

Uncertainty relations for angular momentum

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2015 New J. Phys. 17 093046

(<http://iopscience.iop.org/1367-2630/17/9/093046>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 151.33.205.208

This content was downloaded on 10/06/2016 at 10:57

Please note that [terms and conditions apply](#).



PAPER

Uncertainty relations for angular momentum

OPEN ACCESS

RECEIVED
13 May 2015REVISED
5 August 2015ACCEPTED FOR PUBLICATION
27 August 2015PUBLISHED
25 September 2015

Content from this work
may be used under the
terms of the [Creative
Commons Attribution 3.0
licence](#).

Any further distribution of
this work must maintain
attribution to the
author(s) and the title of
the work, journal citation
and DOI.

Lars Dammeier¹, René Schwonnek and Reinhard F Werner

Institut für Theoretische Physik, Leibniz Universität, Hannover, Germany

¹ Author to whom any correspondence should be addressed.E-mail: lars.dammeier@itp.uni-hannover.de, rene.schwonnek@itp.uni-hannover.de and reinhard.werner@itp.uni-hannover.de**Keywords:** uncertainty relations, angular momentum, quantum mechanicsSupplementary material for this article is available [online](#)**Abstract**

In this work we study various notions of uncertainty for angular momentum in the spin- s representation of $SU(2)$. We characterize the ‘uncertainty regions’ given by all vectors, whose components are specified by the variances of the three angular momentum components. A basic feature of this set is a lower bound for the sum of the three variances. We give a method for obtaining optimal lower bounds for uncertainty regions for general operator triples, and evaluate these for small s . Further lower bounds are derived by generalizing the technique by which Robertson obtained his state-dependent lower bound. These are optimal for large s , since they are saturated by states taken from the Holstein–Primakoff approximation. We show that, for all s , all variances are consistent with the so-called vector model, i.e., they can also be realized by a classical probability measure on a sphere of radius $\sqrt{s(s+1)}$. Entropic uncertainty relations can be discussed similarly, but are minimized by different states than those minimizing the variances for small s . For large s the Maassen–Uffink bound becomes sharp and we explicitly describe the extremalizing states. Measurement uncertainty, as recently discussed by Busch, Lahti and Werner for position and momentum, is introduced and a generalized observable (POVM) which minimizes the worst case measurement uncertainty of all angular momentum components is explicitly determined, along with the minimal uncertainty. The output vectors for the optimal measurement all have the same length $r(s)$, where $r(s)/s \rightarrow 1$ as $s \rightarrow \infty$.

1. Introduction

The textbook literature on quantum mechanics seems to agree that the uncertainty relations for angular momentum, and indeed for any pair of quantum observables A , B should be given by Robertson’s [25] inequality

$$\Delta_{\rho}^2(A) \Delta_{\rho}^2(B) \geq \frac{1}{4} (\text{tr } \rho [A, B])^2, \quad (1)$$

valid for any density operator ρ , with $\Delta_{\rho}^2(A)$ denoting the variance of the outcomes of a measurement of A on the state ρ . Perhaps the main reason for the ubiquity of this relation in textbooks is that it is such a convenient intermediate step to the proof of uncertainty relations for position and momentum. In that case the right-hand side is $\hbar^2/4$, independently of the state ρ . For any pair A , B other than a canonical pair, however, the relation (1) makes a much weaker statement, requiring some prior information about the state. This begs the question: When and with what bounds it is true that

[Preparation uncertainty:] One cannot choose a state ρ so that $\Delta_{\rho}^2(A)$ and $\Delta_{\rho}^2(B)$ simultaneously become arbitrarily small.

Robertson's relation supports no such conclusion, but on the other hand such a statement does hold in many situations. In fact, in a finite dimensional context it is true whenever A and B do not have a common eigenvector. In this paper we will provide optimal bounds for angular momentum components, establishing the methods for deriving optimal bounds in the general case along the way.

The second reason that (1) is unsatisfactory is that it addresses only the preparation side of uncertainty, in the sense loosely described in the italicized sentence above. However, there is always also a measurement aspect to uncertainty, for which Heisenberg's γ -ray microscope [8] is a paradigm. The error disturbance tradeoff would be stated as

[Error-disturbance uncertainty:] An approximate measurement of A of accuracy ΔA disturbs the system in such a way that from the post-measurement state and the measurement result for A the distribution for observable B can only be inferred with accuracy ΔB , where ΔA and ΔB cannot be simultaneously arbitrarily small.

It is often easier to think of the whole experiment as a *joint measurement* of A and B , and state relations of the kind:

[Measurement uncertainty:] For any measurement device with both an A -type and a B -type output, the marginals will have worst case error ΔA , ΔB with respect to ideal measurements of A and B , satisfying a tradeoff relation.

Again, generic observables and angular momenta satisfy non-trivial relations of this kind. Errors $\Delta A = \Delta B = 0$ can occur only if A and B commute, i.e., under an even more stringent condition than for preparation uncertainty. In this paper we will provide some sharp measurement uncertainty relations for angular momentum, establishing along the way some methods which may be of interest in more general cases.

There is a third reason that one should not be satisfied with (1) with $A = L_1$, $B = L_2$: it involves only two of the three components of angular momentum. But there is no reason tradeoff-relations as described above should not be stated for more than two observables. For angular momentum this seems especially natural. Moreover, it seems natural to state relations for all components simultaneously, i.e., not only for the three components along the axes of an arbitrarily chosen Cartesian reference frame, but for the angular momenta along arbitrary rotation axes, restoring the rotational symmetry of the problem.

Indeed the idea that uncertainty should involve just pairs of observables can be traced to Bohr's habit of expressing complementarity as a relation between 'opposite' aspects, like '*in vitro*' and '*in vivo*' biology. This dualistic preference had more to do with his philosophy than with the actual structure of quantum mechanics. Other founding fathers of quantum mechanics did not share this preference. As Wigner said in an interview [34] in 1963:

I always felt also that this duality is not solved and in this I may have been under Johnnys (John von Neumann) influence, who said, 'Well, there are many things which do not commute and you can easily find three operators which do not commute.' I also was very much under the influence of the spin where you have three variables which are entirely, so to speak, symmetric within themselves and clearly show that it isn't two that are complementary; and I still don't feel that this duality is a terribly significant and striking property of the concepts.

In this spirit, an uncertainty relation for triples of canonical operators was recently proposed and proved [16], and further generalizations are clearly possible. However, we will stick to angular momentum in this paper, and particularly seek to establish relations which do not break rotation invariance.

Our paper responds to an increasing interest in quantitative uncertainty relations. This interest is connected to an increasing number of experiments reaching the uncertainty dominated regime, so that rather than qualitative or order-of-magnitude statements one is more interested in the precise location of the quantum limits. Measurement uncertainty was made rigorous in [2, 5, 26]. There is also a controversial [3, 4] state-dependent version [22]. That adequate uncertainty relations are sometimes better stated in terms of the sum of variances rather than their product has been noted repeatedly [10, 14, 20]. There has been some renewed interest also in the uncertainty between angular momentum and angular position [7], angular momentum of certain states [24] and other non-standard complementary pairs [18].

1.1. Setting and notation

In physics angular momentum appears as orbital or as spin angular momentum. Our theory applies to both, but it must be noted that the bounds obtained do depend on the quantum number for \mathbf{L}^2 . For example, there are states with vanishing orbital angular momentum uncertainties (precisely the rotation invariant ones, i.e., $s = 0$) but none for a $s = \frac{1}{2}$ degree of freedom. Therefore, one first has to decompose the given space into irreducible angular momentum components (integer or half integer), and then use the results for the appropriate s . Hence we will consider throughout a system of spin s , with $s = 1/2, 1, 3/2, \dots$ in its $(2s + 1)$ -dimensional Hilbert space $\mathcal{H} = \mathbb{C}^{2s+1}$. The three angular momentum components will be denoted by L_k , $k = 1, 2, 3$, and the component along a unit vector $\mathbf{e} \in \mathbb{R}^3$ by $\mathbf{e} \cdot \mathbf{L}$. We denote by $|m\rangle$ the eigenvectors of L_3 , so that $L_3 |m\rangle = m |m\rangle$ where $-s \leq m \leq s$ and, with $L_{\pm} = L_1 \pm iL_2$,

$$L_{\pm} |m\rangle = \sqrt{s(s+1) - m(m \pm 1)} |m \pm 1\rangle. \quad (2)$$

Rotation matrices, whether they are considered as elements of $\text{SO}(3)$ or of $\text{SU}(2)$, will typically be denoted by R , the corresponding matrix in the spin s representation by U_R , and normalized Haar measure on $\text{SO}(3)$ or $\text{SU}(2)$ by dR . We will always set $\hbar = 1$. Observables are in general always allowed to be normalized positive operator valued measures, with a typical letter F . For a self-adjoint operator A the spectral measure is an observable in this sense, denoted by E_A . For a component of angular momentum, i.e., $A = \mathbf{e} \cdot \mathbf{L}$ we write $E_{\mathbf{e}}$ for short. For the unit vectors \mathbf{e}_k along the axes we further abbreviate this to E_k .

For the variance of the probability distribution obtained by measuring F on a state (density operator) ρ we write

$$\Delta_{\rho}^2(F) = \min_{\xi} \int (x - \xi)^2 \text{tr} \rho F(dx). \quad (3)$$

This minimum is taken over a quadratic expression in ξ , and it is attained when $\xi = \int x \text{tr} \rho F(dx)$ is the mean value of the distribution. The most familiar case is that of the spectral measure for an operator A , in which case we abbreviate the variance by $\Delta_{\rho}^2(A)$. Then the second moment $\int x^2 \text{tr} \rho F(dx) = A^2$ can also be expressed by A and we get

$$\Delta_{\rho}^2(A) = \Delta_{\rho}^2(E_A) = \text{tr}(\rho A^2) - \text{tr}(\rho A)^2. \quad (4)$$

We say that a unit vector $|\phi\rangle \in \mathcal{H}$ is a *maximal weight vector*, if for some direction $\mathbf{e} \in \mathbb{R}^3$ it satisfies $\mathbf{e} \cdot \mathbf{L} |\phi\rangle = s |\phi\rangle$. This is the same as saying that, for some rotation $R \in \text{SU}(2)$, $|\phi\rangle = U_R |s\rangle$ up to a phase. For such a vector we call $\rho = |\phi\rangle\langle\phi|$ a *spin coherent state*. These states are candidates for states of minimal uncertainty.

1.2. Summary of main results

We now describe the structure of our paper and the main results.

Section 2: Preparation uncertainty. The basic object of study is the variance $\Delta_{\rho}^2(\mathbf{e} \cdot \mathbf{L})$ of the angular momentum in direction \mathbf{e} as a function of the unit vector \mathbf{e} , especially properties which hold for an arbitrary state ρ . After clarifying some general features and explicitly solving the two cases $s \leq 1$ (section 2.2), we look at the traditional setting of just two components L_1, L_2 . The set of uncertainty pairs $(\Delta_{\rho}^2(L_1), \Delta_{\rho}^2(L_2))$ is studied, and the fact that not both variances can be small is found to be well expressed by a lower bound not on the product but on the sum of the variances. We compute numerically (and exactly up to $s = 3/2$) the best constants in

$$\Delta_{\rho}^2(L_1) + \Delta_{\rho}^2(L_3) \geq c_2(s) \quad (5)$$

and find that they asymptotically behave like $c_2(s) \sim s^{2/3}$ (section 2.4). For three components the uncertainty region is also studied in some detail. A prominent feature is again given by a linear bound [10]

$$\Delta_{\rho}^2(L_1) + \Delta_{\rho}^2(L_2) + \Delta_{\rho}^2(L_3) \geq s, \quad (6)$$

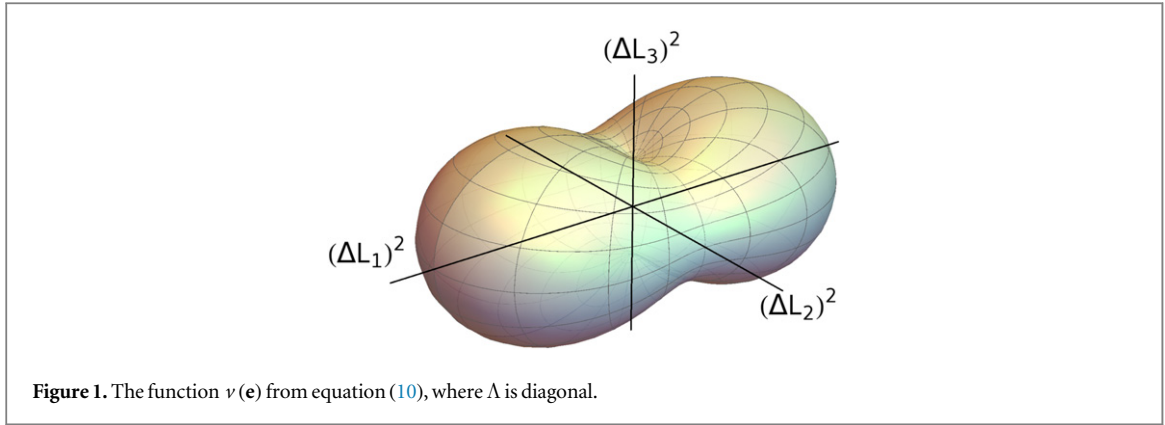
which is very easy to prove (see section 2.1, (17)).

Turning to features of the whole function $\mathbf{e} \mapsto \Delta_{\rho}^2(\mathbf{e} \cdot \mathbf{L})$, we show in section 2.5 that for any ρ there is at least one direction \mathbf{e} such that $\Delta_{\rho}^2(\mathbf{e} \cdot \mathbf{L}) \geq s/2$, i.e.

$$\max_{\mathbf{e}} \Delta_{\rho}^2(\mathbf{e} \cdot \mathbf{L}) \geq \frac{s}{2}. \quad (7)$$

This bound is optimal, since it is saturated by spin coherent states. We generalize from the maximum (seen as the L^{∞} -norm) to all L^p -norms (proposition 1).

For large s equation (6) suggests the scaling $\sim 1/s$ (section 2.7). Indeed, the triples $(\Delta_{\rho}^2(L_1)/s, \Delta_{\rho}^2(L_2)/s, \Delta_{\rho}^2(L_3)/s)$ converge as $s \rightarrow \infty$ (theorem 4). The lower bound on the limit set is obtained by a generalization of Robertson's method for proving (1) section 2.6, which for finite s is (39)



$$\Delta_\rho^2(L_1)(\Delta_\rho^2(L_2) + \Delta_\rho^2(L_3)) \geq \frac{1}{4}(s(s+1) - \Delta_\rho^2(L_1) - \Delta_\rho^2(L_2) - \Delta_\rho^2(L_3)), \quad (8)$$

where the components are ordered so that $\Delta_\rho^2(L_1) \geq \Delta_\rho^2(L_2) \geq \Delta_\rho^2(L_3)$. The upper bound in theorem 4 is provided by a family of states suggested by the Holstein–Primakoff approximation.

Section 3.1: Vector model and moment problems. We revisit the so-called vector model of angular momentum, a classical model which is still found in some textbooks. We show that it can correctly portray the moments up to second order (i.e., means and variances) of the angular momentum observables, but fails on higher moments and, of course, on correlations.

Section 3.2: Entropic uncertainty relations. We discuss entropic uncertainty relations only very briefly. We point out that the criteria ‘variance’ and ‘entropy’ may disagree on which of two distributions is ‘more sharply concentrated’. This effect is illustrated by the uncertainty diagrams for $s = 1$. We show also that the general Maassen–Uffink bound [19] while suboptimal for $s = 1$, becomes sharp for $s \rightarrow \infty$, and determine a family of states saturating it.

Section 4: Measurement uncertainty. We consider two measures for the deviation of an approximate observable from an ideal reference, called metric error and calibration error. We then discuss uncertainty relations for the joint measurement of *all* angular momentum components. The output of such an observable is an angular momentum three-vector $\boldsymbol{\eta}$, from which one can obtain a measurement of the \mathbf{e} -component (for any unit vector \mathbf{e}) simply by taking $\mathbf{e} \cdot \boldsymbol{\eta}$ as the output. Such a marginal observable can in turn be compared with the quantum observable $\mathbf{e} \cdot \mathbf{L}$. The uncertainty relation in this case gives a lower bound on the error in the worst case with respect to \mathbf{e} . Our main result (theorem 12) is a determination of the optimal bound, and an observable saturating it. It turns out that the optimal observable is covariant with respect to rotations, and this implies that it simultaneously minimizes the maximal metric error and the maximal calibration error. All the output vectors have the same length $r_{\min}(s)$, which depends in a non-trivial way on s but is close to s for large s .

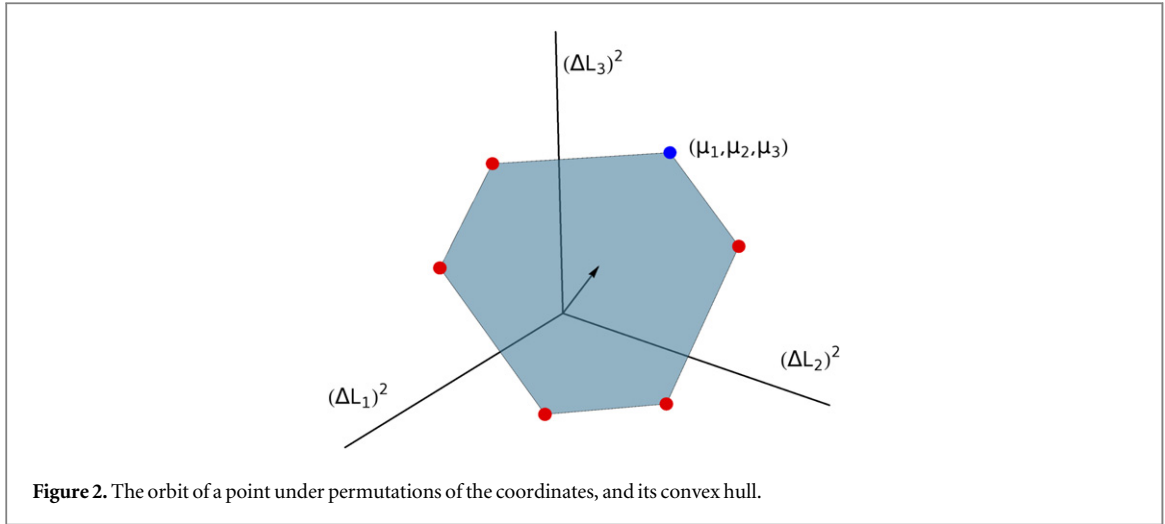
2. Preparation uncertainty

In this section we consider the preparation uncertainty, i.e., a property of a given state ρ . For every unit vector $\mathbf{e} \in \mathbb{R}^3$ we can form the variance of the angular momentum component $\mathbf{e} \cdot \mathbf{L}$, and hence study the function

$$v(\mathbf{e}) = \Delta_\rho^2(\mathbf{e} \cdot \mathbf{L}) = \text{tr}(\rho(\mathbf{e} \cdot \mathbf{L})^2) - (\text{tr} \rho \mathbf{e} \cdot \mathbf{L})^2 \quad (9)$$

on the unit sphere. For the purposes of this section, this function summarizes all the uncertainty properties of the state ρ , and all results in this section are statements about properties of this function, which are valid for all ρ . To visualize the function v , we can use a three-dimensional radial plot, i.e., the surface containing all vectors $v(\mathbf{e})\mathbf{e}$, as \mathbf{e} runs over all unit vectors. A typical radial plot is shown in figure 1. Often we are also interested in the components with respect to some Cartesian reference frame. In this case the best visualization is an *uncertainty diagram*, which represents the possible pairs/triples etc of variances in the same state. In our case this will be the set of pairs $(v(\mathbf{e}_1), v(\mathbf{e}_2))$, or triples $(v(\mathbf{e}_1), v(\mathbf{e}_2), v(\mathbf{e}_3))$. The diagrams for $s = 1$ are shown in figure 4. In this diagram it can be seen that the uncertainty region is not convex in general. Since we are only interested in lower bounds, we therefore always take the *monotone closure* of the uncertainty region, i.e., we also include with every point the whole quadrant/octant of points in which one or more of the coordinates increase. This is described in more detail in section 2.3.

It turns out that after a rotation to suitable principal axes (which has already been carried out in figure 1), the function v depends only on three real parameters μ_1, μ_2, μ_3 . To see this we introduce the 3×3 -matrix $\Lambda = \Lambda(\rho)$ by



$$v(\mathbf{e}) = \sum_{jk} e_j e_k \Lambda_{jk}(\rho) \quad \text{with} \quad (10)$$

$$\Lambda_{jk}(\rho) = \Re e \operatorname{tr}(\rho L_j L_k) - \lambda_j \lambda_k, \quad (11)$$

$$\lambda_j(\rho) = \operatorname{tr}(\rho L_j). \quad (12)$$

Since the L_k transform as a vector operator (i.e., with respect to the spin-1 representation of $SU(2)$) we see that by choice of an appropriate coordinate basis in \mathbb{R}^3 we can diagonalize Λ , i.e., we can choose $\Lambda_{jk}(\rho) = \mu_j \delta_{jk}$ with $\mu_1 \geq \mu_2 \geq \mu_3 \geq 0$.

The eigenvalues (μ_1, μ_2, μ_3) of any matrix $\Lambda(\rho)$ are also a possible triple of variances, namely for a suitably rotated state. In fact, we can find the uncertainty triples for *all* rotated versions of ρ quite easily: When R is the rotation matrix taking the eigenbasis of Λ to the basis $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3$ under consideration, then

$$v(\mathbf{e}_j) = \sum_k R_{jk}^2 \mu_k. \quad (13)$$

Now the squared rotation matrix is doubly stochastic, so by Birkhoff's theorem it is a convex combination of permutation matrices. We therefore find the variance triple in basis \mathbf{e} in the convex hull of the six points, arising from the triple of μ_k by permutation. These six points lie in a plane orthogonal to the vector $(1, 1, 1)$, so they form a hexagon (see figure 2), which degenerates into a triangle if two of the μ_k are equal. One can easily check that the full hexagon is attained by squared rotation matrices.

2.1. Basic bounds

For an L_3 -eigenstate $\rho = |m\rangle\langle m|$ we find $\lambda = (0, 0, m)$ and

$$\Lambda(|m\rangle\langle m|) = \frac{1}{2}(s(s+1) - m^2) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \geq \frac{s}{2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (14)$$

which re-enforces that, for eigenstates, the variances in all directions are smallest for the maximal weight $m = s$. Maximal variance is attained for an equal weight mixture $\rho = 1/2(|s\rangle\langle s| + |-s\rangle\langle -s|)$, which has L_3 -variance s^2 . Hence, for all \mathbf{e} :

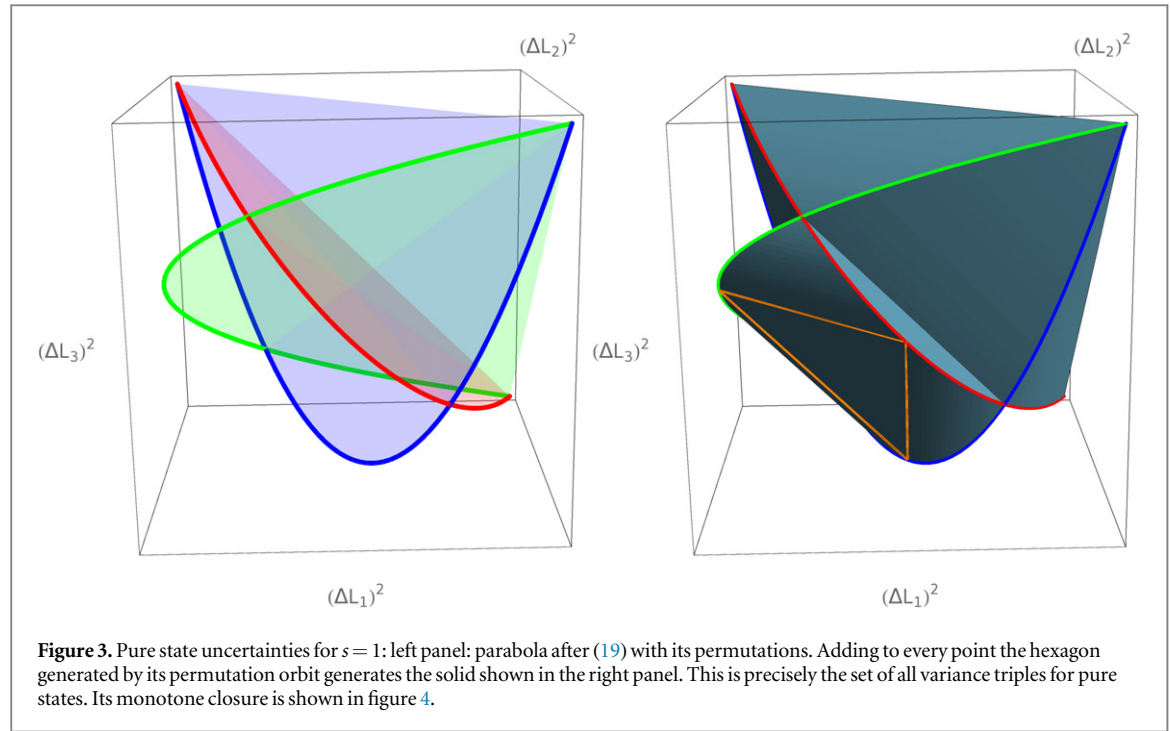
$$0 \leq v(\mathbf{e}) \leq s^2. \quad (15)$$

The average $\overline{v(\mathbf{e})}$ over \mathbf{e} with respect to the surface measure on the sphere, or equivalently the average of $v(R\mathbf{e})$ over Haar-random rotations R , is readily computed from (10), since the average of $e_j e_k$ over the unit sphere is just $\delta_{jk}/3$. Therefore, from (10),

$$\overline{v(\mathbf{e})} = \frac{1}{3} \operatorname{tr} \Lambda(\rho) = \frac{1}{3}(s(s+1) - |\lambda|^2) \geq \frac{1}{3}(s(s+1) - s^2) \geq \frac{s}{3}. \quad (16)$$

In the same simple way we can get an inequality for the variances along the three coordinate directions of a Cartesian coordinate system:

$$\sum_{k=1}^3 v(\mathbf{e}_k) = \operatorname{tr} \Lambda(\rho) \geq s. \quad (17)$$



In both cases equality holds precisely for $|\lambda| = s$, i.e., if ρ is an eigenstate of one of the operators $\mathbf{e} \cdot \mathbf{L}$ for the maximum eigenvalue $m = s$.

2.2. Special features for $s = 1/2$ and $s = 1$

For $s = 1/2$, it happens that L_j and L_k (i.e., up to a factor the Pauli matrices) anticommute for different j, k , so that

$$s = \frac{1}{2} : \quad \Lambda_{jk}(\rho) = \frac{1}{4} \delta_{jk} - \lambda_j \lambda_k. \quad (18)$$

The eigenvalues are $(\mu_1, \mu_2, \mu_3) = 1/4(1, 1, 1 - 4|\lambda|^2)$. Of course, pure states are characterized by $|\lambda| = \frac{1}{2}$. The uncertainties are $1/4 - \lambda_j^2$, and so the uncertainty region is described by a triangle.

The case $s = 1$ is still special because the 3 + 6 operators L_k and $(L_j L_k + L_k L_j)/2$ form a basis of the operators on \mathbb{C}^3 . Therefore, ρ can be reconstructed from (λ, Λ) and, in particular, the set of pure ρ can be characterized in terms of conditions on the eigenvalues μ_k and λ_k . In order to analyze these conditions, let us take the representation of the group $SU(2)$ by real orthogonal matrices. Now consider a vector $\psi \in \mathbb{C}^3$, which we can split into $\psi = \Re\psi + i \Im\psi$ with $\Re\psi, \Im\psi \in \mathbb{R}^3$. Note that the real continuous function $t \mapsto \langle \Re(e^{it}\psi), \Im(e^{it}\psi) \rangle$ changes sign between $t = 0$ and $t = \pi/2$, so we can choose a complex phase for ψ to make $\Re\psi$ and $\Im\psi$ orthogonal. Moreover, we can apply a rotation, so that $\Re\psi$ and $\Im\psi$ are along the first two coordinate axes. Hence, up to a rotation, we have

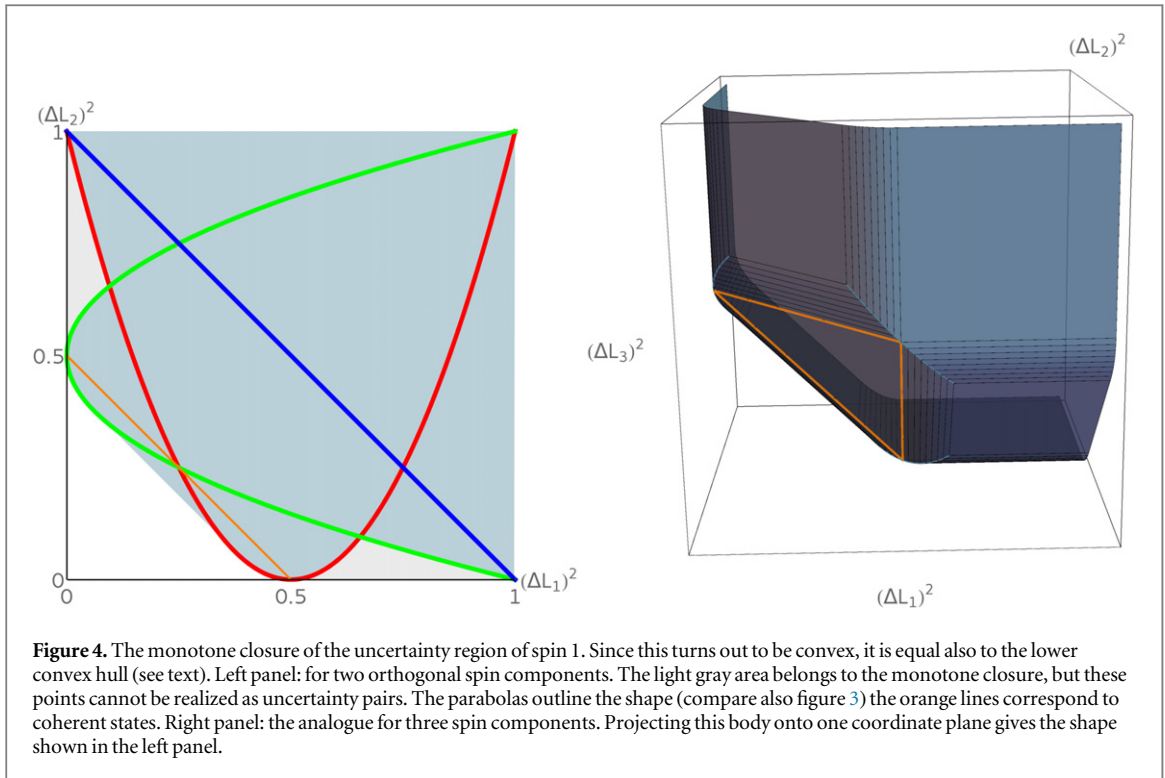
$$\psi = \begin{pmatrix} \cos t \\ i \sin t \\ 0 \end{pmatrix} \quad \lambda = \begin{pmatrix} 0 \\ 0 \\ 2 \sin t \cos t \end{pmatrix} \quad \Lambda = \begin{pmatrix} \tau & 0 & 0 \\ 0 & 1 - \tau & 0 \\ 0 & 0 & 1 - 4\tau(1 - \tau) \end{pmatrix}, \quad (19)$$

where $t \in [-\pi, \pi]$ or $\tau = (\sin t)^2 \in [0, 1]$. In the three-component diagram the curve parameterized by τ is a parabola lying in a diagonal plane. This parabola, and the two copies arising by coordinate permutation are shown in figure 3, as well as the body of uncertainty triples of all pure states, which arises by adding to each point on the parabola the hexagon formed by its permutation orbit. A paper cutout model of this solid is provided as a supplement².

2.3. General minimization method

Consider now, a little more generally, any collection of Hermitian operators A_1, \dots, A_n . We can then form, for any state ρ , the variance n -tuple $(\Delta_\rho^2(A_1), \dots, \Delta_\rho^2(A_n))$, and ask which region Ω in \mathbb{R}^n is filled, when ρ runs over the whole state space. We call Ω the *uncertainty region* of the operator tuple. Typically, this is not a convex set, because $\Delta_\rho^2(A)$ contains a term quadratic in ρ , which consequently does not respect convex mixtures. Ω will be

² Supplementary material—cut-out paper model, available at stacks.iop.org/njp/17/093046/mmedia.



simply connected (as a continuous image of the state space), but beyond that there are few general facts. It can happen that starting from a point in the uncertainty region we can leave the region by increasing one of the coordinates, i.e., the region encodes upper bounds on variances as well as lower bounds. This is clearly not relevant to the theme of uncertainty relations, where we ask for universal lower bounds only. We can therefore consider the *monotone closure* of the uncertainty region, by including all points with larger uncertainties, i.e.

$$\Omega^+ = \left\{ (x_1, \dots, x_n) \mid \exists \forall x_\alpha \geq \Delta_\rho^2(A_\alpha) \right\}. \quad (20)$$

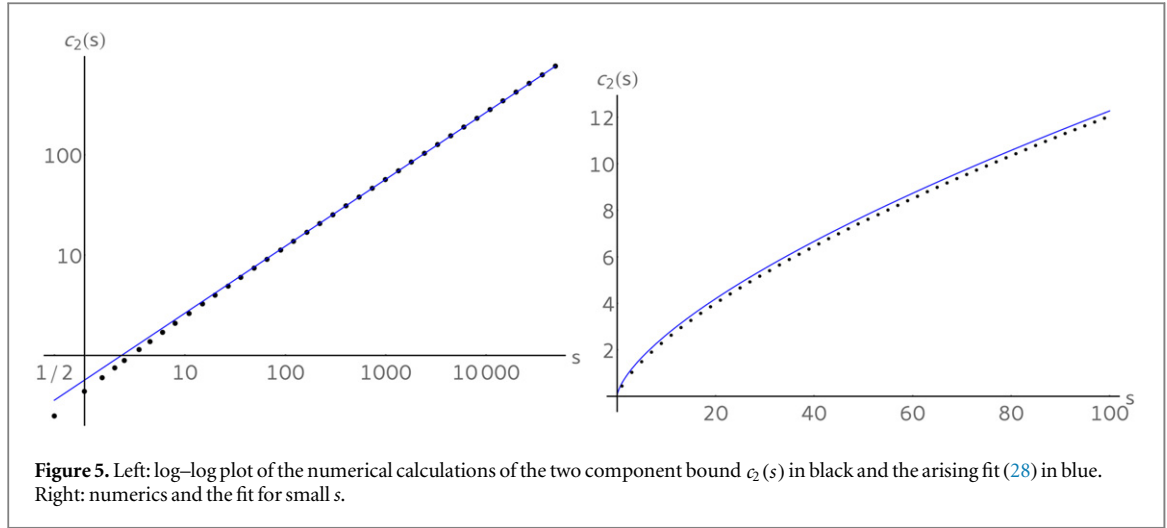
This is still not necessarily a convex set. We will denote the convex hull of Ω^+ by Ω^\vee and call it the *lower convex hull* of Ω (see figure 4). It is this set which has an efficient characterization. Indeed, as a closed convex set it is the intersection of all half spaces containing it, and the monotonicity condition restricts these half spaces to those whose normal vector w has all components non-negative. In other words

$$\Omega^\vee = \left\{ (x_1, \dots, x_n) \mid \forall w_\alpha \geq 0 : \sum_\alpha w_\alpha x_\alpha \geq m(w_1, \dots, w_n) \right\}, \quad (21)$$

$$\begin{aligned} m(w_1, \dots, w_n) &= \inf_\rho \sum_\alpha w_\alpha \Delta_\rho^2(A_\alpha) \\ &= \inf_\rho \inf_{a_1, \dots, a_n} \sum_\alpha w_\alpha \operatorname{tr} \rho (A_\alpha - a_\alpha \mathbb{1})^2. \end{aligned} \quad (22)$$

In this double infimum we can exchange the order, leading to two kinds of operations: With fixed ρ the global minimum over the a_α is obviously at $a_\alpha = \operatorname{tr} \rho A_\alpha$. On the other hand, with fixed a_α the global minimum over ρ is computed by finding the ground state of the positive semidefinite operator $H(a) = \sum_\alpha w_\alpha (A_\alpha - a_\alpha \mathbb{1})^2$. An efficient algorithm is therefore obtained by alternating between these two steps. The upper estimates on m obtained in this way are non-increasing and in practice converge quite well, and independently of the starting value. However, we do not have a theorem to this effect. An analytic consequence of this algorithm (independent of convergence) is that we can restrict the infimum to pure states, since this is sufficient to get the ground state energies.

The algorithm is then run for a suitable set of tuples (w_1, \dots, w_n) , so that for each run, one obtains a tangent plane to Ω^\vee but also the state ρ and with it, the tuple of variances in Ω . We illustrate the results in figure 4 for the case of spin 1, and the operator tuples (L_1, L_2) and (L_1, L_2, L_3) , respectively. For low spin these diagrams can be determined analytically (see the next subsection). The most prominent feature of two-component diagram is the symmetric linear bound, which depends on s and is determined in section 2.4.



2.4. The linear two-component bound

For every s , let $c_2(s)$ be the best constant in the inequality

$$\Delta_\rho^2(L_1) + \Delta_\rho^2(L_3) \geq c_2(s). \quad (23)$$

For $s = 1/2$, 1 it is readily computed from the eigenvalues of Λ given in section 2.2. For arbitrary s we can use a slightly simplified version of the variational principle (21). We have $w_1 = w_3 = 1$, and can assume that $a_1 = 0$ by rotation invariance around the two-axis. Thus

$$c_2(s) = \inf_{\phi} \inf_a \left\langle \phi \left| s(s+1)\mathbb{1} - L_2^2 - 2aL_3 + a^2\mathbb{1} \right| \phi \right\rangle, \quad (24)$$

where the first infimum runs over all pure states (for fixed a a ground state problem) and a over the reals (for fixed ϕ the expectation value of L_3). One notes that in this operator only matrix elements with even $m - m'$ are non-zero, so the problem can be further reduced. For up to $s = 3/2$ it effectively leads to two-dimensional ground state problems. In this way (resp. by using the results of section 2.2) we get

$$c_2\left(\frac{1}{2}\right) = \frac{1}{4}, \quad (25)$$

$$c_2(1) = \frac{7}{16}, \quad (26)$$

$$c_2(3/2) = \frac{9}{4} + \gamma^2 - \sqrt{4\gamma^2 + 2\gamma + 1} \approx 0.600\,933$$

where $\gamma = \cos(\pi/9)$. (27)

Note that the bound $c_2(1)$ was already obtained in [10]. It is readily seen numerically that $c_2(s)$ increases with s , but sub-linearly. This means that if we scale the diagram of Ω^\vee (see figure 4, right) by a factor $1/s$ so that the bottom triangle described by (17) stays fixed, the two-component inequality excludes an asymptotically small prism around the axes. Figure 5 shows the asymptotic behavior of c_2 in a log-log plot, which suggests that

$$c_2(s) \approx 0.569\,524\, s^{2/3} \text{ for large } s. \quad (28)$$

2.5. Power mean and maximal uncertainty

A natural way to characterize states with small variance is to look for the maximum of the variance function $v(\mathbf{e})$ defined in (10). An uncertainty relation would then put a lower bound $c(s)$ on this maximum. In other words we would like to prove the following statement: *for every state ρ there is some direction e such that $\Delta_\rho^2(E_e)$ is larger than c .* By considering coherent states we can immediately see that $c(s) \leq s/2$. The following proposition shows that coherent states in fact have minimal variance in terms of this criterion, and we even have equality.

Such a result can be seen as one end of a one-parameter family of criteria, of which (16) is the other end: we can judge the ‘size’ of the function v by its \mathcal{L}^p norm, of which the maximum is the special case $p = \infty$, and the mean the case $p = 1$. We therefore formulate a proposition to cover all these cases.

Proposition 1. *For every $s \in \mathbb{N}/2$ and every $p \in [1, \infty]$ there is a constant $c(p, s)$ such that, for every density operator ρ in the spin s representation*

$$\|v\|_p = \left(\int \frac{d\mathbf{e}}{4\pi} \Delta_\rho^p(\mathbf{e} \cdot \mathbf{L}) \right)^{\frac{1}{p}} \geq c(p, s) \quad (29)$$

with equality whenever ρ is a spin coherent state. For $p < \infty$ these are the only states with equality. For $p = \infty$ equality holds also for mixtures $\rho = p_+ | +s \rangle \langle +s | + p_- | -s \rangle \langle -s |$, and rotations thereof, provided that $p_+ p_- \geq 1/(8s)$.

The constant is

$$c(p, s) = \frac{s}{2} \left(\frac{\sqrt{\pi} \Gamma(p+1)}{2 \Gamma(p + \frac{3}{2})} \right)^{\frac{1}{p}}, \quad (30)$$

with special values $c(1, s) = s/3$, $c(2, s) = s\sqrt{8/15} \approx 0.73s$, $c(\infty, s) = s/2$.

Proof. Let $\lambda_j = \text{tr } \rho L_j$ be the vector of expectation values, and consider the set of density operators ρ^β arising from ρ by rotation R_β around the vector λ by the angle β . For each ρ^β , we call the variance function $v_\beta(\mathbf{e}) = v(R_\beta \mathbf{e})$. By averaging over β we find a state $\bar{\rho}$, with variance function

$$\bar{v}(\mathbf{e}) = \frac{1}{2\pi} \int d\beta v_\beta(\mathbf{e}), \quad (31)$$

where we used, crucially, that all ρ^β and $\bar{\rho}$ have the same expectations λ_j . By the triangle inequality for the p -norm, we have $\|\bar{v}\|_p \leq \|v\|_p$. Hence we can restrict the search for the ρ with minimal $\|v\|_p$ to those which are rotation invariant around some axis, say the three-axis.

Such a state can be jointly diagonalized with L_3 , and is hence of the form $\rho = \sum_m p_m |m\rangle \langle m|$. Then $\lambda = (0, 0, \sum_m p_m m)$, and $\Lambda(\rho)$ is diagonal with

$$\Lambda_{11} = \Lambda_{22} = \frac{1}{2} \left(s(s+1) - \sum_m p_m m^2 \right) \geq \frac{1}{2} (s^2 + s - s^2) = \frac{s}{2}, \quad (32)$$

$$\Lambda_{33} = \sum_m p_m m^2 - \left(\sum_m p_m m \right)^2 \geq 0, \quad (33)$$

$$v(\mathbf{e}) = (e_1^2 + e_2^2) \Lambda_{11} + e_3^2 \Lambda_{33}. \quad (34)$$

The last equation shows that the function v becomes pointwise smaller (and hence smaller in p -norm) if we decrease some Λ_{ii} . That is, we have to go to the minimum on both Λ_{11} and Λ_{33} . The minimum in (32) is attained precisely when $p_m \neq 0$ only for $m = \pm s$. Then minimality in (32) forces ρ to be a spin coherent state. For $p = \infty$ the norm only sees the maximum, so the pointwise minimum need not be chosen, and we may allow $0 \leq \Lambda_{33} \leq \Lambda_{11}$ without changing the maximum. The latter inequality translates to the one given in proposition 1.

The concrete constants follow easily by integrating the p th power of (34) with $\Lambda_{11} = s/2$ and $\Lambda_{33} = 0$ with respect to the normalized surface measure on the sphere, i.e.

$$c(p, s) := \frac{s}{2} \left(\int_0^\pi \frac{\sin \theta d\theta}{2} (\sin \theta)^{2p} \right)^{\frac{1}{p}}. \quad (35)$$

□

2.6. Robertson's technique: a generalization

We have criticized the Robertson inequality (1) for not giving a state independent bound. However, with only little effort it can be used to derive such a bound. Indeed, abbreviating $v_j = \Delta_\rho^2(L_j)$, and $\lambda_j = \text{tr } \rho L_j$ we can add the three inequalities of the form $v_1 v_2 \geq \lambda_3^2/4$ and use that $\sum_j (v_j + \lambda_j^2) = s(s+1)$ to obtain

$$v_1 v_2 + v_2 v_3 + v_3 v_1 \geq \frac{1}{4} \left(s(s+1) - (v_1 + v_2 + v_3) \right). \quad (36)$$

Clearly, this no longer allows $v_1 = v_2 = 0$, since $v_3 \leq s^2$. The set of variance triples satisfying this is shown in figure 6. Comparison with figure 4 readily shows that this bound is not optimal. However, we can generalize Robertson's technique from two to three components rather than extend his two component result in this trivial way. The basis of the technique is to utilize the observation that for any finite collection of operators X_j (not necessarily Hermitian or normal) the matrix $m_{jk} = \text{tr } \rho X_j^* X_k$ is positive definite, which is the same as saying that for any complex linear combination $X = \sum_j a_j X_j$ the expectation of $X^* X$ must be positive. In order to get Robertson's inequality for L_1, L_2 this idea is applied to the three operators $\mathbb{1}, L_1, L_2$. In fact, this leads to Schrödinger's improvement of the inequality [27] which also contains the square of the covariance matrix element $\Lambda_{12}(\rho)^2$ on the right-hand side.

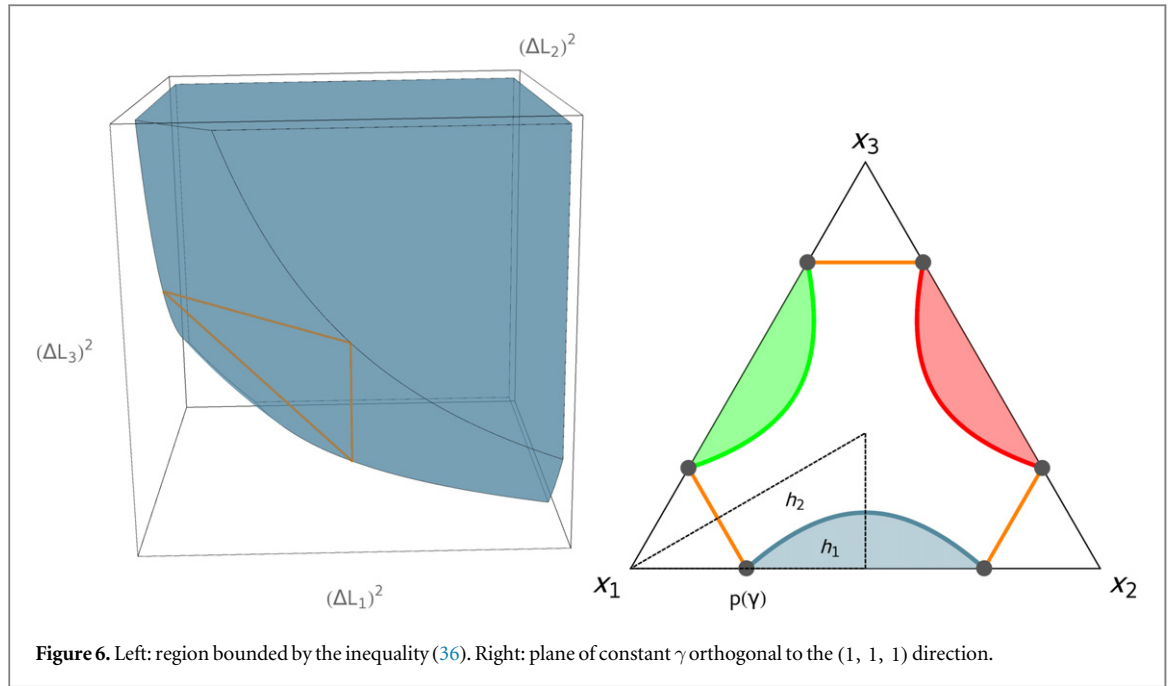


Figure 6. Left: region bounded by the inequality (36). Right: plane of constant γ orthogonal to the $(1, 1, 1)$ direction.

We will apply the method to the four operators $\mathbb{1}$, L_1 , L_2 , L_3 . In order to simplify the expressions, however, we will not look for variances and the off-diagonal elements of $\Lambda(\rho)$, but for inequalities involving the eigenvalues μ_j . As discussed at the beginning of this section, this will contain all the information needed. In other words, we will take the matrix $\Lambda(\rho)$ as the diagonal matrix with entries $\mu_1 \geq \mu_2 \geq \mu_3$. The matrix which then needs to be positive is

$$M = \begin{pmatrix} 1 & \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_1 & \mu_1 + \lambda_1^2 & \lambda_1 \lambda_2 + i\lambda_3/2 & \lambda_1 \lambda_3 - i\lambda_2/2 \\ \lambda_2 & \lambda_1 \lambda_2 - i\lambda_3/2 & \mu_2 + \lambda_2^2 & \lambda_2 \lambda_3 + i\lambda_1/2 \\ \lambda_3 & \lambda_1 \lambda_3 + i\lambda_2/2 & \lambda_2 \lambda_3 - i\lambda_1/2 & \mu_3 + \lambda_3^2 \end{pmatrix}. \quad (37)$$

The positivity of this matrix is equivalent (see e.g. [13 theorem 7.2.5]) to the positivity of the principal minors, i.e., the determinants of the submatrices of the first k rows and columns for $k = 1, 2, 3, 4$. The first three of these are 1 , μ_1 , and $\mu_1 \mu_2 - \lambda_3^2/4$. The positivity of the third one is Robertson's inequality (1). The only remaining condition for the positivity of M is $\det M \geq 0$, which evaluates to

$$\mu_1 \mu_2 \mu_3 - \frac{1}{4}(\lambda_1^2 \mu_1 + \lambda_2^2 \mu_2 + \lambda_3^2 \mu_3) \geq 0. \quad (38)$$

This will be combined with the normalization condition

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = s(s+1) - (\mu_1 + \mu_2 + \mu_3). \quad (39)$$

The condition on the triples (μ_1, μ_2, μ_3) we have to evaluate is the existence of λ_i satisfying both these relations. Since only the squares enter, let us set $x_j = \lambda_j^2$. Then (39) describes a triangle in the positive quadrant with equal intercept $s(s+1) - (\mu_1 + \mu_2 + \mu_3)$ with the axes. The inequality (38) describes a tetrahedron spanned by the origin and the axis intercepts $x_1^0 = 4\mu_2 \mu_3$, and cyclic. Note that Robertson's inequality is automatically satisfied on this tetrahedron. Obviously the tetrahedron and the triangle intersect if and only if one of the axis intercepts of the tetrahedron reaches or lies above the triangle. Since we can take the eigenvalues ordered: $\mu_1 \geq \mu_2 \geq \mu_3$ this means

$$4\mu_1 \mu_2 \geq s(s+1) - (\mu_1 + \mu_2 + \mu_3). \quad (40)$$

This is a bound to the eigenvalue of the Λ -matrix. By Birkhoff's theorem, the variances arising from such Λ also includes all convex combinations of permutations of the μ_i (see beginning of section 2). In order to characterize the set of variance triples generated in this way we need the following Lemma. In its formulation the variables $\sigma \in S_3$ run over the permutation group on three elements, and are applied to the components of a three-vector (see also figure 2).

Lemma 2. *With the notation from above, the following sets K_1 and K_2 are equal*

$$K_1 = \bigcup_{\gamma \geq s} H_1(\gamma), \quad \text{where} \quad (41)$$

$$\begin{aligned} H_1(\gamma) &= \text{conv} \left(\bigcup_{\sigma \in S_3} \sigma(h_1(\gamma)) \right), \quad \text{with} \\ h_1(\gamma) &= \left\{ (\mu_1, \mu_2, \mu_3) \mid \sum_i \mu_i = \gamma, \mu_1 \geq \mu_2 \geq \mu_3 \geq 0, 4\mu_1\mu_2 \geq s(s+1) - \gamma \right\} \\ \text{and} \\ K_2 &= \bigcup_{\gamma \geq s} H_2(\gamma), \quad \text{where} \\ H_2(\gamma) &= \bigcup_{\sigma \in S_3} \sigma(h_2(\gamma)), \quad \text{with} \\ h_2(\gamma) &= \left\{ (v_1, v_2, v_3) \mid \sum_i v_i = \gamma, v_1 \geq v_2 \geq v_3 \geq 0, 4v_1(v_2 + v_3) \geq s(s+1) - \gamma \right\}. \end{aligned} \quad (42)$$

Proof. For the equality of K_1 and K_2 it is sufficient to show that H_1 and H_2 coincide for every γ . The restriction $\sum_i v_i = \sum_i \mu_i = \gamma$, together with the three-fold symmetry of the problem, tells us that $H_1(\gamma)$ and $H_2(\gamma)$ are subsets of the triangle, whose corners lie on the axes at a distance γ from the origin. In this triangle the ordering of the v_i and μ_i reduces h_1 and h_2 to the dashed subset marked in figure 6.

Now the first and last condition in the definition of h_2 can be combined to obtain

$$\begin{aligned} 4v_1(\gamma - v_1) &\geq s(s+1) - \gamma, \\ 0 &\geq v_1^2 - \gamma v_1 - \frac{s(s+1) - \gamma}{4}, \end{aligned} \quad (43)$$

so we get

$$v_1 \leq -\frac{\gamma}{2} \pm \sqrt{\frac{\gamma^2}{4} + \frac{s(s+1) - \gamma}{4}} =: c(\gamma). \quad (44)$$

Because $v_1 \geq 0$ we have to choose the positive sign, which means that $H_2(\gamma)$ is the intersection of the triangle with the three halfspaces $v_i \leq c(\gamma)$, whose boundaries are marked as a orange lines in figure 6, i.e.

$$H_2(\gamma) = \left\{ (v_1, v_2, v_3) \mid \sum_i v_i = \gamma, v_i \leq c(\gamma) \right\}. \quad (45)$$

$H_2(\gamma)$ is clearly a convex polytope. The extremal points of $H_2(\gamma)$ have to saturate at two of the defining inequalities. In the ordered triangle ($v_1 \geq v_2 \geq v_3$) the only extreme point is given by $p(\gamma) := (c(\gamma), \gamma - c(\gamma), 0)$, and all others can be obtained by permutations. Hence H_2 can be described as the hexagon $H_2(\gamma) = \text{conv}(\bigcup_{\sigma \in S_3} \sigma(p(\gamma)))$. On the one hand, by comparing the defining inequalities for h_1 and h_2 , we can see that every triple $\mu_i \in h_1$ is also part of h_2 . So including the permutations and by the fact that $H_2(\gamma)$ is convex, we get $H_1(\gamma) \subseteq H_2(\gamma)$.

On the other hand, the three-component of p is zero, so it is also part of $h_1 \in H_1(\gamma)$. While the point p and its permutations are the extremal points of $H_2(\gamma)$ and $H_1(\gamma)$ is convex, we have $H_2(\gamma) \subseteq H_1(\gamma)$. \square

Therefore we get the following statement:

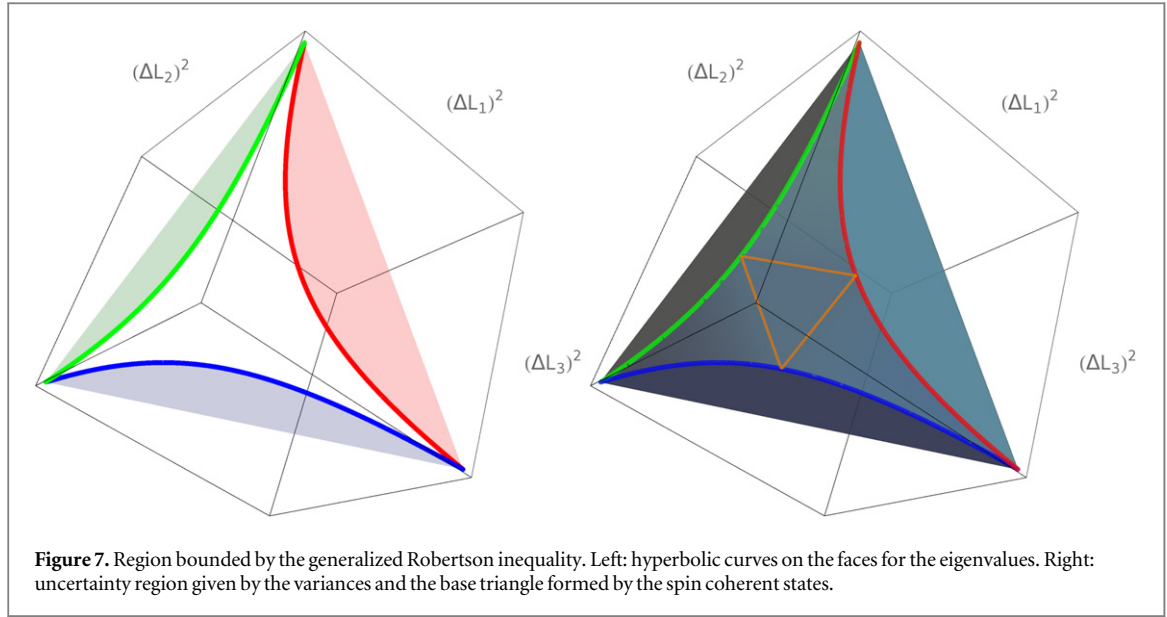
Proposition 3. *Let $v_1 \geq v_2 \geq v_3 \geq 0$ be variances of the angular momentum components, then the following holds:*

$$4v_1(v_2 + v_3) \geq s(s+1) - (v_1 + v_2 + v_3). \quad (46)$$

As one can see in figure 6, the boundaries of the corresponding uncertainty region on the coordinate planes are given by permutations of the hyperbolic curve $4v_1v_2 = s(s+1) - v_1 - v_2$. This uncertainty region is monotonously closed and given by the convex hull of the above hyperbolic curves. This is shown in figure 7.

2.7. Asymptotic case

Now we take a look at the behavior of the asymptotic uncertainty region for $s \rightarrow \infty$. We already know that $\Delta_\rho^2(L_1) + \Delta_\rho^2(L_2) + \Delta_\rho^2(L_3) \geq s$ and hence it is appropriate to scale the problem by $1/s$, which will fix the sum



of the variance in the lower base triangle to 1. We start with the asymptotic behavior of the generalized Robertson inequality derived in the previous section. On the scale of $1/s$, i.e. $\nu_i/s = \nu_i$, and the ordering $\nu_1 \geq \nu_2 \geq \nu_3$ this inequality reads

$$4\nu_1(\nu_2 + \nu_3) \geq 1 + \frac{1}{s} \left(1 - (\nu_1 + \nu_2 + \nu_3) \right) \quad (47)$$

and as s goes to infinity the set of possible variances shrink to

$$4\nu_1(\nu_2 + \nu_3) \geq 1, \quad (48)$$

because $\sum_i \nu_i \geq 1$. Hence the inequality (47) gets stronger for increasing s .

In this section we will show that this bound is attained by states, which will be constructed in the following way. Using the technique described in part 2.3, we look for the states ψ , which minimize the expectation of the operator

$$H(s, \mathbf{w}) = \frac{1}{s} \left(w_1 (L_1 - \lambda_1)^2 + w_2 (L_2 - \lambda_2)^2 + w_3 (L_3 - \lambda_3)^2 \right), \quad (49)$$

for a normal vector \mathbf{w} . We do this using the Holstein–Primakoff transformation [11]:

$$L_+ = \sqrt{2s} \sqrt{1 - \frac{a^*a}{2s}} a \quad L_- = \sqrt{2s} a^* \sqrt{1 - \frac{a^*a}{2s}} \quad L_3 = s - a^*a. \quad (50)$$

Here a and a^* are the creation and annihilation operators, so we have a representation of the angular momentum algebra in the oscillator basis. For large s and appropriate states, this transformation can be reduced to

$$L_+ = \sqrt{2s} a + \mathcal{O}(s^{-\frac{1}{2}}) \quad L_- = \sqrt{2s} a^* + \mathcal{O}(s^{-\frac{1}{2}}) \quad L_3 = s - n. \quad (51)$$

Notice that in the Holstein–Primakoff basis, the spin coherent state $|s\rangle$ is transformed to the ground state $|0\rangle_{\text{HP}}$, hence the state $|n\rangle_{\text{HP}}$ corresponds to $|s - n\rangle$ in the standard angular momentum L_3 eigenbasis. Now we rewrite H using the above transformation and the relation for position $L_1 = \frac{1}{2}(L_+ + iL_-) = \frac{\sqrt{2s}}{2}(a + a^*) = \sqrt{s}X$ and momentum $L_2 = \frac{1}{2}(L_+ - iL_-) = \frac{\sqrt{2s}}{2i}(a - a^*) = \sqrt{s}P$. We arrive at

$$H(s, \mathbf{w})_{\text{HP}} = \left(w_1 (X - \xi)^2 + w_2 (P - \eta)^2 + \frac{w_3}{s} (s - a^*a - \zeta)^2 \right) + \mathcal{O}(s^{-\frac{1}{2}}). \quad (52)$$

Here ξ , η and ζ denote the transformed expectation values. From 2.1 we know that $|s\rangle$ has minimal uncertainty for $w \sim (1, 1, 1)$ and arbitrary s . Based on this observation we make the assumption that we are close to the L_3 spin coherent state. We thus have $s \gg \langle a^*a \rangle$ and $\lambda_3 \approx s$, hence ζ is linear in s . Furthermore we can order the weights, such that $w_1 \leq w_2 \leq w_3$ to minimize the expectation value. Now we take the limit and let s becomes large, the operator converges to harmonic oscillator

$$H(\mathbf{w})_{\text{HP}} = w_1 X^2 + w_2 P^2. \quad (53)$$

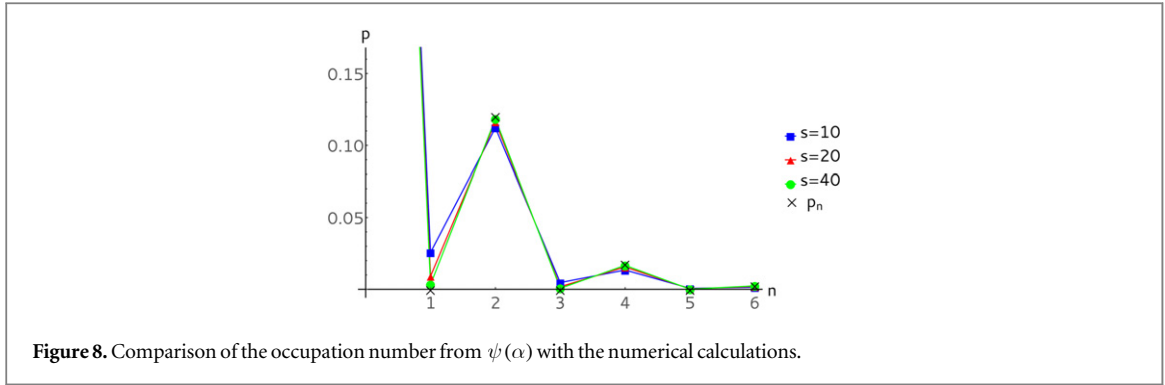


Figure 8. Comparison of the occupation number from $\psi(\alpha)$ with the numerical calculations.

Here we use that the expectation value of the harmonic oscillator is translation-invariant in phase space, so that we can choose ξ and η to be zero. The state which minimizes the expectation of this operator is simply the harmonic oscillator ground state $\psi(m, \omega)$, with $m = \frac{1}{2w_2}$ and $\omega = \sqrt{4w_1w_2}$. In the following these will be combined in the parameter $\alpha := m\omega = \sqrt{\frac{w_1}{w_2}}$. For the comparison of this result with numerical calculations using the above described algorithm, we must express these ground states in a common basis $|n\rangle_{\text{HP}}$, i.e. decomposing $\psi(\alpha)$ in the basis of a harmonic oscillator with $\alpha = 1$. This transformation is given by

$$\psi_n := \langle n | \psi(\alpha) \rangle = \frac{(\alpha)^{\frac{1}{4}}}{\sqrt{2^n \pi n!}} \int H_n(x) \exp\left(-\frac{(1+\alpha)}{2}x^2\right) dx, \quad (54)$$

which is zero for odd n and can be solved for even n through

$$\int H_n(x) \exp(-cx^2) dx = \frac{i^{3n} \sqrt{\pi} n! (c-1)^{\frac{n}{2}}}{\frac{n}{2}! c^{\frac{n+1}{2}}}. \quad (55)$$

The corresponding probability distribution is given by

$$p_n := |\psi_n|^2 = \frac{\sqrt{\alpha}}{1+\alpha} \frac{(1-\alpha)^n}{(2+2\alpha)^n} \frac{n!}{\left(\frac{n}{2}!\right)^2} (1+(-1)^n). \quad (56)$$

Because this is zero for odd n , we can set $n = 2k$ and get

$$p_{2k} = \frac{2\sqrt{\alpha}}{1+\alpha} \frac{(1-\alpha)^{2k}}{(2+2\alpha)^{2k}} \binom{2k}{k}. \quad (57)$$

The above approximation does not necessarily yield the optimal states and it is not rigorously justified so far. As a first step, we compare the distribution p_n with numerically determined ones for finite s . These tend to converge as shown in figure 8.

Theorem 4. *The lower bound of the asymptotic uncertainty region on a scale of $1/s$ is fully described by the generalized Robertson inequality (48) and is saturated by the states $\psi(\alpha)$.*

Proof. First we will show that the approximation (51) is justified for $\psi(\alpha)$ and evaluate the corresponding asymptotic variances. While the generalized Robertson inequality gets stronger for increasing s , every extremal point of the corresponding boundary is attained by $\psi(\alpha)$, which will prove the above statement. Moreover, by truncating the sequence $\psi_n(\alpha)$ at $n = 2s + 1$ and renormalizing, we get a sequence of spin- s states well approximating $\psi(\alpha)$ as s goes to infinity. With this in mind, we will prove the above statement in two steps:

(i) On the one hand we have to verify that $\lim_{s \rightarrow \infty} \frac{1}{\sqrt{s}} L_+ |\psi(\alpha)\rangle = \sqrt{2} a |\psi(\alpha)\rangle$ and

$\lim_{s \rightarrow \infty} \frac{1}{\sqrt{s}} L_- |\psi(\alpha)\rangle = \sqrt{2} a^* |\psi(\alpha)\rangle$ which is true if $\psi(\alpha)$ is in the domain of a^*a . On the other hand we have to show that the term $\frac{w_3}{s}(s - a^*a - \zeta)^2$ from (52) will vanish for $\psi(\alpha)$ with $\zeta = \langle \psi(\alpha) | s - a^*a | \psi(\alpha) \rangle$ as s goes to infinity. Both requirements are fulfilled if the moments $\langle \psi(\alpha) | a^*a | \psi(\alpha) \rangle$ and $\langle \psi(\alpha) | (a^*a)^2 | \psi(\alpha) \rangle$ are finite. In the Holstein–Primakoff occupation basis $|n\rangle_{\text{HP}}$, these moments are given by series of the form

$$\sum_{n=0}^{\infty} n^c p_n = \sum_{k=0}^{\infty} (2k)^c p_{2k}, \quad (58)$$

and can be computed as derivatives using the generating function

$$\frac{c_1}{\sqrt{1-4x}} = c_1 \sum_{n=0}^{\infty} x^n \binom{2n}{n} \quad (59)$$

of the probability distribution p_{2k} (57). By straightforward calculations we get

$$\langle \psi(\alpha) | a^* a | \psi(\alpha) \rangle = \frac{(1-\alpha)^2}{4\alpha}, \quad (60)$$

$$\langle \psi(\alpha) | (a^* a)^2 | \psi(\alpha) \rangle = \frac{3(1-\alpha)^4}{16\alpha^2} + \frac{(1-\alpha)^2}{2\alpha}, \quad (61)$$

which is finite for $\alpha > 0$.

(ii) Now the asymptotic variances for $\psi(a)$ can be determined. For $\psi(\alpha)$ the operators $\frac{1}{\sqrt{s}}L_1$, $\frac{1}{\sqrt{s}}L_2$ and $\frac{1}{\sqrt{s}}L_3$ converge to P , Q and a multiple of the identity, we obtain

$$\lim_{s \rightarrow \infty} \frac{1}{s} \Delta_{\psi(\alpha)}^2(L_1) = \Delta_{\psi(\alpha)}^2(Q) = \frac{1}{2\alpha}, \quad (62)$$

$$\lim_{s \rightarrow \infty} \frac{1}{s} \Delta_{\psi(\alpha)}^2(L_2) = \Delta_{\psi(\alpha)}^2(P) = \frac{\alpha}{2}, \quad (63)$$

$$\lim_{s \rightarrow \infty} \frac{1}{s} \Delta_{\psi(\alpha)}^2(L_3) = \lim_{s \rightarrow \infty} \frac{1}{s} \Delta_{\psi(\alpha)}^2(s - a^* a) = 0. \quad (64)$$

This set of variance triples $\left(\frac{1}{2\alpha}, \frac{\alpha}{2}, 0\right)$ saturates the asymptotic generalized Robertson bound (48). Moreover they describe the extremal boundary curves, see the proof of theorem 2, of the associated uncertainty region. \square

3. Preparation uncertainty: special topics

3.1. The vector model and moment problems

This may be a good place to comment on the so-called vector model of angular momentum, as it was suggested by old quantum theory. It still seems to be quite popular in teaching, although theoreticians tend to deride it as ridiculously classical and obviously inconsistent. Indeed, its two-particle version gives manifestly false predictions even for spin-1/2, as witnessed by Bell's (CHSH) inequality. Since *any* local classical model fails this test, not much can be learned about angular momentum from this observation. Therefore we consider here only the one-particle version, and try to sort out how far it can be trusted.

The basic rationale of the vector model is shown in figure 9: angular momentum is thought of as a classical random variable taking values on a sphere of radius $r_s = \sqrt{s(s+1)}$. For an eigenstate $|m\rangle$ the corresponding classical distribution is supposed to be concentrated at latitude m , and uniform with respect to rotations around the three-axis. The expectation value of this distribution is $(0, 0, m)$. Moreover, its matrix of second moments is also diagonal, since the coordinate axes are clearly the inertial axes of a mass uniformly distributed on a circle of fixed latitude. One readily checks that all second moments are the same as for the corresponding quantum state. This can be generalized:

Proposition 5. *For any quantum state ρ on \mathcal{H} there is a classical probability distribution μ on the sphere of radius $\sqrt{s(s+1)}$ which has the same first and second moments as the angular momentum in ρ , i.e.*

$$m_j = \int \mu(dx) x_j = \text{tr } \rho L_j \quad \text{and} \quad M_{jk} = \int \mu(dx) x_j x_k = \Re \text{tr } \rho L_j L_k. \quad (65)$$

For the proof we only need a characterization of the moments (m_j, M_{jk}) of probability measures on a sphere of radius r , which turns out to be quite simple. This in turn provides an immediate proof of proposition 5, since the quantum moments $M_{jk} = \Re \text{tr } \rho L_j L_k$, $m_j = \text{tr } \rho L_j$ obviously have the required properties with radius $r^2 = s(s+1)$.

Lemma 6. *Let $(m_j) \in \mathbb{R}$ for $j = 1, 2, 3$, and $(M_{jk})_{j,k=1}^3$ a real symmetric 3×3 -matrix. Then these numbers are the first and second moments of a probability distribution on the sphere of radius r if and only if $\text{tr } M = r^2$ and the covariance matrix $M_{jk} - m_j m_k$ is positive semi-definite.*

Proof. Necessity is obvious because covariance matrices are always positive and the function $v \mapsto \sum_j v_j^2 = r^2$ is constant on the sphere. For the converse consider the set K of moments (m, M) of probability measures on the

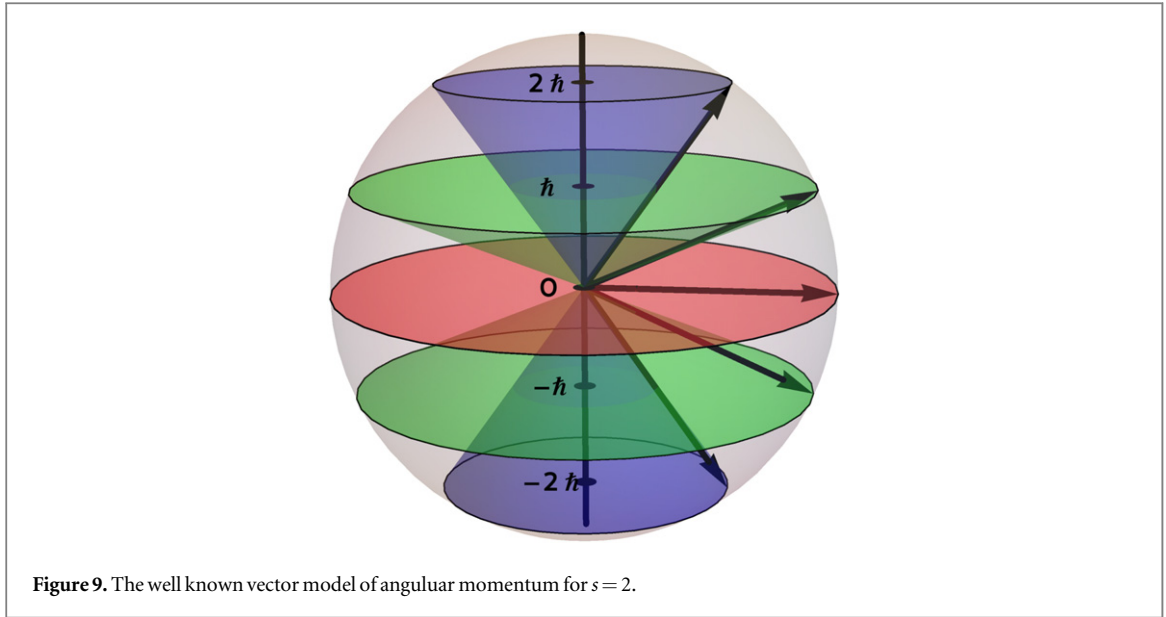


Figure 9. The well known vector model of angular momentum for $s = 2$.

sphere. This is a compact convex set, which we can think of as embedded into $3 + 6 - 1$ dimensional real space, because the real symmetric matrix M is specified by six parameters, and we have an additional linear constraint $\text{tr } M = r^2$. By the separation theorems for compact convex sets the set K is therefore completely characterized by a collection of affine inequalities

$$f(m, M) = \text{tr } AM - b \cdot m + \gamma \geq 0, \quad (66)$$

where A is real symmetric, $b \in \mathbb{R}^3$ with the dot indicating scalar product, and $\gamma \in \mathbb{R}$. The functionals for which these inequalities have to be satisfied are precisely those for which the above inequality holds for all *pure* probability measures, i.e., for $M_{jk} = v_j v_k m_j = v_j$ for some $v \in \mathbb{R}^3$ on the sphere. In this case we slightly abuse notation and write $f(m, M) = f(v)$.

Not all inequalities are needed to characterize K , but only the extremal ones, which furnish a minimal subset from which all the others follow as linear combinations with positive scalar factors. In particular, we can assume that f is not strictly positive, so has a zero $f(u) = 0$, which then also has to be a minimum. The extremality condition gives $2Au - b = 2\lambda u$, where $\lambda \in \mathbb{R}$ is a Lagrange multiplier. This determines b , and from $f(u) = 0$ we get γ , so that we can rewrite

$$f(v) = (v - u) \cdot A(v - u) + 2\lambda u \cdot (v - u). \quad (67)$$

Now since u, v lie on a sphere of radius r , we can write $2u \cdot (v - u) = -(v - u)^2 + (v^2 - u^2) = -(v - u)^2$, so we can combine the two terms, and obtain again the form (67), with A modified by a multiple of the identity, and $\lambda = 0$.

It remains to determine all real symmetric matrices A such that $(v - u) \cdot A(v - u) \geq 0$, whenever u, v lie on a sphere of radius r . Equivalently, $\xi \cdot A\xi \geq 0$ for all multiples of vectors of the form $v - u$. But this set is dense in \mathbb{R}^3 . Hence the desired condition is just the positive semi-definiteness of A . The resulting inequality for (m, M) can be rewritten in terms of the covariance matrix $V_{jk} = M_{jk} - m_j m_k$ as

$$f(m, M) = \text{tr } VA + (m - u) \cdot A(m - u) \geq 0. \quad (68)$$

Since the second term is anyhow positive, the positive semidefiniteness of V is sufficient for all these inequalities. This shows the sufficiency of the conditions stated in the lemma. \square

Let us make some remarks, which all fit into a fruitful analogy here with the phase space case, i.e., the case of two canonical operators P, Q , and *moment problems* posed in the respective contexts.

- (1) The phase space analogue of proposition 5 is the statement that for any quantum state the first and second moments can also be realized by a classical probability distribution on phase space. Of course, not all classically allowed first and second moments can arise in this way: This is just the theme of preparation uncertainty relations.
- (2) The classical probability measure μ is not uniquely defined by ρ . For example, the density operator $\rho = (2s + 1)^{-1} \mathbb{1}$ can either be represented by the uniform distribution on the sphere, or by an equal-weight mixture of the distributions with constant latitude m (in any direction). In the phase space case it is well-

known that with the given, quantum-realizable moments one can always find a *Gaussian state*, which is defined as the distribution with the maximal entropy given those moments. The same idea also works for angular momentum, and it gives probability densities which are the exponential of a quadratic form in the variables. In contrast to the phase space case, when approaching eigenstates (any direction, any m) this entropy will go to $-\infty$, since for eigenstates only the singular measures depicted in figure 9 can be used.

- (3) Proposition 5 is certainly false if we include higher than second moments. For example, consider a pure qubit state with $m = +1/2$. Without loss of generality we can choose the measure μ invariant under rotations around the three-axis. Since μ must be concentrated on $m = 1/2$, this uniquely fixes the measure μ , and hence the moments to all orders. Now consider a direction \mathbf{e} which is at an angle strictly between 0 and $\pi/2$ to \mathbf{e}_3 . Then the quantum expectation of $(\mathbf{e} \cdot \mathbf{L})^3 = \mathbf{e} \cdot \mathbf{L}/4$ is $(\mathbf{e} \cdot \mathbf{e}_3)/8$, but the classical expectation of $(x \cdot \mathbf{e})^3$ is larger, reflecting the nonlinearity of the cube function.
- (4) The quantum analogue of the classical Hamburger moment would be to reconstruct a quantum state from the set of moments, i.e., the expectations of the monomials in the basic operators (P , Q , or L_1 , L_2 , L_3). Commutation relations impose some constraints on these moments, so that in the end only monomials like $L_1^{n_1} L_2^{n_2} L_3^{n_3}$ need to be considered. Of course, the expectation values of such operators will generally be complex numbers. Can we do the reconstruction for arbitrary states in the spin- s representation? Indeed, we can, and it is actually much easier than in the phase space case since only finitely many moments suffice. The basic observation is that the moments fix all expectations on the von Neumann algebra \mathcal{A} generated by the L_i . Because the representation is irreducible, the commutant of this algebra consists of the multiples of the identity. Hence \mathcal{A} must be the full matrix algebra, and the state is uniquely determined. That finitely moments suffice is clear because $\dim \mathcal{A} < \infty$.
- (5) Noncommutative moment problems are plagued by ‘operator ordering’ issues. But in some sense we have already adopted a standard ‘symmetrized’ solution for operator ordering, namely to form moments only of the operators $\mathbf{e} \cdot \mathbf{L}$ for all fixed \mathbf{e} . This is analogous in the phase space case to considering the moments of linear combinations of P and Q . Now, famously, the full distributions of all such combinations are correctly rendered by the Wigner distribution function, which is itself hardly ever positive [15]. The analogy to the angular momentum case is immediate. So what do we get if we accept ‘quasi-probability distributions’? Can every state be represented like that? This is answered by the following proposition.

Proposition 7. *Let ρ be a quantum state in the spin- s representation. Then there is a unique tempered distribution \mathcal{W}_ρ on \mathbb{R}^3 , which is formally real, has support in a ball of radius s and satisfies, for all \mathbf{e} and $n \in \mathbb{N}$*

$$\int d\eta \mathcal{W}_\rho(\eta) (\mathbf{e} \cdot \eta)^n = \text{tr} \left(\rho (\mathbf{e} \cdot \mathbf{L})^n \right). \quad (69)$$

Proof. We can compute the Fourier transform of \mathcal{W}_ρ directly from (69), by multiplying with $(i\mathbf{k})^n/n!$ and summing over n . This turns the left side into the Fourier integral over \mathcal{W}_ρ allowing the sum to be evaluated also on the right-hand side:

$$\widehat{\mathcal{W}}_\rho(\mathbf{k}) = \int d\eta \mathcal{W}_\rho(\eta) e^{i\mathbf{k} \cdot \eta} = \text{tr} \left(\rho e^{i\mathbf{k} \cdot \mathbf{L}} \right), \quad (70)$$

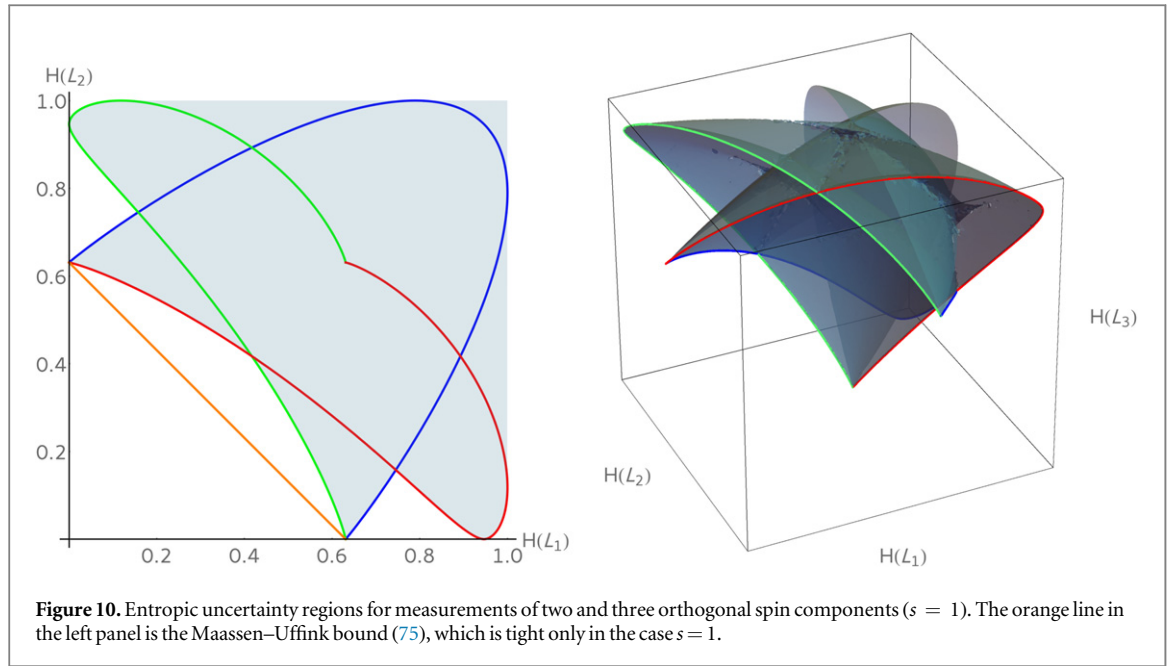
where $\mathbf{k} = k\mathbf{e}$. Strictly speaking this computation should be regularized by multiplying with an arbitrary test function before summation, but this would lead to the same explicit representation of the Fourier transform $\widehat{\mathcal{W}}_\rho$ as a bounded \mathcal{C}^∞ -function. This shows that the desired tempered distribution \mathcal{W}_ρ is essentially unique, and can be defined for every ρ . It is formally real, because $\widehat{\mathcal{W}}_\rho(-\mathbf{k}) = \overline{\widehat{\mathcal{W}}_\rho(\mathbf{k})}$. For the claim about the support we invoke the distributional version of the Paley–Wiener–Schwartz theorem [12 theorem 7.3.1]. According to that theorem we need to show only that for real vectors \mathbf{k} , κ the estimate

$$\left| \text{tr} \left(\rho e^{i(\mathbf{k} + i\kappa)\mathbf{L}} \right) \right| \leq C e^{s|\kappa|} \quad (71)$$

holds for some constant C . Treating the sum in the exponential by the Trotter formula, which we may, because these are finite dimensional matrices, we get

$$\| e^{i(\mathbf{k} + i\kappa)\mathbf{L}} \| = \lim_{N \rightarrow \infty} \left\| \left(e^{i(\mathbf{k}/N)\mathbf{L}} e^{(-\kappa/N)\mathbf{L}} \right)^N \right\| \leq \lim_{N \rightarrow \infty} \left(e^{(\kappa/N)\|\mathbf{L}\|} \right)^N = e^{s|\kappa|}. \quad (72)$$

Clearly this implies the desired estimate. \square



This proposition is remarkable in comparison to proposition 5: if we insist on positivity but require only the first two moments to be correct, the vector model requires a sphere of radius $\sqrt{s(s+1)}$. On the other hand, if we waive instead the positivity of the classical distribution, we are forced to use a ball of radius s .

One might ask at this point, in continuation of the above analogies to the phase space case, whether there are ‘classical’ states, for which the Wigner function \mathcal{W}_ρ is a positive function. However, this is easily seen to be impossible. Indeed, when the marginal of a proper probability distribution has support on $\{-s, \dots, s\}$, the measure itself has to have support on a union of hyperplanes $\{\eta \mid \eta \cdot \mathbf{e} = m\}$. But these families of hyperplanes, drawn for various \mathbf{e} have empty intersection, contradicting the normalization of the measure.

The main use of Wigner functions on phase space is the visualization of quantum states. Unfortunately, the much more singular nature of \mathcal{W}_ρ for angular momentum will prevent these Wigner functions from becoming similarly popular. This irregularity can be tamed by replacing on the right-hand side of (70)

$$e^{i\mathbf{k} \cdot \mathbf{L}} \mapsto e^{ik_1 L_1} e^{ik_2 L_2} e^{ik_3 L_3}. \quad (73)$$

The corresponding distribution is then a sum of point measures sitting on a finite cubical grid [6]. This may actually be useful in quantum information, where it relates to a discrete phase space structure over the cyclic group of $d = 2s + 1$ elements. However, for angular momentum proper we find this breaking of rotational symmetry abhorrent.

3.2. Entropic uncertainty

In this section we will have a look at the entropic uncertainty relations. Given a measurement of a Hermitian operator $A = \sum_i a_i P_i$, with eigenprojectors P_i , the probability of obtaining the i th measurement outcome will be denoted by $\pi_i(\rho) = \text{tr}(\rho P_i)$ and the associated probability distribution as $\pi(\rho) = \{\pi_1(\rho), \dots, \pi_d(\rho)\}$. Then the output entropy of A in the state ρ is defined as the Shannon entropy of $\pi(\rho)$

$$H(A, \rho) := H(\pi(\rho)) = -\sum_{i=1}^d \pi_i(\rho) \log_d(\pi_i(\rho)), \quad (74)$$

which serves as an uncertainty measure. Note that we normalize the Shannon entropy by its maximal value $\log d$ so that all occurring entropies are bounded by 1. In contrast to the variance, the entropy of a probability distribution does not change by permuting or rescaling the measurement outcomes and so only depends on the choice of the P_i (up to permutations) and not on the eigenvalues a_i . This implies that an entropic uncertainty relation, which constrains the output entropies of two (for simplicity non-degenerate) observables A, B , only depends on the unitary operator U connecting the respective eigenbases. A well-known bound in this setting is the general Maassen–Uffink bound [19]

$$H(A, \rho) + H(B, \rho) \geq -2 \log_d(c), \quad (75)$$

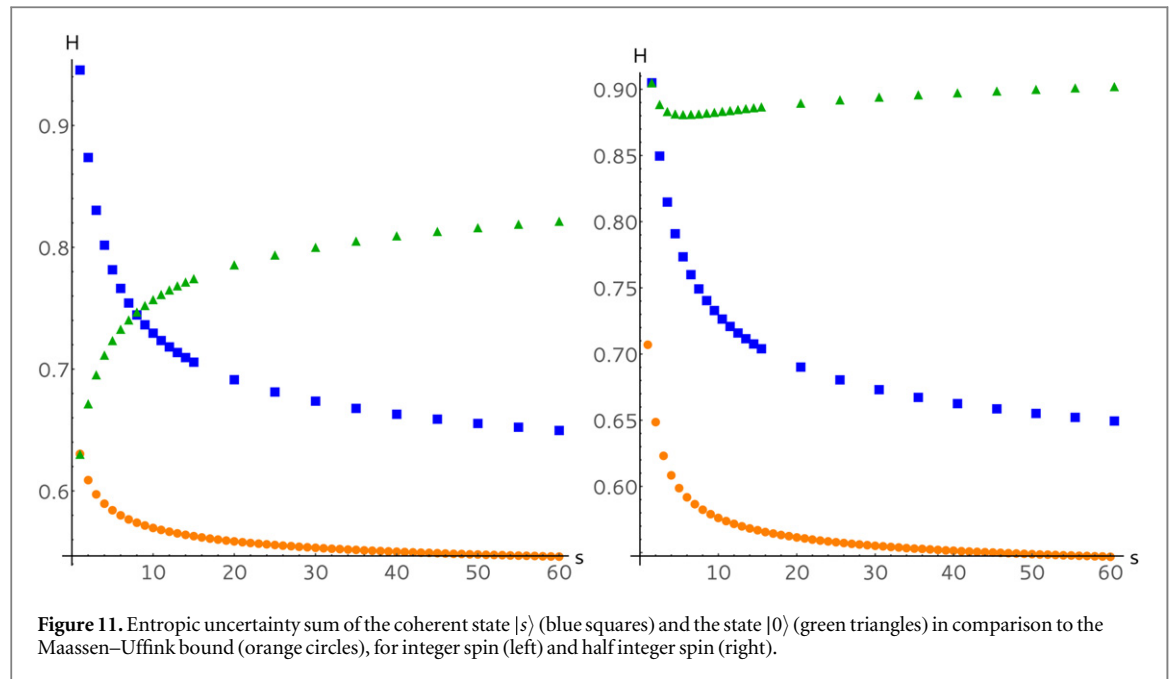


Figure 11. Entropic uncertainty sum of the coherent state $|s\rangle$ (blue squares) and the state $|0\rangle$ (green triangles) in comparison to the Maassen–Uffink bound (orange circles), for integer spin (left) and half integer spin (right).

$$c = c(U) = \max_{ij} |U_{ij}|. \quad (76)$$

For angular momentum measurements in two orthogonal directions the connecting unitary operators are rotations, i.e., given as a rotation by $\pi/2$ around the third coordinate axis according to the spin- s representation of $SO(3)$ on \mathbb{C}^{2s+1} . For arbitrary angles these representations are called Wigner- D matrices [1] and will also be used in section 4.5. It turns out that the Maassen–Uffink bound is in general not optimal, but describes precisely the uncertainty region for $s \rightarrow \infty$.

For spin $s = 1$, the uncertainty region can still be reliably investigated by parameterizing the set of pure states in the L_3 eigenbasis. Numerics suggests that real valued states and their permutations characterize the lower bound of the uncertainty region. The resulting uncertainty regions for two and three components are shown in figure 10, which should be compared directly to figures 4(left) and 3(right).

The marked lines in figure 10 correspond to states of a form $|\phi(t)\rangle := \left(\frac{\cos(t)}{\sqrt{2}}, \sin(t), \frac{\cos(t)}{\sqrt{2}} \right)$, written in L_3 -eigenbasis, and their permutations. They correspond exactly to the extremal curves found for variance uncertainty as shown in figure 3. Remarkably, however, the ordering of ‘uncertainties’ turns out to be different in the two cases. Consider the L_2 -eigenstates, which are shown in the left panels of both 2D figures 4 and 10 as the points on the horizontal axis. The respective L_1 -probability distribution are $(1/4, 1/2, 1/4)$ for $m = \pm 1$ and $(1/2, 0, 1/2)$ for $m = 0$. The former distribution has the larger entropy $((3/2)\log_2 2 > \log_2 2)$ but the smaller variance $(1/2 < 1)$.

For larger s this inversion no longer holds. Figure 11 shows this effect. Because we can exchange the roles of L_1 and L_2 by a unitary rotation, the uncertainty diagrams are symmetric with respect to the diagonal. Therefore the optimal linear bound must be of the form (75), with a suitable c . The entropy sums for the eigenstates $|m\rangle$ with minimal and maximal $|m|$ are shown in figure 11. For all half-integer s and for integer $s > 7$ the coherent state $|s\rangle$ produces not only the lowest variance, but also the lowest entropy. Figure 11 also shows the Maassen–Uffink bound, which has been computed by Sánchez–Ruiz [29]. It is attained for the overlap of two spin coherent states and is given by

$$c^2 = 2^{-2s} \binom{2s}{[s + 1/2]}, \quad (77)$$

$s = 1$ seems to be the only case in which the bound is tight.

However, for large s , the bound is again optimal, as the following result shows.

Proposition 8. *In the limit $s \rightarrow \infty$ the optimal lower bound on the entropic uncertainty region of L_1 and L_2 is given by the Maassen–Uffink inequality, which converges to*

$$H(L_1, \rho) + H(L_2, \rho) \geq \frac{1}{2}. \quad (78)$$

Proof. As a first step we will compute the asymptotic behavior of the bound $-\log_{2s+1}[c^2]$, with c given in (77). Expanding the central binomial coefficient in factorials and using the Stirling approximation up to order $\log_{2s+1}(s)$ gives

$$\lim_{s \rightarrow \infty} -\log_{2s+1}[c^2] = \lim_{s \rightarrow \infty} -\log_{2s+1} \left[2^{-2s} \binom{2s}{s+1/2} \right] = 1/2, \quad (79)$$

which proves the convergence of the Maassen–Uffink bound to the right-hand side of (78).

In order to show that this bound describes the asymptotic uncertainty region, we have to exhibit sequences of states saturating for every point on the boundary curve. We first show that the endpoint $(0, 1/2)$ is asymptotically attained by the L_1 -eigenstates $|s\rangle$: the output entropy of $|s\rangle$ in the L_1 basis is always zero, whereas the output entropy in the L_2 basis can be evaluated as $H(L_2, |s\rangle) = H(L_1, U_3 |s\rangle)$, with $U_3 = e^{-\frac{i\pi}{2}L_3}$. Because $|s\rangle$ has maximal quantum number m , the probability amplitudes of $U_3 |s\rangle$ are given by the last column of the Wigner- D Matrix $U_3 = D(0, \pi/2, 0)$. By expanding the Wigner- D matrix in terms of Jacobi polynomials [1], one can verify that $\pi_m(U_3 |s\rangle) = |\langle m | U_3 |s\rangle|^2$ is a binomial distribution in m , symmetric on the domain $\{-s, s\}$. The entropy of this distribution is $-1/2 \log_{2s+1} \left(2\pi e \frac{2s+1}{4} \right) + \mathcal{O}(\frac{1}{n})$, which converges to $\frac{1}{2}$ as s goes to infinity, hence (78) can be saturated:

$$\lim_{s \rightarrow \infty} (H(L_1, |s\rangle), H(L_2, |s\rangle)) = \left(0, \frac{1}{2}\right). \quad (80)$$

Finally we have to construct states saturating every point of this bound. For $|s\rangle$, still in the eigenbasis of L_1 , and arbitrary s , we define a family of unit vectors $|\psi_\alpha\rangle$ as

$$|\psi_\alpha\rangle = c_\alpha (\cos(\alpha)|s\rangle + \sin(\alpha)U_3 |s\rangle). \quad (81)$$

The L_1 -outcome probability distribution associated with this vector is

$$\pi_m(\psi_\alpha) \approx \cos(\alpha)^2 \delta_{ms} + \sin(\alpha)^2 |\langle m | U_3 |s\rangle|^2, \quad (82)$$

because the two probability distributions have practically no overlap for large s , and hence also $c_\alpha \approx 1$. For the same reason we can evaluate the entropy of $\pi(\Psi_\alpha)$ as a sum, obtaining

$$H(L_1, |\psi_\alpha\rangle) \approx \sin^2(\alpha) H(L_1, U_3 |s\rangle) + \frac{H_2(\cos^2(\alpha), \sin^2(\alpha))}{\log_2(2s+1)} \approx \frac{1}{2} \sin^2(\alpha), \quad (83)$$

where H_2 is the binary entropy function. In the L_2 -basis the roles of the two terms in ψ_α are exchanged, and we get

$$H(L_2, |\psi_\alpha\rangle) \approx \frac{1}{2} \cos^2(\alpha). \quad (84)$$

Hence the sequence ψ_α realizes the point $\frac{1}{2}(\sin^2(\alpha), \cos^2(\alpha))$ on the boundary. \square

4. Measurement uncertainty

4.1. Introduction

As mentioned in the introduction, a measurement uncertainty relation is a quantitative bound on the accuracy with which two observables can be measured approximately on the same device. Already in Kennard's 1927 paper [17] it is clearly stated that in quantum mechanics the notion of a 'true value' loses its meaning, so that we should not think of 'measurement error' as the deviation of the observed value from a true value. What we can always do, however, is to compare the performance of two measuring devices, one of which is an (perhaps hypothetical) 'ideal' measurement and the other an approximate one. The only requirement is that these two measurements give outputs which lie in the same space X and whose distance is somehow defined. A good approximate measurement is then one which will give, on every input state, almost the same output distribution as the ideal one. This operational focus on the output distributions is also in keeping with the way one would detect a disturbance of the system. Consider how we discover that trying to detect through which of two slits the

particles pass disturbs them: the interference pattern, i.e., the output distribution of the interferometer is changed and fringes washed out.

Two related ways to build up a quantitative comparison of distributions, and thereby a quantitative approximation measure between observables, were introduced in the papers [2, 5] and applied to the standard situation of a position and a momentum operator. These two notions, called calibration error and metric error will be described in the following subsections. Either way we get a natural figure of merit for an observable F jointly measuring two or more components of angular momentum. In fact, we will only treat the case where F jointly measures *all* components. By this we simply mean an observable whose output is not a single number but a vector $\boldsymbol{\eta}$. From this, one derives a ‘marginal measurement’ F^e of the \mathbf{e} -component by post-processing, i.e., by taking the \mathbf{e} -component $\mathbf{e} \cdot \boldsymbol{\eta}$ of the output vector as the output of F^e . These marginals can then be compared with the standard projection valued measurement of the angular momentum component $\mathbf{e} \cdot \mathbf{L}$. When $D(G, E)$ is the quantity chosen to characterize the error of an observable G with respect to the ideal reference E we get, in our special case

$$D_{\max}(F) = \max_{\mathbf{e}} D(F^e, \mathbf{e} \cdot \mathbf{L}) \quad (85)$$

is the desired figure of merit. This is the quantity which we will minimize. But first we have to be more explicit about the two choices for the error quantity $D(G, E)$. This will be done in the next two subsections.

4.2. Calibration error

The simpler one assumes that the ‘ideal’ observable is projection valued, so that we can produce states which have a very narrow distribution around one of its eigenvalues (or points in the continuous spectrum). In other words, we have some states available which come close to having a ‘true value’ in the sense that the ideal distribution is sharp around a known value. A good approximate measurement should then have an output distribution, which is also well peaked around this value. Thus we only have to compare probability distributions to δ -function like distributions, i.e., point measures δ_x with $x \in X$. This is straightforward, and we set, for any probability measure μ on the space X , $1 \leq \alpha < \infty$

$$D_\alpha(\mu, \delta_x) = \left(\int_X \mu(dy) D(y, x)^\alpha \right)^{1/\alpha}, \quad (86)$$

where D under the integral is the given metric on X . This could be called the power- α deviation of μ from the point x . We are mostly interested in *quadratic deviations*, i.e., $\alpha = 2$. However, in this section we keep α general, which causes no extra difficulty, but makes clear which numbers ‘2’ arise directly from the role of the averaging power α in (86) and similar equations.

We apply this now to F_ρ the output distribution F_ρ obtained by measuring the observable F on the input state ρ , and its ideal counterpart E_ρ . The ε -deviation or ε -calibration error of the observable F with respect to the ideal observable E is

$$\Delta_\alpha^\varepsilon(F, E) = \sup \left\{ D_\alpha(F_\rho, \delta_x) \mid D_\alpha(E_\rho, \delta_x) \leq \varepsilon \right\}, \quad (87)$$

where the supremum is over all $x \in X$ and ‘calibration states’ ρ , which are sharply concentrated on x up to quality ε . Note that as a function of ε this expression is decreasing as $\varepsilon \rightarrow 0$, because the supremum is taken over smaller and smaller sets. Therefore the limit exists, and we define the *calibration error* of F with respect to E by

$$\Delta_\alpha^c(F, E) = \lim_{\varepsilon \rightarrow 0} \Delta_\alpha^\varepsilon(F, E). \quad (88)$$

For observables E with discrete spectrum (like angular momentum components) we can also take $\varepsilon = 0$, in (87), and directly get $\Delta_\alpha^c(F, E) = \Delta_\alpha^0(F, E)$.

4.3. Metric error

A possible issue with the calibration error is that it describes the performance of F only on a very special subclass of states. On the one hand this makes it easier to determine it experimentally, but on the other hand we get no guarantee about the performance of the device on general inputs. Classically this problem does not arise, because broad distributions can be represented as mixtures of sharply peaked ones, and this allows us to give an estimate also on the similarity of output distributions for general inputs. The form of this estimate gives a good hint towards how to define the distance of probability distributions both of which are diffuse. Indeed suppose ρ is an input state such that $\rho = \int \mu(dx) \rho_x$, where ρ_x is an ε -calibration state at point x , and μ is an arbitrary probability measure. Then we can define measure γ on $X \times X$ by $\gamma(dx dy) = \mu(dx) F_{\rho_x}(dy)$, which gives the probability of the joint event of having a ‘true value’ $x \in dx$ and finding $y \in dy$. If one integrates out the x variable one gets the output distribution for ρ , because F_ρ is linear in ρ , and if one integrates out y one gets μ ,

because each F_{ρ_x} is normalized. To within ε this is the output distribution E_ρ , and with known calibration error we get the bound

$$\int \gamma(dx dy) D(x, y)^\alpha = \int \mu(dx) \int F_{\rho_x}(dy) D(x, y)^\alpha = \int \mu(dx) D_\alpha(F_{\rho_x}, \delta_x)^\alpha \leq \Delta_\alpha^\varepsilon(F, E)^\alpha. \quad (89)$$

This suggests the following definitions. For two probability distributions μ and ν on X we define a *coupling* to be a measure γ on $X \times X$ whose first marginal is μ and whose second marginal is ν . The set of couplings will be denoted by $\Gamma(\mu, \nu)$, and is always non-empty because it contains the product measure. We then define the *Wasserstein α -distance* of μ and ν as

$$D_\alpha(\mu, \nu) = \inf_{\gamma \in \Gamma(\mu, \nu)} \left(\int \gamma(dx dy) D(x, y)^\alpha \right)^{1/\alpha}. \quad (90)$$

This is also called a transport distance, because of the following interpretation, first seen by Gaspar Monge in the 18th century who considered the building of fortifications. We consider μ and ν as some distribution of earth, and the task of a builder who wants to transform distribution μ into distribution ν . The workers are paid by the bucket and the power α of the distance travelled with each bucket (giving a bonus pay on long distances). The builder's plan is precisely the coupling γ saying how many units are to be taken from x to y , and the integral is the total cost. The infimum is just the price of the optimal transport plan. The theory of such metrics is well developed, and we recommend the book of Villani [31] on the subject, but in the present context we only need some simple observations.

With a metric between probability distributions we define the distance of two observables as the worst case distance of their output distributions:

$$D_\alpha(F, E) := \sup_\rho D_\alpha(F_\rho, E_\rho). \quad (91)$$

For the connection between this *metric error* and the calibration error introduced above, note first that when ν is the point measure δ_x , and μ is arbitrary the product is the only coupling, and the two definitions $D_\alpha(\mu, \delta_x)$ from equations (86) and (90) coincide. Therefore, if $D_\alpha(E_\rho, \delta_x) \leq \varepsilon$, we have

$$D_\alpha(F_\rho, \delta_x) \leq D_\alpha(F_\rho, E_\rho) + D_\alpha(E_\rho, \delta_x) \leq D_\alpha(F, E) + \varepsilon. \quad (92)$$

By taking the supremum (87) and letting $\varepsilon \rightarrow 0$, we hence have

$$\Delta_\alpha^\varepsilon(F, E) \leq D_\alpha(F, E). \quad (93)$$

Intuitively, this merely indicates that for calibration we test deviations only in the small subset of highly concentrated states. Then (89) is a partial converse: if ρ has a convex decomposition into ε -concentrated states, $D_\alpha(F_\rho, \mu) \leq \Delta_\alpha^\varepsilon(F, E)$, and since $D_\alpha(\mu, E_\rho) \leq \varepsilon$ we get $D_\alpha(F_\rho, E_\rho) \leq \Delta_\alpha^\varepsilon(F, E) + \varepsilon$. In the classical case such a decomposition always exists, so we have equality $D_\alpha(F_\rho, E_\rho) = \Delta_\alpha^\varepsilon(F, E)$. In the quantum case, however, we not only have convex mixtures of sharply concentrated states but also coherent superpositions. Using these it is easy to build examples in which (94) is strict.

There is a second 'quasi-classical' setting, in which calibration and metric error coincide, and this will actually be used below. This is the case when F and E differ only by classical noise generated in the measuring apparatus. More formally this is described by a transition probability kernel $P(x, dy)$, which is for every x the probability measure in y describing the output of F , given that E has been given the value x . We can think of this as classical probabilistic post-processing or noise. It is, of course, not necessary that F actually operates in two steps, but only that it could be simulated in this way, i.e., the relation $F(dy) = \int E(dx) P(x, dy)$ holds. This is enough to conclude $\Delta_\alpha^\varepsilon(F, E) = D_\alpha(F, E)$, and to give a formula for both in terms of the size of the noise kernel P . In the following lemma the E -essential supremum of a measurable function f with respect to a measure E (denoted $E\text{-ess sup}_{x \in X} f(x)$) is the supremum of all λ such that the upper level set $\{x | f(x) > \lambda\}$ has non-zero E -measure. In our application E is the spectral measure of a component $\mathbf{e} \cdot \mathbf{L}$, so it is concentrated on the finite set $\{-s, \dots, s\}$. The essential supremum is then simply the maximum of f over this set.

Lemma 9. *Let E be a projection valued observable on a separable metric space (X, D) . Let F be an observable arising from E by post-processing with a transition probability kernel P . Then, for all α ,*

$$\Delta_\alpha^\varepsilon(F, E) = D_\alpha(F, E) = \left(E\text{-ess sup}_{x \in X} \int P(x, dy) D(x, y)^\alpha \right)^{1/\alpha}. \quad (94)$$

Proof. Let I, II, III be the three terms in this equation. Then I \leq II is given by (94). To show II \leq III, note that for any state ρ we get a coupling γ between F_ρ and E_ρ by $\gamma(dx dy) = E_\rho(dx)P(x, dy)$. Hence

$$D_\alpha(F_\rho, E_\rho)^\alpha \leq \int \gamma(dx dy) D(x, y)^\alpha = \int E_\rho(dx) \int P(x, dy) D(x, y)^\alpha.$$

We introduce the function $f(x) = \left(\int P(x, dy) D(x, y)^\alpha\right)^{1/\alpha}$ and split the integral with respect to E_ρ into an integral over $X_{>} = \{x | f(x) > t\}$ and an integral over its complement X_{\leq} , where and $t > E\text{-ess sup}_x f(x)$. Then, by definition of the essential supremum, E_ρ vanishes on $X_{>}$ and on X_{\leq} the integrand is bounded by t^α . Hence $D_\alpha(F_\rho, E_\rho)^\alpha < t^\alpha$. Taking the supremum over ρ and the α th root we get $D_\alpha(F, E) < t$ for every $t > \text{III}$, proving $\text{II} \leq \text{III}$.

It remains to show that $\text{III} \leq \text{I}$. This time we pick a $t < E\text{-ess sup}_x f(x)$. By definition, this means that $E(\{x | f(x) > t\}) \neq 0$. Now let $\varepsilon > 0$. Then because we have assumed X to be separable, it is covered by a countable collection of ε -balls $B_\varepsilon(x_i) = \{y \in X | D(x_i, y) \leq \varepsilon\}$, $i = 1, 2, \dots$. Hence due to countable additivity of E

$$0 \leq E(\{x | f(x) > t\}) \leq \sum_i E(B_\varepsilon(x_i) \cap \{x | f(x) > t\}). \quad (95)$$

Hence for some $z = x_i$ we have $E(B_\varepsilon(z) \cap \{x | f(x) > t\}) \neq 0$. Since E is projection valued, we find a state ρ such that the probability measure $E_\rho(dx)$ is concentrated on this set. In particular, $D_\alpha(E_\rho, \delta_z) \leq \varepsilon$. Moreover, $m^\alpha = \int E_\rho(dx) f(x)^\alpha \geq t^\alpha$ and

$$D_\alpha(F_\rho, \delta_z)^\alpha = \int E_\rho(dx) \int P(x, dy) D(z, y)^\alpha = \int E_\rho(dx) g(x)^\alpha, \quad (96)$$

where the second function defines a function g . We interpret these quantities as L^α -norms with respect to E_ρ , i.e., $m = \|f\|_\alpha$ and $D_\alpha(F_\rho, \delta_z) = \|g\|_\alpha$. Then by the triangle inequality

$$D_\alpha(F_\rho, \delta_z) \geq \|f\|_\alpha - \|f - g\|_\alpha \geq t - \|f - g\|_\alpha. \quad (97)$$

To get an upper bound on $\|f - g\|_\alpha$ note that the expression for $\|f - g\|_\alpha^\alpha$ is the integral over $E_\rho(dx)$ of

$$\left(\left(\int P(x, dy) D(x, y)^\alpha \right)^{1/\alpha} - \left(\int P(x, dy) D(z, y)^\alpha \right)^{1/\alpha} \right)^\alpha. \quad (98)$$

Again we can read the outer parenthesis as a difference of norms, namely the L^α -norms of the functions $h_x(y) = D(x, y)$ and $h_z(y) = D(z, y)$ with respect to integration by $P(x, dy)$ where x is considered a fixed parameter. But by the triangle inequality for the metric D we have $|h_x - h_z| \leq D(x, z)$ independent of y . Since the transition kernel P is a probability measure with respect to dy , we find that (98) is bounded above by $(\|h_x\|_\alpha - \|h_z\|_\alpha)^\alpha \leq \|h_x - h_z\|_\alpha^\alpha \leq D(x, z)^\alpha$. Hence in (97) we have

$$\|f - g\|_\alpha^\alpha \leq \int E_\rho(dx) D(x, z)^\alpha \leq \varepsilon^\alpha, \quad (99)$$

because by construction E_ρ has support in $B_\varepsilon(z)$. Combining the estimates we get $D_\alpha(F_\rho, \delta_z) \geq t - \varepsilon$. The supremum over all calibrating states can only increase the left-hand side, and on the right we use that the only condition on t was that $t < E\text{-ess sup}_x f(x)$, so that

$$\Delta_\alpha^\varepsilon(F, E) \geq E\text{-ess sup}_x f(x) - \varepsilon. \quad (100)$$

Now $\text{III} \leq \text{I}$ follows in the limit $\varepsilon \rightarrow 0$. □

In [5] a special case of this Lemma was used to show $\Delta^c = D$ for the position and momentum marginals of a covariant phase space measurement. In that case the noise kernel P is even translation invariant, i.e., the output of the marginal observable can be simulated by just adding some state-independent noise to the output of the ideal position or momentum observable. Such translation invariance makes no sense in the case of angular momentum, since the range of the outputs $m \in \{-s, \dots, s\}$ of the ideal observable is bounded. This is why the above generalization was needed, in which the noise can depend on the ideal output value. The reason for the existence of a post-processing kernel, however, will be the same as in the phase space case: the covariance of the joint measurement. Roughly speaking this makes the marginal corresponding to $\mathbf{e} \cdot \mathbf{L}$ invariant under rotations around the \mathbf{e} -axis, which in an irreducible representation means that it must be a function of $\mathbf{e} \cdot \mathbf{L}$. It is therefore crucial to argue that the optimal joint measurement is covariant, which will be done in the next section.

4.4. Covariant observables

Consider a general observable F with outcome space X . Suppose some group G acts on X , with the action written $(g, x) \mapsto gx$ as usual. Suppose that the group also acts as a symmetry group of the quantum system. That is, there is a representation $g \mapsto U_g$ of G by operators U_g , which are unitary or antiunitary, and satisfy the group law (possibly up to a phase factor). The observable F is then called covariant if $U_g^* F(S) U_g = F(g^{-1}S)$ for all $g \in G$ and every measurable set S . In other words, shifting the input state by U_g will result in the entire output distribution shifted by g . For our purposes it will be convenient to express this in terms of an action $F \mapsto T_g F$ of

G on the set of observables:

$$(T_g F)(S) = U_g F(g^{-1}S) U_g^*. \quad (101)$$

Then the covariant observables are precisely those for which $T_g F = F$ for all $g \in G$.

For angular momentum the group will be the rotation group with its action on the three-vectors ($X = \mathbb{R}^3$). The representation U is then up to a factor ± 1 . Alternatively, we can take G as the covering group $SU(2)$. Since the covariant observables are exactly the same this choice is completely equivalent.

Covariance is certainly a reasonable condition to impose on a ‘good’ observable, so it would make sense to study uncertainty relations just for these. However, there is no need for such an ad hoc restriction, because the minimum of uncertainty over all observables is anyway attained on a covariant one. The basic reason for this is that our figure of merit (85) does not single out a direction in space, so that it is invariant under the action T_g . We therefore only have to show that there is no symmetry breaking, i.e., the symmetric variational problem has a symmetric solution. This will be done in the following lemma.

Lemma 10. *For any observable F with an outcome set $X = \mathbb{R}^3$, and $1 \leq \alpha < \infty$ let*

$$D_{\max}(F) = \max_{\mathbf{e}} D_{\alpha}(F^{\mathbf{e}}, \mathbf{e} \cdot \mathbf{L}), \quad (102)$$

$$\Delta_{\max}(F) = \max_{\mathbf{e}} \Delta_{\alpha}^c(F^{\mathbf{e}}, \mathbf{e} \cdot \mathbf{L}). \quad (103)$$

Then

- (1) Both these functionals are invariant under the action T , and $D_{\max}(F)^{\alpha}$ and $\Delta_{\max}(F)^{\alpha}$ are convex.
- (2) Each of the infima $\inf_F D_{\max}(F)$ and $\inf_F \Delta_{\max}(F)$ is independent of whether it is taken over all observables or just the covariant ones.

Proof. By definition of $F^{\mathbf{e}}$ we have $(T_R F)^{\mathbf{e}} = U_R F^{R^{-1}\mathbf{e}} U_R^*$. When $E^{\mathbf{e}}$ denotes the spectral measure of $\mathbf{e} \cdot \mathbf{L}$, the relation $U_R \mathbf{L} U_R^* = R^{-1} \mathbf{L}$ similarly implies $U_R E^{R^{-1}\mathbf{e}} U_R^* = E^{\mathbf{e}}$. Moreover, due to the supremum over all states, $D_{\alpha}(F, E)$ does not change, if both observables are rotated with the same unitary. Hence

$$D_{\alpha}((T_R F)^{\mathbf{e}}, E^{\mathbf{e}}) = D_{\alpha}(U_R F^{R^{-1}\mathbf{e}} U_R^*, U_R E^{R^{-1}\mathbf{e}} U_R^*) = D_{\alpha}(F^{R^{-1}\mathbf{e}}, E^{R^{-1}\mathbf{e}}). \quad (104)$$

Hence the supremum over \mathbf{e} is unchanged and $D_{\max}(T_R F) = D_{\max}(F)$. For Δ note that we can carry out the limit $\varepsilon \rightarrow 0$ directly, because $\mathbf{e} \cdot \mathbf{L}$ has finite spectrum and states with $D(E^{\mathbf{e}}, \delta_x)$ small are norm-close to eigenstates with $x = m$. Hence

$$\Delta_{\alpha}^c(F^{\mathbf{e}}, \mathbf{e} \cdot \mathbf{L})^{\alpha} = \max \left\{ D(F_{|\psi\rangle\langle\psi|}^{\mathbf{e}}, \delta_m)^{\alpha} \mid m, \psi: \mathbf{e} \cdot \mathbf{L}\psi = m\psi \right\} = \max_{m, \psi} \int \langle \psi | F(\mathbf{x}) | \psi \rangle (\mathbf{e} \cdot \mathbf{x} - m)^{\alpha}. \quad (105)$$

If we now insert the definition of $T_{R^{\mathbf{e}}}$, rewrite the maximum over ψ in terms of $\psi' = U_R^* \psi$, and substitute $\mathbf{x}' = R^{-1} \mathbf{x}$ in the integral, we find $\Delta_{\alpha}^c((T_R F)^{\mathbf{e}}, \mathbf{e} \cdot \mathbf{L}) = \Delta_{\alpha}^c(F^{R^{-1}\mathbf{e}}, E^{R^{-1}\mathbf{e}})$, and again $\Delta_{\max}(T_R F) = \Delta_{\max}(F)$.

Convexity for D_{α} follows from the corresponding property of transport distances. Indeed let $\mu = \sum_k \lambda_k \mu_k$ and $\nu = \sum_k \lambda_k \nu_k$ be convex combinations of measures with the same weights λ_k , and let γ_k be a coupling between μ_k and ν_k . Then $\gamma = \sum_k \lambda_k \gamma_k$ is a coupling of the convex combinations. Moreover

$$D_{\alpha}(\mu, \nu)^{\alpha} \leq \int \gamma(\mathbf{x}, \mathbf{y}) D_{\alpha}(\mathbf{x}, \mathbf{y})^{\alpha} = \sum_k \lambda_k \int \gamma_k(\mathbf{x}, \mathbf{y}) D_{\alpha}(\mathbf{x}, \mathbf{y})^{\alpha}. \quad (106)$$

Here the first inequality holds because D_{α} is defined as the infimum over couplings. Then if we take the infimum over each of the γ_k , we get $D_{\alpha}(\mu, \nu)^{\alpha} \leq \sum_k \lambda_k D_{\alpha}(\mu_k, \nu_k)^{\alpha}$. In particular, $F \mapsto D_{\alpha}(F_{\rho}^{\mathbf{e}}, E_{\rho}^{\mathbf{e}})$ is convex. Since the pointwise supremum of convex functions is convex, D_{\max} , as the supremum with respect to ρ and \mathbf{e} , is convex. The same observation, applied to (105) with an additional supremum over \mathbf{e} , shows that Δ_{\max} is likewise convex.

Given any observable F we now form its average $\bar{F} = \int dR T_R F$ with respect to the normalized Haar measure dR . Then \bar{F} is covariant and $D_{\max}(\bar{F}) \leq \int dR D_{\max}(T_R F) = D_{\max}(F)$. Taking the infimum here, and using that \bar{F} is covariant, and all covariant observables are such averages ($\bar{F} = F$) gives the second inequality in

$$\inf_F D_{\max}(F) \leq \inf_{F \text{ covariant}} D_{\max}(F) \leq \inf_F D_{\max}(F), \quad (107)$$

while the first follows trivially because the covariant observables are a subset. Hence the two infima are the same, and the same argument also applies to Δ_{\max} , proving the second claim. \square

We could have included also a statement that the infima in this Lemma are all attained. The argument for that is the compactness of the set of observables (in a suitable topology) and the lower semi-continuity of D_{\max} and Δ_{\max} which follows, like convexity, from the representation of these functionals as the pointwise supremum of continuous functionals. However, since we will later anyhow explicitly exhibit minimizers, we will skip the abstract arguments. We also remark that one of the main difficulties in the position/momentum case [5] does not arise here: in contrast to the group of phase space translations, the rotation group is compact, so the average is an integral and not an ‘invariant mean’, which has the potential of producing singular measures with some support on infinitely far away points.

The main importance of this Lemma is to make the variational problem much more tractable. For covariant observables we have a fairly explicit parameterization, which allows us to explicitly compute the minimizers. In contrast, for the seemingly easier case of joint measurement of just two components covariance gives only a very weak constraint, and we were not able to complete the minimization.

To develop the form of covariant observables, let us first consider the case when the output vectors have a fixed length r . A plausible value would be $r = s$, but we will leave this open. In this case X reduces to a sphere of radius r , which is a homogeneous space for the rotation group. We could thus apply the covariant version [28, 32] of the Naimark [21] dilation theorem to obtain a complete classification. But we do not need this machinery in this elementary case. Note first, that $\text{tr } F(\text{d}\mathbf{x})$ is an invariant measure on the sphere, and all probability densities $\text{tr } \rho F(\text{d}\mathbf{x})$ are bounded by this measure (since ρ is bounded). Hence, for every ρ there is a bounded probability density with respect to the uniform measure. Since this depends linearly on ρ , it is given by a bounded operator depending on \mathbf{x} . The \mathbf{x} -dependence is then completely resolved by covariance, and it is sufficient to know this density at one point, say the north pole $\boldsymbol{\nu}$. Moreover, by covariance this density must commute with the rotations around $\boldsymbol{\nu}$, and is hence a linear combination of the eigenstates $|n\rangle\langle n|$ with $n \in \{-s, \dots, s\}$. The only choices to be made are hence the coefficients F_n of this linear combination. We write the resulting observable in terms of its integrals with an arbitrary function h on the sphere:

$$\begin{aligned} \int F(\text{d}\mathbf{x}) h(\mathbf{x}) &= (2s + 1) \int \text{d}R \, U_R \left(\sum_n F_n |n\rangle\langle n| \right) U_R^* h(r R \boldsymbol{\nu}) \\ &= (2s + 1) \int \frac{\sin \theta \, \text{d}\theta \, \text{d}\phi}{4\pi} \sum_n F_n U_{\theta, \phi} |n\rangle\langle n| U_{\theta, \phi}^* h(r \mathbf{e}_{\theta, \phi}). \end{aligned} \quad (108)$$

Here the first integral is over Haar measure on the rotation group (or $\text{SU}(2)$), whereas the second is expressed in polar coordinates (θ, ϕ) on the sphere, with $\mathbf{e}_{\theta, \phi}$ the corresponding unit vector, and $U_{\theta, \phi}$ some rotation rotating the north pole $\boldsymbol{\nu}$ to $\mathbf{e}_{\theta, \phi}$. It does not matter which rotation we choose, because $|n\rangle\langle n|$ is invariant with respect to rotations around the three-axis. The two expressions are related by introducing Euler angles on the rotation group and integrating out the initial rotation around the three-axis. The normalization factor $(2s + 1)$ is chosen so that the constraints on F_n are exactly $F_n \geq 0$ and $\sum_n F_n = 1$, i.e., the observable is represented as a convex combination of observables using only one fixed $|n\rangle\langle n|$ as the density. What changes when r is not fixed is simply that we get an additional integration over r , where the F_n may also depend on r . Effectively, we get a probability measure $F_n(\text{d}r)$ on $\{-s, \dots, s\} \times \mathbb{R}_+$ and the second version of (108) just becomes

$$\int F(\text{d}\mathbf{x}) h(\mathbf{x}) = \sum_n \int F_n(\text{d}r) (2s + 1) \int \frac{\sin \theta \, \text{d}\theta \, \text{d}\phi}{4\pi} U_{\theta, \phi} |n\rangle\langle n| U_{\theta, \phi}^* h(r \mathbf{e}_{\theta, \phi}). \quad (109)$$

The criterion for joint measurability does not depend on the full observable, but only on the marginals along the various directions \mathbf{e} . It is one of the direct consequences of covariance, evident from the proof of lemma 10, that $D_\alpha(F^\mathbf{e}, E^\mathbf{e})$ and $\Delta_\alpha(F^\mathbf{e}, E^\mathbf{e})$ do not depend on \mathbf{e} . We will therefore only consider the case $\mathbf{e} = \boldsymbol{\nu}$ in the following. In (109) this just means that we specialize to functions of the form $h(\mathbf{x}) = h_1(\mathbf{x} \cdot \boldsymbol{\nu})$ with $h_1: \mathbb{R} \rightarrow \mathbb{R}$. Thus in the integrand we get $h_1(r \mathbf{x}_{\theta, \phi} \cdot \boldsymbol{\nu}) = h_1(r \cos \theta)$, which no longer depends on ϕ . We can therefore carry out the ϕ -integration. The resulting operator will commute with rotations around the three-axis, so we can express it as a linear combination of operators $|m\rangle\langle m|$:

$$\begin{aligned} \int F^\nu(\text{d}\mathbf{x}) h_1(x) &= \sum_m |m\rangle\langle m| \int P(m, \text{d}\mathbf{x}) h_1(x) \quad \text{with} \\ \int P(m, \text{d}\mathbf{x}) h_1(x) &= \sum_n \int F_n(\text{d}r) (2s + 1) \int \frac{\sin \theta \, \text{d}\theta}{2} |\langle m | U_\theta | n \rangle|^2 h_1(r \cos \theta). \end{aligned} \quad (110)$$

The first line establishes the connection with the premise of lemma 9: for covariant observables the marginals can be simulated by an exact measurement of m , with a post-processing kernel P . Therefore, for covariant F we have $D_{\max}(F) = \Delta_{\max}(F)$. Since by lemma 10 the infimum of this quantity, say $\Delta_{\min}(s)$, is the same as the minimization over all observables. Therefore we can state the measurement uncertainty in the forms

$$D_{\max}(F) \geq \Delta_{\min}(s) \quad \text{and} \quad \Delta_{\max}(F) \geq \Delta_{\min}(s) \quad (111)$$

for *all* observables F , whether covariant or not. We will now compute $\Delta_{\min}(s)$, and show that both minima are attained for a unique covariant observable.

4.5. Minimal uncertainty

While the above holds for arbitrary exponent α , we will now restrict to the standard variance case, i.e. $\alpha = 2$. So far, we have derived that the optimal observable \mathbf{F} is covariant, leading to the parametrization (110). In particular, F^e arises from $\mathbf{e} \cdot \mathbf{L}$ by a transition probability kernel, so that metric and calibration error coincide. In the sequel, we will therefore only consider the calibration error, which is easier to evaluate. By covariance the calibration error $\Delta_2^e(F^e, \mathbf{e} \cdot \mathbf{L})$ is independent of \mathbf{e} , so we can choose $\mathbf{e} = \mathbf{e}_3$. Observing that for discrete valued observables we can take $\varepsilon = 0$ in (87), and we get from (110) the basic figure of merit

$$\begin{aligned} \Delta_2^e(F^e, \mathbf{e} \cdot \mathbf{L})^2 &= \Delta_2^e(F^3, L_3)^2 = \max_m \int \text{tr}(|m\rangle \langle m| F(d\eta)) (\mathbf{e}_3 \cdot \boldsymbol{\eta} - m)^2 \\ &= \max_m \sum_{n=-s}^s \int_0^\infty E_n(dr) (2s+1) \int_0^\pi \frac{\sin \theta d\theta}{2} |\langle n| U_\theta |m\rangle|^2 (r \cos \theta - m)^2. \end{aligned} \quad (112)$$

Before calculating the optimal case, we introduce the following lemma, which provides a more manageable expression $I(s, r, n, m)$ for the integral over θ , such that (112) reads

$$\Delta_2^e(F^3, L_3)^2 = \max_m \sum_{n=-s}^s \int_0^\infty E_n(dr) I(s, r, n, m). \quad (113)$$

Lemma 11. For $s \geq 1$ the integral $I(s, r, n, m)$ can be written as

$$I(s, r, n, m) = A_s(r, n) + m^2 B_s(r, n), \quad (114)$$

where

$$A_s(r, n) = \frac{r^2 s(s+1)(-2n^2 + 2s(s+1) - 1)}{s(s+1)(2s-1)(2s+3)} \quad (115)$$

and

$$B_s(r, n) = \frac{6n^2 r^2 - 2nr(4s(s+1) - 3) + s(s+1)(-2r^2 + 4s(s+1) - 3)}{s(s+1)(2s-1)(2s+3)}. \quad (116)$$

For $s = \frac{1}{2}$ we have

$$I\left(\frac{1}{2}, r, \pm \frac{1}{2}, m\right) = \frac{r^2}{3} \mp \frac{r}{3} + \frac{1}{4}. \quad (117)$$

Proof. Here we have to solve the integral

$$I(s, r, n, m) = (2s+1) \int_0^\pi \frac{\sin \theta d\theta}{2} |d_{nm}^{(s)}(\theta)|^2 (r \cos \theta - m)^2, \quad (118)$$

where $d_{nm}^{(s)}(\theta) = \langle n| U_\theta |m\rangle$ is the small Wigner d -matrix [1]. First we expand $d_{nm}^{(s)}(\theta)$ in terms of the Jacobi polynomials:

$$d_{nm}^{(s)}(\theta) = \left[\frac{(s+m)!(s-m)!}{(s+n)!(s-n)!} \right]^{\frac{1}{2}} \left(\sin \frac{\theta}{2} \right)^{m-n} \left(\cos \frac{\theta}{2} \right)^{n+m} P_{s-m}^{(m-n, m+n)}(\cos \theta). \quad (119)$$

In the following we use a recurrence relation for the Jacobi polynomials. This three-term relation does not hold for $s = 1/2$, so that we have to treat this case separately: the integrals $I\left(\frac{1}{2}, r, n, \pm \frac{1}{2}\right)$ can indeed be calculated directly from the above expression, and the results are given in the statement of the lemma. From now on we assume $s \geq 1$.

We can simplify some case distinctions by introducing $k = \min(s+m, s-m, s+n, s-n)$ and substitute the arising positive integers $s-m, m-n, m+n$ according to table 1, which is possible due to the symmetries of the Wigner d -matrix [1]. Our expression then depends implicitly on (s, m, n) through (μ, ν, k) :

Table 1. Wigner- d matrix substitution.

	k	μ	ν
(I)	$s + m$	$n - m$	$-n - m$
(II)	$s - m$	$m - n$	$n + m$
(III)	$s + n$	$m - n$	$-n - m$
(IV)	$s - n$	$n - m$	$n + m$

$$d_{nm}^{(s)}(\theta) = \left[\frac{k!(2s-k)!}{(k+\mu)!(k+\nu)!} \right]^{\frac{1}{2}} \left(\sin \frac{\theta}{2} \right)^{\mu} \left(\cos \frac{\theta}{2} \right)^{\nu} P_k^{(\mu,\nu)}(\cos \theta). \quad (120)$$

Substituting $x = \cos \theta$ yields

$$I(s, r, n, m) = \kappa(j, k, \mu, \nu) \int_{-1}^1 dx (1-x)^{\mu} (1+x)^{\nu} P_k^{(\mu,\nu)}(x) P_k^{(\mu,\nu)}(x) (rx - m)^2, \quad (121)$$

where

$$\kappa(s, k, \mu, \nu) = \left[\frac{(2s+1)k!(2s-k)!}{2^{\mu+\nu+1}(k+\mu)!(k+\nu)!} \right]. \quad (122)$$

This integral can be solved by expanding the factor $(rx - m)^2$, using the Jacobi polynomial orthogonality relation [30]

$$\int_{-1}^1 (1-z)^a (1+z)^b P_m^{(a,b)}(z) P_n^{(a,b)}(z) dz = \frac{(2^{a+b+1} \Gamma(a+n+1) \Gamma(b+n+1))}{\underbrace{n! (a+b+2n+1) \Gamma(a+b+n+1)}_{\omega(n,a,b)}} \delta_{m,n} \quad (123)$$

and the recurrence relation [30]

$$\begin{aligned} z P_n^{(a,b)}(z) &= \frac{(b^2 - a^2)}{\underbrace{(a+b+2n)(a+b+2n+2)}_{\alpha(n,a,b)}} P_n^{(a,b)}(z) \\ &+ \frac{(2(a+n)(b+n))}{\underbrace{(a+b+2n)(a+b+2n+1)}_{\beta(n,a,b)}} P_{n-1}^{(a,b)}(z) \\ &+ \frac{(2(n+1)(a+b+n+1))}{\underbrace{(a+b+2n+1)(a+b+2n+2)}_{\gamma(n,a,b)}} P_{n+1}^{(a,b)}(z). \end{aligned} \quad (124)$$

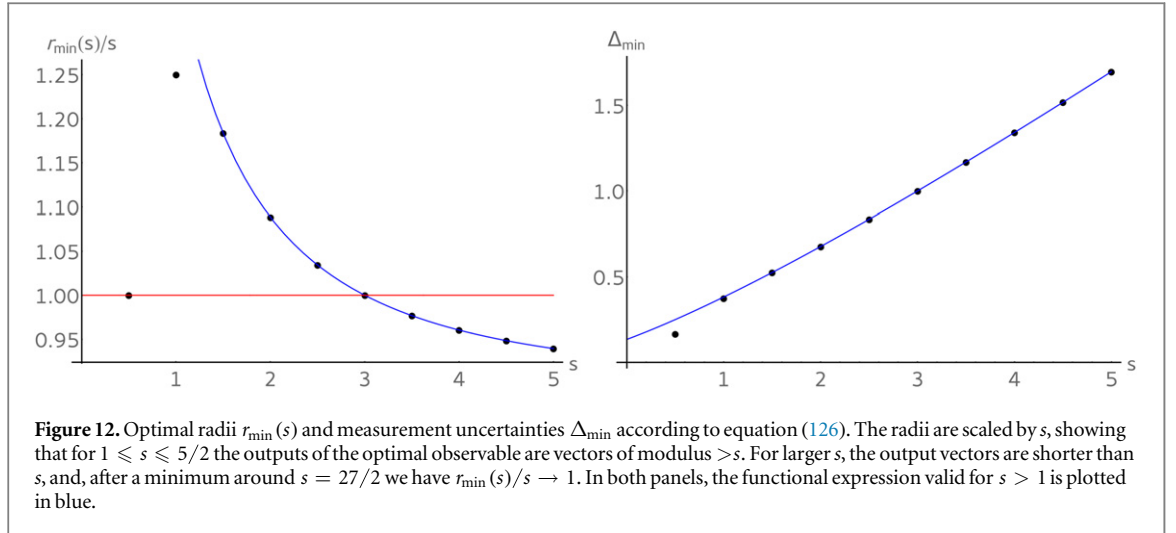
Then the expressions in the lemma arise by simplifying the corresponding polynomials:

$$\begin{aligned} I(s, r, n, m) &= \kappa(s, k, \mu, \nu) \left[r^2 (\alpha(k, \mu, \nu)^2 \omega(k, \mu, \nu) \right. \\ &\quad + \beta(k, \mu, \nu)^2 \omega(k-1, \mu, \nu) + \gamma(k, \mu, \nu)^2 \omega(k+1, \mu, \nu)) \\ &\quad \left. - 2rm \alpha(k, \mu, \nu) \omega(k, \mu, \nu) + m^2 \omega(k, \mu, \nu) \right] = A_s(r, n) + m^2 B_s(r, n). \end{aligned} \quad (125)$$

□

We will use this lemma to simplify the minimization over m . Moreover, the integral over r and the sum over n can be seen as taking a convex combination over two-dimensional vectors $(A_s(r, n), B_s(r, n))$. Hence optimizing F can be analyzed geometrically in terms of the set of such pairs (see figures 13). This is solved in the next theorem, whose results are visualized in figure 12.

Theorem 12. *The minimal measurement uncertainty $\Delta_{\min}(s)$ in the sense of (111) is attained at a unique covariant observable, for which $F_n(dr)$ is a point measure at $n = s$ and $r = r_{\min}$, with*



$$r_{\min}(s) = \begin{cases} 1/2 \\ 5/4 \\ (2s - \sqrt{2s+3} + 3)/2 \end{cases} \quad \text{and} \quad \Delta_{\min}(s) = \begin{cases} 1/6 & s = 1/2 \\ 3/8 & s = 1 \\ (s - \sqrt{2s+3} + 2)/2 & s > 1 \end{cases}. \quad (126)$$

Except for $s = 1$, the maximum over m in (113) is trivial for the optimal observable, i.e., the calibration error is the same for all calibration inputs m .

Proof. We consider first the case $s \geq 1$. For arbitrary $s \geq 1$ we reformulate the problem using that $\Delta(F^3, L_3)^2$ is a convex combination of the functions $A_s(r, n)$ and $B_s(r, n)$. Here we must find the best n as well as the probability distribution $F_n(dr)$ for the worst m . We denote the convex set of all possible combinations by $\Omega \in \mathbb{R}^2$, i.e.

$$\begin{aligned} \Omega &:= \left\{ (a, b) \mid a = \int F_n(dr) A_s(r, n) \text{ and } b = \int F_n(dr) B_s(r, n) \right\} \\ &= \text{conv} \left\{ (A_s(r, n), B_s(r, n)) \mid r > 0, n = -s, \dots, s \right\}. \end{aligned} \quad (127)$$

All information about a possible observable is now contained in $(a, b) \in \Omega$. Furthermore a maximum over m is part of the definition of $\Delta_c^2(F^3, L_3)^2$, so we can rewrite it as a functional on Ω :

$$K : \Omega \rightarrow \mathbb{R} \quad K(a, b) = \begin{cases} a + s^2 b & \text{if } b > 0 \\ a & \text{if } b \leq 0 \text{ and } s \in \mathbb{N} \\ a + \frac{1}{4} b & \text{if } b \leq 0 \text{ and } s + \frac{1}{2} \in \mathbb{N} \end{cases}. \quad (128)$$

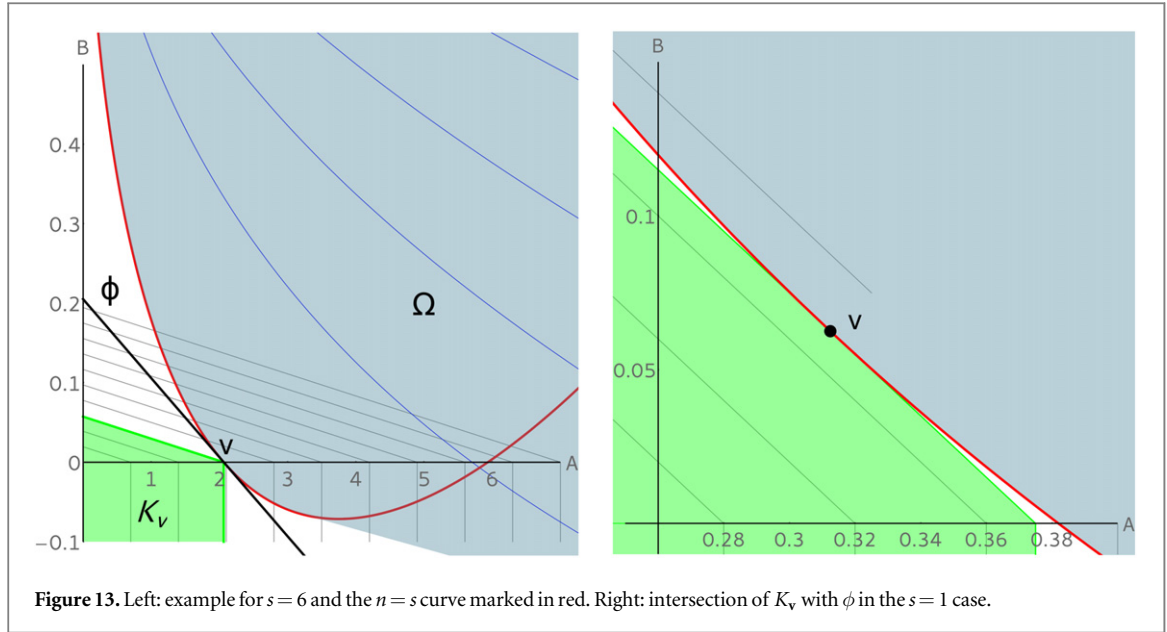
The problem is now to minimize the functional $K(a, b)$. Since, for general n and s , Ω is hard to describe, we choose the following strategy, which is illustrated in the left panel of figure 13.

We will show that, for $s > 1$, K takes its minimum at

$$\mathbf{v} = (A_s(r_{\min}(s), s), 0) = \left(\frac{1}{2}(s - \sqrt{2s+3} + 2), 0 \right) \quad (129)$$

by constructing a line ϕ which separates the set Ω from the convex level set $K_{\mathbf{v}} := \{(a, b) \in \mathbb{R}^2 \mid K(a, b) \leq K(\mathbf{v})\}$.

For this we will take the line ϕ to be the tangent of the curve $r \mapsto (A(r, s), B(r, s))$ at the point \mathbf{v} . The normal \mathbf{u} of ϕ is



$$\mathbf{u} = \left(\frac{2\sqrt{2s+3}}{2s^2+5s+3}, \frac{2s-\sqrt{2s+3}+3}{2s+3} \right) \quad (130)$$

and $\Phi := \{\mathbf{x} \in \mathbb{R}^2 | \mathbf{x} \cdot \mathbf{u} \geq \mathbf{v} \cdot \mathbf{u}\}$ is the half plane above ϕ .

Now we show that $\Omega \subset \Phi$, i.e.

$$g_s(n, r) := (A_s(r, n), B_s(r, n)) \cdot \mathbf{u} - (A_s(r_{\min}(s), s)) \cdot \mathbf{u} \geq 0, \quad (131)$$

with equality iff $n = s$ and $r = r_{\min}(s)$. Note that the function $g_s(n, r)$ is quadratic in r , so for verifying (131) it is sufficient to show that $\partial_r^2 g_s(n, r) \geq 0$ and $g_s(s, r_n) \geq 0$, where r_n is the stationary point determined by $\partial_r g_s(n, r_n) = 0$.

Indeed we have

$$\begin{aligned} \partial_r^2 g_s(n, r) &= 4s^2\sqrt{2s+3} - 4s^2 + 12n^2 - 4n^2\sqrt{2s+3} - 4s \\ &= 4((s^2 - n^2)(\sqrt{2s+3} - 3) + 2s^2 - s). \end{aligned} \quad (132)$$

Now for $s \geq 3$ the factor with the square root is positive, and since $|n| < s$ we can estimate the expression (132) as $\geq 2s^2 - s \geq 0$. For $1 \leq s \leq 3$, the minimum with respect to n is assumed at $n = 0$, for which (132) can be evaluated explicitly, and shown to be positive.

The stationary points r_n can be straightforwardly computed as

$$r_n = \frac{n(2s-1)(-2s+\sqrt{2s+3}-3)}{2(n^2(\sqrt{2s+3}-3)+s(-\sqrt{2s+3}s+s+1))} \quad (133)$$

with

$$g_s(n, r_n) = \frac{s^2 - n^2}{s(s\sqrt{2s+3} - s - 1) - n^2(\sqrt{2s+3} - 3)}. \quad (134)$$

The denominator of $g_s(n, r_n)$ in (134) is again the quadratic coefficient of the expression (131), i.e., (132) which was already shown to be positive. Hence $\Omega \subset \Phi$.

Finally we have to certify that $K_v \cap \Phi = \{\mathbf{v}\}$. Using the gradient of ϕ and comparing it to the linear boundaries of K_v we get the conditions

$$\begin{aligned} -4 < -\frac{1+\sqrt{2s+3}}{(1+s)^2} \quad \text{and} \quad \frac{1+\sqrt{2s+3}}{(1+s)^2} < -\frac{1}{s^2} \quad \text{for } s + \frac{1}{2} \in \mathbb{N} \\ -\infty < -\frac{1+\sqrt{2s+3}}{(1+s)^2} \quad \text{and} \quad -\frac{1+\sqrt{2s+3}}{(1+s)^2} < -\frac{1}{s^2} \quad \text{for } s \in \mathbb{N} \end{aligned} \quad (135)$$

which are true for $s > 1$. This concludes the proof for $s > 1$.

For $s = 1$, this last step of the above proof fails. Indeed, the level set K_v which was determined by taking that point v on the horizontal axis which is also on the boundary of Ω does intersect Ω , as can be seen in figure 13. We therefore have to take a level set of K for a slightly smaller value. Since the tangents of the level sets are all the same for $b > 0$, we can readily find the level set which is tangent to Ω . This gives the optimal radius $r_{\min}(s = 1) = 5/4$, and the

$$\mathbf{v} = \left(A_s(r_{\min}(s), s), B_s(r_{\min}(s), s) \right) = (5/16, 1/16). \quad (136)$$

Analogous to the above arguments, one verifies easily that $\{\mathbf{v}\} = K(\mathbf{v}) \cap \Omega$.

Finally, for $s = 1/2$ one can draw the conclusion directly from the form of $I(1/2, r, n, m)$ given in lemma 11: this expression does not depend on m , and has a unique global minimum at $r_{\min}(1/2) = 1/2$ and $n = +1/2$, where the optimal probability measure F must therefore be concentrated.

In all cases, the optimal value $\Delta_{\min}(s)$ is computed by substituting the obtained optimal $r_{\min}(s)$ and $n = s$ in (113). \square

5. Conclusions and outlook

Uncertainty relations can be built for any collection of observables. In this paper we provided some methods, which work in a general setting, but chiefly looked at angular momentum as one of the paradigmatic cases of non-commutativity in quantum mechanics.

The basic mathematical methods are well-developed for the case of preparation uncertainty, so that even in a general case the optimal tradeoff curves can be generated efficiently. We resorted to numerics quite often, since it turns out that the salient optimization problems can rarely be solved analytically for general s . One of the features one might hope to settle analytically in the future is the asymptotic estimate $c_2(s) \propto s^{2/3}$ which comes out with a precision that suggests an exact result.

Much is left to be done for entropic uncertainty. Here we gave only some basic comparisons to the variance case. It would be interesting to see whether the entropic relations can be refined to the point that they can be used to derive sharp variance inequalities as Hirschman did in the phase space case [9].

For measurement uncertainty the general situation is not so favourable, perhaps due to the much more recent introduction of the subject. At this point we know of no efficient way to derive sharp bounds for generic pairs of observables. Nevertheless, we were able to treat the case of a joint measurement of all components in arbitrary directions, because in this case rotational symmetry is not broken and leads to considerable simplification. One of these simplifications is the observation that the two basic error criteria, namely metric uncertainty and calibration error lead to the same results. This was already familiar from the phase space case. However, a further simplification one might have expected from this analogy definitely does not hold: there seems to be no quantitative link between preparation and measurement uncertainty for angular momentum. Further research will show whether useful general connections between the two faces of the uncertainty coin can be established.

The limit large $s \rightarrow \infty$ can be understood as a mean field limit [23], when the spin- s representation is considered as $2s$ copies of a spin- $1/2$ system in a symmetric state. We can also see this as a classical limit $\hbar \rightarrow 0$ [33], in the sense that the angular momentum in physical units, i.e., \hbar is fixed, and hence the dimensionless half-integral representation parameter s has to diverge. This offers a way to treat not just the uncertainty aspects of this limit, but also the limit of the whole theory of angular momentum.

Acknowledgments

The authors thank Ashley Milsted, Ciara Morgan and David Reeb for critically reading our manuscript. We also acknowledge the financial support from the ERC grant DQSIM, RTG1991 funded by the DFG and the collaborative research project Q.com-Q funded by the BMBF.

We acknowledge support by Deutsche Forschungsgemeinschaft and the Open Access Publishing Fund of Leibniz Universität Hannover.

References

- [1] Biedenharn L and Louck J 1981 *Angular Momentum in Quantum Physics* (Reading, MA: Addison-Wesley)
- [2] Busch P, Lahti P and Werner R F 2013 Proof of Heisenberg's error-disturbance relation *Phys. Rev. Lett.* **111** 160405
- [3] Busch P, Lahti P and Werner R F 2014 Quantum root-mean square error and measurement uncertainty relations *Rev. Mod. Phys.* **86** 1261–81

- [4] Busch P, Lahti P and Werner R F 2014 Heisenberg uncertainty for qubit measurements *Phys. Rev. A* **89** 012129
- [5] Busch P, Lahti P and Werner R F 2014 Measurement uncertainty relations *J. Math. Phys.* **55** 042111
- [6] Cunha M T, Man'ko V and Scully M 2001 Quasiprobability and probability distributions for spin-1/2 states *Found. Phys. Lett.* **14** 103–17
- [7] Franke-Arnold S, Barnett S M, Yao E, Leach J, Courtial J and Padgett M 2004 Uncertainty principle for angular position and angular momentum *New J. Phys.* **6** 103
- [8] Heisenberg W 1927 Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik *Z. Phys.* **43** 172–98
- [9] Hirschman I I Jr 1957 A note on entropy *Am. J. Math.* **79** 152–6
- [10] Hofmann H F and Takeuchi S 2003 Violation of local uncertainty relations as a signature of entanglement *Phys. Rev. A* **68** 032103
- [11] Holstein T and Primakoff H 1940 Field dependence of the intrinsic domain magnetization of a ferromagnet *Phys. Rev.* **58** 1098–113
- [12] Hörmander L 2003 *The Analysis of Linear Partial Differential Operators I* (Berlin: Springer)
- [13] Horne R A and Johnson C R 1985 *Matrix Analysis* (Cambridge: Cambridge University Press)
- [14] Huang Y 2012 Variance-based uncertainty relations *Phys. Rev. A* **86** 024101
- [15] Hudson R 1974 When is the wigner quasi-probability density non-negative? *Rep. Math. Phys.* **6** 249–52
- [16] Kechrimparis S and Weigert S 2014 A Heisenberg uncertainty relation for three canonical observables *Phys. Rev. A* **90** 062118
- [17] Kennard E 1927 Zur Quantenmechanik einfacher Bewegungstypen *Z. Phys.* **44** 326–52
- [18] Kiukas J, Ruschhaupt A, Schmidt P O and Werner R F 2012 Exact energy–time uncertainty relation for arrival time by absorption *J. Phys. A: Math. Theor.* **45** 185301
- [19] Maassen H and Uffink J B M 1988 Generalized entropic uncertainty relations *Phys. Rev. Lett.* **60** 1103–6
- [20] Maccone L and Pati A K 2014 Stronger uncertainty relations for all incompatible observables *Phys. Rev. Lett.* **113** 260401
- [21] Naimark M A 1943 On a representation of additive operator set functions *C.R. (Doklady) Acad. Sci. URSS (N.S.)* **41** 359–61
- [22] Ozawa M 2004 Uncertainty relations for joint measurements of noncommuting observables *Phys. Lett. A* **320** 367–74
- [23] Raggio G A and Werner R F 1989 Quantum statistical mechanics of general mean field systems *Helv. Phys. Acta.* **62** 980–1003
- [24] Rivas A and Luis A 2008 Characterization of quantum angular-momentum fluctuations via principal components *Phys. Rev. A* **77** 022105
- [25] Robertson H 1929 The uncertainty principle *Phys. Rev.* **34** 163–4
- [26] Scholz V B and Renes J M 2014 Operationally-motivated uncertainty relations for joint measurability and the error-disturbance tradeoff (arXiv:1402.6711)
- [27] Schrödinger E 1930 Zum Heisenbergschen Unschärfeprinzip *Sitzungsberichte der Preußischen Akademie der Wissenschaften. Physikalisch-mathematische Klasse* 296–303
- [28] Scutaru H 1979 Some remarks on covariant completely positive linear maps on C^* -algebras *Rep. Math. Phys.* **16** 79–87
- [29] Sánchez-Ruiz J 1993 Maassen–Uffink entropic uncertainty relation for angular momentum observables *Phys. Lett. A* **181** 193–8
- [30] Szegő G 1975 *Orthogonal Polynomials* (Providence, RI: American Mathematical Society)
- [31] Villani C 2009 *Optimal Transport* (Berlin: Springer)
- [32] Werner R F 1986 Screen observables in relativistic and nonrelativistic quantum mechanics *J. Math. Phys.* **27** 793
- [33] Werner R F and Wolff M P H 1995 Classical mechanics as quantum mechanics with infinitesimal \hbar *Phys. Lett. A* **202** 155–9
- [34] Wigner E 1963 Interview with Thomas S Kuhn (www.aip.org/history-programs/neils-bohr-library/oral-histories/4963-3)