite the non-existence

of a uniform prior distribution in this case, one can still define covariant time estimates and show, for example, that one cannot typically extract more than 1 bit of information from such an estimate [32]. It is therefore expected that a non-covariant estimate is required to exploit any prior information.

An alternative resource of interest for bounding time resolution, particularly if the spectral structure is complex, is the number of energy eigenstates accessible to the probe state. For example, if the probe state is a D -level system, then the entropy of its energy distribution must satisfy $H(E) \leq \ln D$, implying via the entropic bound (8) that $\delta \hat{T} \geq (2\pi e)^{-1/2} K_I / D$. Hence, via equation (12), one has the relative resolution bound

$$\frac{\delta \hat{T}}{\tau} \geq \frac{(2\pi e)^{-1/2}}{mD} \approx \frac{0.242}{mD} \quad (22)$$

for a time shift uniformly distributed over $[0, \tau]$.

Note, however, that it is reasonable to expect that the actual time resolution of a discrete system will have a strong dependence on the detailed structure of the energy spectrum. Hence, the above bounds may be well below what is actually achievable.

Finally, note that for a prior distribution uniform over an interval of length τ , Tsang has very recently given the lower bound [23]

$$\frac{\delta \hat{T}}{\tau} \geq \frac{0.154\hbar}{\tau \langle E - \epsilon_0 \rangle} \sqrt{1 - \frac{0.329\hbar}{\tau \langle E - \epsilon_0 \rangle}} \quad (23)$$

for the relative resolution, under the constraint that $\tau \langle E - \epsilon_0 \rangle \geq 0.690\hbar$. For $\tau \Delta_E \leq 0.154(2\pi e^3)^{1/2}\hbar \approx 1.73\hbar$ this is asymptotically stronger, as a function of $\langle E - \epsilon_0 \rangle^{-1}$, than the relative resolution bound (21) (with $m = 1$), and is weaker otherwise. Hence it provides an improved relative resolution bound for the case of a sufficiently small energy gap, or a sufficiently small interval τ still satisfying the constraint.

4. Examples: continuous shift generators

4.1. Time estimation for continuous Hamiltonians

For Hamiltonians having a continuous spectrum, equation (2) (recalling the factor of 2 can be dropped for $g = \epsilon_0$), together with equation (10), yields

$$\delta \hat{T} \geq \frac{(2\pi e^3)^{-1/2} K_I}{\langle E \rangle - \epsilon_0}, \quad (24)$$

where K_I is given in equation (14). Further, for any covariant time estimate one has

$$\delta \hat{T}_{\text{cov}} \geq \frac{\sqrt{\pi/(2e)}}{m} \frac{\hbar}{\langle E \rangle - \epsilon_0} \approx \frac{0.76 \hbar / m}{\langle E \rangle - \epsilon_0} \quad (25)$$

from equation (16). Finally, if conjecture (15) is correct, then (again dropping the factor of 2) the relative resolution bound

$$\frac{\delta \hat{T}}{\tau} \geq \frac{\sqrt{\pi/2e} \hbar}{m\tau \langle E - \epsilon_0 \rangle} \quad (\text{conjecture})$$

holds for a prior probability density $q(t)$ uniform over an interval of length τ .

It would be of great interest to determine how closely the bound (25) can be approached, via a canonical time measurement on a suitable probe state. Note that the variance of the canonical time distribution does not exist if the energy distribution of the probe state has a non-zero groundstate component [4]. Hence, such probe states would require a different measure of time resolution—e.g. the ensemble length of the error distribution, $\exp[H(\hat{T} - T)]$, which may be bounded from below via inequalities (7) and (9).

4.2. Spatial displacement estimation

As a final example, consider the case of estimation of the displacement of a quantum system in some direction, corresponding to the generator $G = P/\hbar$ for the momentum in that direction. The general resolution bound

$$\delta \hat{X} \geq \frac{\hbar}{2} \frac{(2\pi e^3)^{-1/2} K_I}{\langle |P - p| \rangle} \quad (26)$$

follows from equations (2) and (10), while for any covariant estimate one has

$$\delta \hat{X}_{\text{cov}} \geq \frac{\hbar}{2} \frac{\sqrt{\pi/(2e)}}{\langle |P - p| \rangle} \quad (27)$$

from equation (16), for the case of a single copy, $m = 1$. The conjecture (15) implies a similar bound for the case of a prior distribution uniform over any finite interval, suggesting that measurement of the position observable Q conjugate to P , with POVM elements $M_q = |q\rangle\langle q|$, is always optimal in this case.

The above bounds are valid for all values of the reference momentum p . However, a variational calculation shows they are strongest when p is chosen to be the median value p_0 of the momentum distribution $p(k)$, i.e. when $\int_{-\infty}^{p_0} dk p(k) = 1/2$. Note that the mean and median values are identical for the case of symmetric distributions.

It is of interest to note that the covariant resolution bound (27) can be weaker or stronger than the Cramer–Rao related bound, $\hbar/(2\Delta P)$, following from equation (1) [2–5]. For example, for probe states with a Gaussian momentum distribution, $p(k) \sim \exp[-k^2/(2\sigma^2)]$, one finds $\Delta P = (\pi/2)^{1/2} \langle |P| \rangle$, implying that the bound (1) is stronger by a factor of 1.05. Conversely, for a probe state with an exponential momentum distribution, $p(k) \sim \exp[-\beta|k|]$, one finds $\Delta P = \sqrt{2} \langle |P| \rangle$, implying that the bound (1) is weaker by a factor of 1.08.

5. Including the effects of noise

The presence of noise is expected to decrease the accuracy of any estimate, and hence to increase the lower bounds in the previous sections. Consider, for example, the very simple case in which independent noise is added to the measurement outcome \hat{X} . Denoting the noise variable by Z , the entropy power inequality [28] and equation (7) imply

$$e^{2H(\hat{X}+Z-X)} \geq e^{2H(\hat{X}-X)} + e^{2H(Z)} \geq K_I^2 e^{-2H(G)} + e^{2H(Z)}.$$

Hence, as per the derivation of equation (8), it follows that

$$\delta \hat{X}_{\text{noisy}} \geq [(\delta \hat{X}_{\text{bound}})^2 + (2\pi e)^{-1} e^{2H(Z)}]^{1/2}, \quad (28)$$

where $\hat{X}_{\text{noisy}} := \hat{X} + Z$ and $\delta \hat{X}_{\text{bound}}$ denotes any of the lower bounds of the previous sections. Noise thus increases the minimum possible resolution.

A more physical approach is to consider processes that add noise directly to the probe state, and to use stronger entropic uncertainty relations which depend on the probe state. For example, for a rank-1 *discrete* generator G , inequality (A.8) of appendix A effectively replaces K_I by $K_I e^{S[\rho_0]}$, where $S[\rho_0]$ is the von Neumann entropy of the probe state. Hence, replacing ρ_0 by the noisy probe state $\Pi(\rho_0)$ for some completely positive map Π , the generic lower bound (2) generalizes to

$$\delta \hat{X} \geq \frac{k_I}{1 + 2\Delta^{-1} \langle |G - g| \rangle_{\Pi}} e^{S[\Pi(\rho_0)]} \quad (29)$$

for such generators, where $\langle \cdot \rangle_{\Pi}$ denotes an average with respect to $\Pi(\rho_0)$. Similarly, the relative resolution bound (3) generalizes to

$$\frac{\delta \hat{X}}{L} \geq \frac{(2\pi e^3)^{-1/2}}{1 + 2\Delta^{-1} \langle |G - g| \rangle_{\Pi}} e^{S[\Pi(\rho_0)]} \quad (30)$$

(one is limited to the case $m = 1$, since $G_T = G_1 + \dots + G_m$ is not rank-1 for $m > 1$). For a rank-1 *continuous* generator G , the covariant bound (16) generalizes to

$$\delta \hat{X}_{\text{cov}} \geq \frac{\sqrt{2\pi/e^3}}{2 \langle |G - g| \rangle_{\Pi}} e^{S[\Pi(\rho_0)]}, \quad (31)$$

via uncertainty relation (A.10) of appendix A. As always, the factor of 2 in the above denominators may be removed for the choice $g = g_{\min}$.

For example, let N be the photon number operator of a single mode field subject to Gaussian noise, where the noise is described by the completely positive map [33]

$$\Lambda(\rho) = \int d^2\alpha (\pi n_{\lambda})^{-1} e^{-|\alpha|^2/n_{\lambda}} D(\alpha)\rho D(\alpha)^{\dagger},$$

and $D(\alpha) = \exp(\alpha a^{\dagger} - \alpha^* a)$ denotes the Glauber displacement operator. The parameter n_{λ} characterizes the average number of photons added to the field, i.e. $\langle N \rangle_{\Lambda} = \langle N \rangle + n_{\lambda}$, while the entropy of the field is bounded, both for pure and mixed states, by [34]

$$S[\Lambda(\rho)] \geq \ln(1 + n_{\lambda}) + n_{\lambda} \ln(1 + 1/n_{\lambda}) \geq \ln(1 + n_{\lambda}).$$

Combining this with the relative resolution bound (30) then yields a ‘noisy’ bound

$$\frac{\delta \hat{\Phi}}{L} \geq \frac{(2\pi e^3)^{-1/2}}{1 + \langle N \rangle / (n_{\lambda} + 1)} \quad (32)$$

for the relative resolution of any phase estimate, for a prior distribution uniform over some interval of width L . It is seen that the resolution becomes poor for sufficiently large noise.

6. Discussion

The results of the paper establish a rigorous, non-asymptotic and constraint-free lower bound for parameter estimation which is in the form of the Heisenberg limit and which takes prior information into account. The fundamental bound (2) implies that asymptotic scaling better than $1/\langle |G| \rangle$ is impossible, while bound (3) for discrete generators further demonstrates that, for shifts randomly distributed over some interval, asymptotic scaling better than $1/m$ is impossible for the relative resolution, where m is the number of probe states (and entangling joint measurements are permitted). Bound (16) for continuous generators implies a similar $1/m$

limit is unavoidable for the case of covariant estimates. It has also been shown how the effects of noise may be quantified in section 5, including resolution bound (32) for phase estimates on a single mode field subjected to Gaussian noise.

Examples have been given for estimates of phase shifts, time shifts and spatial displacement in sections 3 and 4. These sections also give examples of how the basic method of section 2 may be applied to obtain resolution bounds in terms of alternative resources, such as the total available photon number in equation (19) and the energy support of the probe state in equation (22). For the case of discrete generators with a finite minimum eigenvalue, the corresponding relative resolution bound (21) may be stronger or weaker than the recent constrained bound (23) due to Tsang (section 3.2).

The fundamental tool used to obtain the above resolution bounds is (7) for the entropy of the error in the estimate, $H(\hat{X} - X)$. As noted briefly in section 2.2, the exponential of this entropy may be useful as an alternative possible measure of resolution [35]. Further, this measure has tighter corresponding bounds, as it avoids the use of the relation between $\delta\hat{X}$ and $H(\hat{X} - X)$ (which is only saturated for Gaussian distributions), required for obtaining (8).

Another interesting measure of resolution to consider is the mutual information between the shift and its estimate, $H(\hat{X} : X)$ [28]. While mutual information is not dealt with directly in this paper, the relative resolution bounds (3), (18), (19), (21), (22), (30) and (32) do allow an approximate upper bound to be derived for $H(\hat{X} : X)$, for discrete generators, whenever the prior distribution is uniform over some sufficiently large interval. In particular, the number of distributions of width $\delta\hat{X}$ that can be distinguished without error, over an interval of width L , is approximately $L/\delta\hat{X}$. The corresponding mutual information, i.e. the corresponding number of bits that can be encoded by the shift parameter X and distinguished by the estimate \hat{X} [28], is therefore $H(\hat{X} : X) \approx \log_2[L/\delta\hat{X}]$, i.e. the logarithm of the reciprocal of the relative resolution. Thus the above-mentioned relative resolution bounds place an approximate upper bound on the mutual information. For example, for estimates of phase shifts uniform over an interval of length L , generated by the photon number of a single mode field subjected to Gaussian noise, one has the approximate upper bound

$$H(\hat{\Phi} : \Phi) \lesssim \log_2 [1 + \langle N \rangle / (n_\lambda + 1)] + \frac{1}{2} \log_2 [2\pi e^3] \quad (33)$$

for mutual information from equation (32). If the conjectured bound (15) is correct, one may similarly obtain estimates of mutual information for continuous generators.

As noted above, the resolution can scale no better than inversely with the number of probe states, m , even when entangling joint measurements are permitted. As noted in the introduction, this contrasts with the $m^{-1/2}$ scaling of the Cramer–Rao related bound (1). Note that for $m = 1$, the bounds of this paper can be numerically weaker or stronger than (1) (section 4.2).

It has been seen in sections 2–4 that covariant estimates do not exploit any prior information that may be available. Hence it is only possible to approach the generic resolution bounds (2) and (3) via non-covariant estimates. Further, as noted in section 3.1, a necessary condition for making use of prior information is that the measurement scheme must return a non-uniform distribution when some eigenstate of the generator is input as a probe state. While the recently proposed scheme of Luis and Rivas meets this condition, it does not approach the corresponding bound (18) (section 3.1). It therefore remains an important challenge for future work to determine how closely the various lower bounds of this paper can be approached.

It is also hoped that future work will settle the conjectures made regarding the relative resolution bound (15) for continuous generators in section 2.3, the improvement in scaling

factors for phase estimates in section 3.1, and the strong entropic uncertainty bound (A.9) in appendix A.

Finally, it is noted that the extensions of various entropic uncertainty relations to continuous observables, obtained in appendix A, will find application beyond the realm of quantum metrology.

Acknowledgments

This work was supported by the ARC Centre of Excellence CE110001027.

Appendix A. Entropic uncertainty relations involving continuous positive-operator-valued measures

A.1. One continuous observable

On a finite-dimensional Hilbert space, the entropies of two observables A and B , corresponding to finitely-valued POVMs $\{A_j\}$ and $\{B_k\}$, satisfy the entropic uncertainty relation [25, 36]

$$H(A) + H(B) \geq -2 \ln \max_{j,k} \left\| A_j^{1/2} B_k^{1/2} \right\|_{\infty}, \quad (\text{A.1})$$

where $\|X\|_{\infty}$ denotes the largest singular value of X , i.e. the square root of the largest eigenvalue of $X^{\dagger} X$.

To extend this relation to the case where one of the observables is continuously valued, first consider some observable C taking continuous values in some *compact* set, with corresponding POVM $\{C_{\theta}\}$, and partition the range of θ into a finite number of non-overlapping bins $\{P_k\}$ of equal size ϵ . Define the discrete observable C^P associated with the partition via the POVM $\{C_k^P\}$ with $C_k^P := \int_{P_k} d\theta C_{\theta}$. Then, for any probability density $p(\theta)$ of C , there is a corresponding well-defined discrete probability distribution $p_k := \int_{P_k} d\theta p(\theta)$ (equal to the probability of $\theta \in P_k$), and an associated piecewise-continuous probability density $\tilde{p}(\theta)$ given by replacing $p(\theta)$ by its average value over the bin P_k for $\theta \in P_k$, i.e.

$$\tilde{p}(\theta) := \epsilon^{-1} \int_{P_k} d\theta p(\theta) = \epsilon^{-1} p_k \text{ for } \theta \in P_k.$$

Note that the probability of $\theta \in P_k$ is identical for both $p(\theta)$ and $\tilde{p}(\theta)$, implying the latter converges in distribution to the former in the limit $\epsilon \rightarrow 0$. Note also that the entropy of C^P , $H(C^P) = -\sum_k p_k \ln p_k$, can be rewritten using $\sum_k p_k = 1$ as

$$H(C^P) = -\ln \epsilon - \int d\theta \tilde{p}(\theta) \ln \tilde{p}(\theta).$$

The entropic uncertainty relation (A.1) for observables A and C^P gives

$$\begin{aligned} H(A) + H(C^P) &\geq -\max_{j,k,\psi} \ln \langle \psi | A_j^{1/2} C_k^P A_j^{1/2} | \psi \rangle \\ &= -\max_{j,k,\psi} \ln \left[\langle \psi_j | C_k^P | \psi_j \rangle \langle \psi | A_j | \psi \rangle \right] \\ &= -\max_{j,\theta,\psi} \ln \left[\epsilon \tilde{p}_{\psi_j}(\theta) \langle \psi | A_j | \psi \rangle \right] \\ &= -\ln \epsilon - \max_{j,\theta,\psi} \ln \left[\tilde{p}_{\psi_j}(\theta) \langle \psi | A_j | \psi \rangle \right], \end{aligned}$$

where $|\psi_j\rangle := A_j^{1/2}|\psi\rangle/\langle\psi|A_j|\psi\rangle^{1/2}$, and $\tilde{p}_{\psi_j}(\theta)$ is defined analogously to $\tilde{p}(\theta)$ above, with respect to the probability density $p_{\psi_j}(\theta) = \langle\psi_j|C_\theta|\psi_j\rangle$ (and thus converges in distribution to $p_{\psi_j}(\theta)$ in the limit $\epsilon \rightarrow 0$). Using the above expression for $H(C^P)$ then yields, taking the limit $\epsilon \rightarrow 0$,

$$\begin{aligned} H(A) + H(C) &\geq - \sup_{j,\theta,\psi} \ln [\langle\psi_j|C_\theta|\psi_j\rangle\langle\psi|A_j|\psi\rangle] \\ &= - \sup_{j,\theta,\psi} \ln \langle\psi|A_j^{1/2}C_\theta A_j^{1/2}|\psi\rangle \end{aligned} \quad (\text{A.2})$$

$$= -2 \sup_{j,\theta} \ln \|A_j^{1/2}C_\theta^{1/2}\|_\infty, \quad (\text{A.3})$$

whenever $H(C)$ exists, thus generalizing (A.1).

Uncertainty relations (A.2) and (A.3) may be further extended to the case of infinite-dimensional Hilbert spaces, whenever the left hand side exists, by considering the limit of a series of projections of the observables onto finite Hilbert spaces. They similarly extend to the case of a countably infinite POVM $\{A_j\}$, whenever $H(A)$ exists, by considering the limit of the sequence of finite POVMs $\{A_1, A_2, \dots, A_d, 1 - \sum_{j=1}^d A_j\}$ as $d \rightarrow \infty$. Finally, they also extend to the case of a non-compact range of C , whenever the left hand side exists, by representing the range as the limit of a series of compact sets S , and replacing $\{C_\theta\}$ by a corresponding series of POVMs $\{C_\theta, \theta \in S\} \cup \{1 - C_S\}$, where $C_S := \int_S d\theta C_\theta$. Thus, equation (A.2) is valid for any discrete-valued observable A and continuously-valued observable C , whenever $H(A) + H(C)$ exists, and similarly for equation (A.3) if $C_\theta^{1/2}$ is well defined.

For example, for an angular momentum component J_z and its corresponding conjugate angle Φ , inequality in equation (A.2) reduces to the known result

$$H(J_z) + H(\Phi) \geq - \sup_{j,\phi,\psi} \ln [|\langle\psi|j\rangle|^2 |\langle j|\phi\rangle|^2] = \ln 2\pi\hbar,$$

first given by Bialynicki-Birula and Mycielski [37], which is saturated by eigenstates of J_z .

A.2. Two continuous observables

Uncertainty relation (A.3) provides the basis for extending to the case of two continuous-valued observables X and Y corresponding to POVMs $\{X_x\}$ and $\{Y_y\}$ respectively. The procedure is similar to the foregoing. In particular, partitioning the range of x into bins P_j of equal size ϵ , one has the corresponding discrete POVM $\mathcal{A} \equiv \{A_j\}$ with $A_j := \int_{P_j} (d)x X_x$. The entropy of the corresponding observable A , given a continuous probability density $p(x)$ of X , is then that of the discrete distribution $p_j := \int_{P_j} dx p(x) = \epsilon \tilde{p}(x)$, where the second expression defines the piecewise continuous probability density $\tilde{p}(x)$. Substitution into equation (A.3), with $C \equiv Y$ and assuming $Y_y^{1/2}$ is well-defined, gives

$$\begin{aligned} -\ln \epsilon - \int dx \tilde{p}(x) \ln \tilde{p}(x) + H(Y) &\geq -2 \sup_{j,y} \ln \|A_j^{1/2}Y_y^{1/2}\|_\infty \\ &= - \sup_{j,y,\psi} \ln \langle\psi|Y_y^{1/2}A_j Y_y^{1/2}|\psi\rangle \\ &= - \sup_{j,y,\psi} \ln [\epsilon \tilde{p}_{\psi_y}(x) \langle\psi|Y_y|\psi\rangle], \end{aligned}$$

where $|\psi_y\rangle := Y_y^{1/2}|\psi\rangle/\langle\psi|Y_y|\psi\rangle^{1/2}$ and $\tilde{p}_{\psi_y}(x)$ is defined analogously to $\tilde{p}(x)$ above, with respect to the probability density $p_{\psi_y}(x) = \langle\psi_y|X_x|\psi_y\rangle$. Taking the limit $\epsilon \rightarrow 0$ then gives

$$\begin{aligned} H(X) + H(Y) &\geq - \sup_{x,y,\psi} \ln[\langle\psi_y|X_x|\psi_y\rangle \langle\psi|Y_y|\psi\rangle] \\ &= - \sup_{x,y,\psi} \ln\langle\psi|Y_y^{1/2}X_xY_y^{1/2}|\psi\rangle = -2 \sup_{x,y} \ln \|X_x^{1/2}Y_y^{1/2}\|_\infty, \end{aligned} \quad (\text{A.4})$$

whenever the entropies and the relevant square roots are well defined.

Indeed, this uncertainty relation can also be applied in some instances when the square roots are not well defined, via taking appropriate limits. For example, for conjugate position and momentum observables Q and P , with eigenkets $|q\rangle$ and $|p\rangle$ respectively, and any $\epsilon > 0$, define the ‘averaged’ momentum observable \tilde{P} with POVM $\{\tilde{P}_p\}$ via $\tilde{P}_p := (2\epsilon)^{-1} \int_{-\epsilon}^{\epsilon} dk |p+k\rangle\langle p+k|$. Then, $\tilde{P}^{1/2} = (2\epsilon)^{1/2} \tilde{P}_p$ from the spectral theorem, yielding

$$\begin{aligned} \langle\psi|\tilde{P}_p^{1/2}Q_q\tilde{P}_p^{1/2}|\psi\rangle &= |\langle q|\tilde{P}_p^{1/2}|\psi\rangle|^2 = \frac{1}{2\epsilon} \left| \int_{-\epsilon}^{\epsilon} dk \langle q|p+k\rangle\langle p+k|\psi\rangle \right|^2 \\ &\leq \frac{1}{2\epsilon} \int_{-\epsilon}^{\epsilon} dk |\langle q|p+k\rangle|^2 \int_{-\epsilon}^{\epsilon} dk |\langle p+k|\psi\rangle|^2, \end{aligned}$$

where the last line follows from the Schwarz inequality. The first integral evaluates to $2\epsilon/(2\pi\hbar)$, while the second is never greater than unity for any normalized state ψ . Hence, substituting $X = Q$ and $Y = \tilde{P}$ into the first inequality of uncertainty relation (A.4), and taking the limit $\epsilon \rightarrow 0$, yields

$$H(Q) + H(P) \geq \ln 2\pi\hbar. \quad (\text{A.5})$$

Note that the lower bound is not optimal, although it is close to the optimal bound $\ln \pi e\hbar$, saturated by Gaussian pure states [36, 37]. However, the same lower bound is optimal for the related tight uncertainty relation (A.10) below.

More generally, if the POM $\{Y_y\}$ is a continuous projection-valued measure corresponding to some Hermitian operator, then although $Y_y^{1/2}$ is not well defined, an entropic uncertainty relation may be obtained via a similar limiting approach. In particular, in such a case $Y_yY_{y'} = \delta(y-y')Y_y$, implying that $\tilde{Y}_y := (2\epsilon)^{-1} \int_{-\epsilon}^{\epsilon} dz Y_{y+z}$ satisfies $\tilde{Y}^{1/2} = (2\epsilon)^{1/2}\tilde{Y}_y$. Applying the first inequality in equation (A.4) to X and \tilde{Y} yields

$$H(X) + H(\tilde{Y}) \geq - \sup_{x,y,\psi} \ln[\langle\tilde{\psi}_y|X_x|\tilde{\psi}_y\rangle \langle\psi|\tilde{Y}_y|\psi\rangle],$$

with $|\tilde{\psi}_y\rangle := \tilde{Y}_y^{1/2}|\psi\rangle/\langle\psi|\tilde{Y}_y|\psi\rangle^{1/2} = (2\epsilon)^{1/2}\tilde{Y}_y|\psi\rangle/\langle\psi|\tilde{Y}_y|\psi\rangle^{1/2}$. Noting

$$\langle\psi|\tilde{Y}_y|\psi\rangle = (2\epsilon)^{-1} \int_{-\epsilon}^{\epsilon} \langle\psi|Y_y|\psi\rangle \leq (2\epsilon)^{-1},$$

the limit $\epsilon \rightarrow 0$ gives $H(X) + H(Y) \geq - \sup_{x,y,\psi} \ln\langle\psi|Y_yX_xY_y|\psi\rangle/\langle\psi|Y_y|\psi\rangle$. Finally, $Y_y|\psi\rangle$ is always proportional to some unit eigenket of Y , i.e. $Y_y|\psi\rangle = \alpha|y, d\rangle$, where d is a degeneracy index in some orthogonal expansion $Y_y = \sum_d |y, d\rangle\langle y, d|$ of Y_y (noting such expansions are invariant under unitary transformations of the degeneracy basis), and the uncertainty relation

$$H(X) + H(Y) \geq - \sup_{x,y} \ln \langle y|X_x|y\rangle \quad (\text{A.6})$$

immediately follows, where $|y\rangle$ ranges over all unit eigenkets of Y_y . This generalization of equation (A.5), holding for any projection-valued measure $\{Y_y\}$, is of particular relevance to generators with continuous spectra (section 2.3.2).

A.3. One rank-1 observable

When the observable A in relation (A.1) is rank 1, i.e. when $A_j = |a_j\rangle\langle a_j|$ for some set of (not necessarily normalized) kets $\{|a_j\rangle\}$, then one has the stronger uncertainty relation [26]

$$H(A) + H(B) \geq -\ln \max_{j,k} \left\| A_j^{1/2} B_k^{1/2} \right\|_{\infty} + S[\rho], \quad (\text{A.7})$$

where $S[\rho]$ denotes the von Neumann entropy $-\text{tr}[\rho \ln \rho]$ of the density operator ρ generating the statistics of A and B . This recent result, by Coles *et al* [26], generalizes an earlier version by Berta *et al* [27] for the case of two rank-1 projection valued observables A and B .

Using the same methods as in appendix A.1, this relation can be similarly generalized to infinite Hilbert spaces and one continuous observable, to give

$$H(A) + H(C) \geq -\sup_{j,\theta,\psi} \ln \langle \psi | A_j^{1/2} C_{\theta} A_j^{1/2} | \psi \rangle + S[\rho] \quad (\text{A.8})$$

whenever the left hand side exists, for any discrete-valued rank-1 observable A and any continuously-valued observable C .

Unfortunately, one cannot analogously generalize equation (A.4) via the methods of appendix A.2, as these methods rely on use of an observable A which is not rank 1. However, it is conjectured here that such a generalization exists, with

$$H(X) + H(Y) \geq -\sup_{x,y,\psi} \ln \langle \psi | Y_y^{1/2} X_x Y_y^{1/2} | \psi \rangle + S[\rho] \quad (\text{A.9})$$

for any two continuously valued observables X and Y , providing X is rank 1, both entropies exist, and $Y_y^{1/2}$ is well defined.

The above conjecture can be proved for the special case of conjugate position and momentum observables, using an approach of Pegg *et al* [38] in which Q and P are represented by approximating them as discrete rank-1 observables on a D -dimensional Hilbert space and taking the limit $D \rightarrow \infty$. In particular, substituting the discrete observables into equation (A.7) and taking this limit yields

$$H(Q) + H(P) - S[\rho] \geq \ln 2\pi\hbar, \quad (\text{A.10})$$

whenever the left hand side exists. The same method generalizes to the case of conjugate n -vectors \mathbf{Q} and \mathbf{P} , with the right hand side of the above relation being multiplied by n . This result proves the conjecture in equation (47) of [35], which was made on the basis of a semiclassical argument. Note that, in contrast to inequality (A.5), the bound in inequality (A.10) is tight, being saturated in the limit of equilibrium states in the high temperature limit [35].

A.4. Covariant observables

The entropic uncertainty relation

$$H(G) + H(M) \geq \ln \pi e \quad (\text{A.11})$$

will be obtained here, for any observable M covariant with respect to a continuous Hermitian observable G , and also a generalization to estimates based on m repeated measurements, as required for equation (16) of the text.

First, let F denote the observable canonically conjugate to G , with POVM $\{F_x\}$ given by [4, 32]

$$F_x := e^{-iGx} F_0 e^{iGx}, \quad F_0 := (2\pi)^{-1} \sum_d \int dg dg' |g, d\rangle \langle g', d|,$$

where the projection Γ_g has the orthogonal expansion $\Gamma_g = \sum_d |g, d\rangle \langle g, d|$. Now, any density operator ρ may be formally mapped to a density operator ρ^* of a 1-dimensional particle, with position coordinate Q and eigenkets $|q\rangle$, via $\rho^* := \sum_d \int dq dq' |q\rangle \langle q'| \langle q, d| \rho |q', d\rangle$, (where one takes $|g, d\rangle := 0$ for g outside the spectrum of G). By construction, the probability distributions of F and G for ρ are identical to the probability distributions of Q and P for ρ^* , where P is the momentum observable conjugate to Q in units such that $\hbar = 1$. From the known entropic uncertainty relation for Q and P [36, 37], it immediately follows that

$$H(F) + H(G) = H(Q) + H(P) \geq \ln \pi e.$$

Further, the probability distribution of any covariant observable M for some state ρ is equal to the probability distribution of the conjugate observable F for some corresponding state ρ' , where ρ and ρ' have the same probability distribution for the observable G [32]. Hence the above bound also holds with F replaced by M , yielding equation (A.11) as desired.

It is of interest to consider a class of covariant estimates m based on independent measurements of the covariant observable M , made on m respective copies of the probe state. Let $\hat{x} = f(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_m)$ denote the corresponding estimate of the shift parameter, where \hat{x}_j denotes the individual estimate given by the j -th measurement. It will be assumed that the estimate is shift-invariant, i.e. that the function f satisfies the identity

$$f(x_1 + y, x_2 + y, \dots, x_m + y) = f(x_1, x_2, \dots, x_m) + y. \quad (\text{A.12})$$

This is satisfied, for example, by any weighted mean $f = w_1 x_1 + \dots + w_m x_m$ with $w_1 + \dots + w_m = 1$. Differentiating with respect to y and taking $y = 0$ gives the equivalent condition $\sum_j (\partial f / \partial x_j) = 1$.

To show that such an estimate is covariant with respect to the corresponding shift generator $G_T = G_1 + \dots + G_m$, let \mathbf{x} denote the vector (x_1, \dots, x_m) , and define $S_x := \{\mathbf{x} : f(\mathbf{x}) = x\}$. The POVM $\{M_x^J\}$ corresponding to the joint estimate is then given by $M_x^J = \int_{S_x} d\mathbf{x} M_{\hat{x}_1} \otimes \dots \otimes M_{\hat{x}_m}$ and therefore

$$\begin{aligned} e^{-iG_T x} M_x^J e^{iG_T x} &= \int_{S_y} d\mathbf{y} e^{-iG_1 x} M_{y_1} e^{iG_1 x} \otimes \dots \otimes e^{-iG_m x} M_{y_m} e^{iG_m x} \\ &= \int_{S_y} d\mathbf{y} M_{x+y_1} \otimes \dots \otimes M_{x+y_m} \\ &= \int_{S_{x+y}} d\mathbf{y} M_{y_1} \otimes \dots \otimes M_{y_m} = M_{x+y}^J, \end{aligned}$$

where the last line follows via the change of variables $y_m \rightarrow y_m - x$ and using the above identity for f . Hence M^J is covariant with respect to G_T .

Appendix B. Entropy bounds for fixed $\langle |G - g| \rangle$

B.1. Discrete generators

First, consider the special case G where the spectrum of G is some subset of the integers. Now, the maximum entropy of any distribution over the integers for a fixed value of $\langle |n - n_0| \rangle$ corresponds to maximizing the variational quantity

$$J = - \sum_n p_n \ln p_n + \alpha \sum_n p_n + \beta \sum_n |n - n_0| p_n,$$

where α and β are Lagrange multipliers. It is convenient to work with the displaced distribution $q_n := p_{n+n_0}$, for which the variational equation $\delta J / \delta q_n = 0$ has the solution $q_n = A v^{|n|}$, for suitable positive constants A and v determined by the constraints. One easily finds that $A = (1 - v) / (1 + v)$ and $\bar{n} := \langle |n - n_0| \rangle = \sum_n |n| q_n = 2v / (1 - v^2)$. Further, $H_{\max}(G) = - \sum_n q_n [\ln A + |n| \ln v] = - \ln A - \bar{n} \ln v$. Inverting the relation between \bar{n} and v gives $v = (1 + 1/\bar{n}^2)^{1/2} - 1/\bar{n} = e^{-x}$, where $\sinh x = 1/\bar{n}$, and the maximum entropy simplifies to

$$H_{\max}(G) = \ln \left[(\sqrt{\bar{n}^2 + 1} + \bar{n}) \right] + x / \sinh x \leq \ln[2\bar{n} + 1] + 1, \quad (\text{B.1})$$

where the final inequality follows by adding a term $2\bar{n}$ under the square root and noting that $x \leq \sinh x$ (e.g., from the Taylor series expansion of $\sinh x$). This corresponds to the bound (9) with $\Delta = 1 = c$. Further, if the spectrum of G is bounded below by n_{\min} , a similar calculation for the choice $n_0 = n_{\min}$ leads to $H_{\max}(G) \leq \ln(\bar{n} + 1) + 1$ [19], i.e. to replacement of the factor of 2 in equation (B.1) by unity.

More generally, let G have an arbitrary discrete spectrum, and let $\{g_n\}$ denote the distinct eigenvalues of G in increasing order. Thus, the minimum spectral gap is given by $\Delta = \min_j (g_{j+1} - g_j)$. Hence, if $n \geq n_0$ then

$$|g_n - g_{n_0}| = (g_n - g_{n-1}) + \dots + (g_{n_0+1} - g_{n_0}) \geq |n - n_0| \Delta,$$

and the same relation also follows for $n \leq n_0$, implying for any probability distribution $\{p_n\}$ of G that

$$\langle |G - g_{n_0}| \rangle = \sum_n |g_n - g_{n_0}| p_n \geq \sum_n |n - n_0| \Delta p_n = \bar{n} \Delta,$$

where $\bar{n} := \langle |n - n_0| \rangle$ as before. Since inequality (B.1) bounds the maximum possible entropy of any discrete distribution $\{p_n\}$ for a fixed value of \bar{n} , it follows immediately that the entropy of G is bounded by

$$H(G) \leq \ln[2\bar{n} + 1] + 1 \leq \ln[2\Delta^{-1} \langle |G - g_{n_0}| \rangle + 1] + 1.$$

This establishes the bound (9) for an arbitrary discrete spectrum. Again, if the spectrum of G is bounded below by g_{\min} , the factor of 2 may be replaced by unity for the choice $g_{n_0} = g_{\min}$.

B.2. Continuous generators

If G has a continuous spectrum, then a bound on the maximum entropy of G , under the constraint of a fixed value of $\langle |G - g'| \rangle$, is obtained by maximizing the variational quantity

$$J = - \int dg p(g) \ln p(g) + \alpha \int dg p(g) + \beta \int dg |g - g'| p(g)$$

over all probability distributions on the real line. It is convenient to work with the displaced distribution $q(g) := p(g + g')$, having the same entropy as $p(g)$, for which the variational equation $\delta J/\delta q = 0$ yields $q(g) = A e^{-\beta|g|}$. The constants A and β are determined by the constraints to be $A = 1/(2\bar{g})$ and $\beta = 1/\bar{g}$, where $\bar{g} := \langle |G - g'| \rangle = \int dg |g| q(g)$, allowing the maximum possible entropy to be calculated as

$$H_{\max}(G) = - \int dg q(g) [\ln A - \beta|g|] = -\ln A + \beta\bar{g} = \ln[2\bar{g}] + 1,$$

establishing the bound (9) for an arbitrary continuous spectrum. If the spectrum of G is bounded below by g_{\min} , a similar calculation for the choice $g' = g_{\min}$ leads to $H_{\max}(G) \leq \ln(\bar{g} + 1) + 1$, i.e. to replacement of the factor of 2 by unity.

References

- [1] Giovanetti V, Lloyd S and Maccone L 2004 *Science* **306** 1330
- [2] Wiseman H M and Milburn G J 2010 *Quantum Measurement and Control* (Cambridge: Cambridge University Press)
- [3] Helstrom C W 1976 *Quantum Detection and Estimation Theory* (New York: Academic)
- [4] Holevo A S 1982 *Probabilistic and Statistical Aspects of Quantum Theory* (Amsterdam: North-Holland)
- [5] Braunstein S L and Caves C M 1994 *Phys. Rev. Lett.* **72** 3439
- [6] Paris M G A 2009 *Int. J. Quantum Inf.* **7** 125
- [7] Giovanetti V, Lloyd S and Maccone L 2006 *Phys. Rev. Lett.* **96** 010401
- [8] Boixo S, Flammia S T, Caves C M and Geremia J M 2007 *Phys. Rev. Lett.* **98** 090401
- [9] Modi K, Cable H, Williamson M and Vedral V 2011 *Phys. Rev. X* **1** 021022
- [10] Holland M J and Burnett K 1993 *Phys. Rev. Lett.* **71** 1355
- [11] Summy G S and Pegg D T 1990 *Opt. Commun.* **77** 75
- [12] Hall M J W 1993 *J. Mod. Opt.* **40** 809
- [13] Luis A and Peřina J 1996 *Phys. Rev. A* **54** 4564
- [14] Wiseman H M and Killip R B 1998 *Phys. Rev. A* **57** 2169
- [15] Berry D, Wiseman H M and Zhang Z -X 1999 *Phys. Rev. A* **60** 2458
- [16] Berry D W *et al* 2009 *Phys. Rev. A* **80** 052114
- [17] Zwierz M, Pérez-Delgado C A and Kok P 2010 *Phys. Rev. Lett.* **105** 180402
- [18] Giovannetti V, Lloyd S and Maccone L 2011 arXiv:1109.5661v1
- [19] Hall M J W, Berry D M, Zwierz M and Wiseman H M 2011 arXiv:1111.0788v1 [quant-ph]
- [20] Xiang G Y *et al* 2011 *Nature Photonics* **5** 43
- [21] Rivas A and Luis A 2011 arXiv:1105.6310v2 [quant-ph]
- [22] Gill R D and Massar S 2000 *Phys. Rev. A* **61** 042312
- [23] Tsang M 2011 arXiv:1111.3568v3 [quant-ph]
- [24] Giovannetti V and Maccone L 2012 arXiv:1201.1878v1 [quant-ph]
- [25] Krishna M and Parthasarathy K R 2002 *Indian J. Stat. A* **64** 842
- [26] Coles P J, Yu L, Gheorgiu V and Griffiths R B 2011 *Phys. Rev. A* **83** 062338
- [27] Berta M, Christandl M, Colbeck R, Renes J and Renner R 2009 *Nature Phys.* **6** 659
- [28] Cover T M and Thomas J A 1991 *Elements of Information Theory* (New York: Wiley)
- [29] Hickey R J 1984 *J. Appl. Prob.* **21** 924
- [30] Berry D M, Hall M J W, Zwierz M and Wiseman H M in preparation
- [31] Yurke B, McCall S L and Klauder J R 1986 *Phys. Rev. A* **33** 4033
- [32] Hall M J W 2008 *J. Phys. A: Math. Theor.* **41** 255301

- [33] Vourdas A 1986 *Phys. Rev. A* **34** 3466
- [34] Hall M J W and O'Rourke M J 1993 *Quantum Opt.* **5** 161
- [35] Hall M J W 1999 *Phys. Rev. A* **59** 2602
- [36] Wehner S and Winter A 2010 *New J. Phys.* **12** 025009
- [37] Bialynicki-Birula I and Mycielski J 1975 *Commun. Math. Phys.* **44** 129
- [38] Pegg D M, Vaccaro J A and Barnett S M 1990 *J. Mod. Opt.* **37** 1703