Properties of the Wigner distribution for $n$ arbitrary operators

René Schwonnek$^{1,*}$ and Reinhard F. Werner$^{1,†}$

$^1$Institut für Theoretische Physik, Leibniz Universität, Hannover, Germany

(Dated: February 26, 2018)

We study a generalization of the Wigner function to arbitrary tuples of hermitian operators, which is a distribution uniquely characterized by the property that the marginals for all linear combinations of the given operators agree with the quantum mechanical distributions. Its role as a joint quasi-probability distribution is underlined by the property that its support always lies in the set of expectation value tuples of the operators. We characterize the set of singularities and positivity, and provide some basic examples.

I. INTRODUCTION

It is well known that non-commuting hermitian operators do not admit a joint measurement, i.e., their spectral measures cannot be written as the marginals of some other positive operator valued measure. That is, there is no way to linearly assign to each quantum state (density operator) a positive joint distribution, whose marginals are the standard probability distributions of the given operators. The positivity in this statement is crucial, and Wigner’s quasiprobability function [1] is the paradigm of a non-positive operator valued measure, whose marginals are the momentum and position observables, respectively. One can hope to learn more about the properly quantum, non-commutative structures by looking at just how much negativity is required. Indeed, the non-negativity of the Wigner function is seen by many as the hallmark of quantum effects in quantum optics. When we turn to more general pairs or tuples of operators, very little is known. The first obstacle to such an investigation is the non-uniqueness. Even in the position-momentum case a whole range of quasi-probability functions, and even gives a unique result for arbitrary operator tuples. This basic idea amounts to Weyl quantisation, and has been suggested in work of Cohen and Scully [3, 4]. However, it apparently has never been worked out in any detail. The aim of this paper is to provide some examples and to begin building an intuition concerning these general Wigner distributions. One prominent feature in the case of finite dimensional matrices, and in all cases where the original observables have discrete spectrum, are the singularities. We will provide a semialgebraic characterization of the singular set, and show how it behaves for nearly commuting operators. Moreover, we establish some basic properties concerning support, positivity and symmetries. An overview with more examples is to be found in [5]. The special case of the three angular momentum operators in an irreducible spin representation, and its classical limit will be analyzed in depth elsewhere.

II. DEFINITION

Let $A_1, \ldots, A_n$ be bounded hermitian operators on a Hilbert space. We typically take the Hilbert space dimension $d = \dim \mathcal{H}$ to be finite, as well. Linear combinations with real coefficients $\xi = (\xi_1, \ldots, \xi_n)$ are written as $\xi \cdot A = \sum_k \xi_k A_k$. Such an operator tuple will be assumed to be fixed throughout the paper, as well as a density operator $\rho$.

A. by the marginal property

The Wigner distribution is a distribution (generalized function) satisfying the equation

$$\int d^n a \ W_{\rho}(a_1, \ldots, a_n) \ f(a) = \text{tr} \ \rho f(\xi \cdot A), \tag{1}$$

for every bounded infinitely differentiable function $f: \mathbb{R} \to \mathbb{C}$. If $W_{\rho}$ were a probability density this would just equate the classical expectation of some function of the random variable $\xi \cdot a$ with the quantum expectation of the same function, applied to the operator $\xi \cdot A$. However, $W_{\rho}$ is usually not a probability density, not just because it may be negative, but because it may be a generalized function/distribution. From this definition it is not immediately clear that such a distribution exists and is even unique. However, this will be clear from the next one.

B. by Fourier transform

When we take $f(t) = \exp(it)$ in (1) we directly get an expression for the Fourier transform $\hat{W}_{\rho}$ of $W_{\rho}$, namely

$$\hat{W}_{\rho}(\xi) = \text{tr} \ \rho e^{i \xi \cdot A}. \tag{2}$$

---

* rene.schwonnek@itp.uni-hannover.de
† reinhard.werner@itp.uni-hannover.de
This fixes $\hat{W}_\rho$ uniquely as a distribution, which we consider as a tempered distribution (i.e., with the Schwartz functions for a test function space). The reverse Fourier transform is then also a tempered distribution, which we denote by $W_\rho$. We will establish below (Property 1) that this distribution has compact support, which makes it possible to integrate it with differentiable, but unbounded “test functions”.

C. by ordered moments

Let $r = (r_1, \ldots, r_n) \in \mathbb{N}^n$, and $R = \sum_{k=1}^n r_k$. The we define the Weyl-ordered moment of order $r$ as

$$M^r_W(A) := \frac{1}{R!} \sum_{\pi} A_{i_1} A_{i_2} \cdots A_{i_R}$$

where $\pi$ runs over all permutations of $\{1, \ldots, R\}$, where the labels $i_1, \ldots, i_R \in \{1, \ldots, n\}$ are chosen so that when $\pi$ is the identity the product is simply $A_1 \cdot A_2 \cdots A_R$. That is, each product contains $r_1$ factors equal to $A_1$, etc, and $M^r_W$ is the average over all such products. This form of the definition follows a pattern that can be used also for other for ordered moments $[6, 7]$, like the “normal” or “Wick” ordering where the average contains only the term with $\pi$ the identity, or the “antinormal” with only the reversing permutation $\pi k = R - k + 1$. All these moments coincide with the ordinary ones when the $A_k$ commute. The importance of the Weyl ordering lies in the non-commutative multinomial theorem: For any coefficients $\xi_1, \xi_2, \ldots, \xi_n$ one has

$$\sum_{r} \binom{R}{r} \xi_1^{r_1} \cdots \xi_n^{r_n} M^r_W(A) = (\xi \cdot A)^R.$$  

Now consider any polynomial $f(a_1, \ldots, a_n)$, expand, and replace the monomial $a_1^{r_1} a_2^{r_2} \cdots a_n^{r_n}$ by $M^r_W(A)$. We denote the resulting operator by $W^*_f$, the Weyl-ordered operator version of $f(A)$. Then

$$\int d^n \rho \mathcal{W}_\rho(a) f(a) = \text{tr} \rho W^*_f.$$  

(D. Graphical representation)

A distribution like the derivative of the $\delta$-function cannot be plotted. However, visualization is still possible if we take into account that any graphical representation of a function requires a choice of resolution. So at each point we plot instead the average of the distribution over a Gaussian of some width $\varepsilon$, i.e., we take the convolution with such a Gaussian. Equivalently, we multiply $\hat{W}_\rho(\xi)$ by the Fourier transform of the Gaussian, which is $\exp(-\varepsilon \xi^2)$. The product is then a well-decaying integrable function, and we can take its Fourier transform with standard numerical methods (Fast Fourier transform). This is how all the plots of Wigner functions in this paper were generated. We usually tuned the regularization parameter $\varepsilon$ for best visibility. For too large $\varepsilon$ one gets a featureless bump approximately equal to the regularizing Gaussian. For too small $\varepsilon$ one sees only the singularities in the form of infinitely high walls. Fig. 1 shows the dependence on $\varepsilon$ for the example of two random matrices in $d = 4$. 

FIG. 1. Dependence on the regularization parameter $\varepsilon$. Depicted for $\varepsilon = 1, 10^{-1}, 10^{-2}, \cdots, 10^{-5}$.
III. PROPERTIES

A. Support

**Property 1.** $W_\rho$ has support in a compact convex set, the joint numerical range $\mathcal{R}$ of the given operators:

$$\mathcal{R} = \{ a \in \mathbb{R}^n \mid a_k = \text{tr} \, \rho A_k \} \tag{6}$$

where $\rho$ runs over all density operators.

The support of a distribution is defined as the smallest closed set $\mathcal{R}$ such that for any test function with support disjoint from $\mathcal{R}$ the integral vanishes. The definition of joint numerical range used here includes mixed states, so it is automatically compact and convex as the continuous linear image of the state space. In the mathematical literature a definition of joint numerical range based on the pure states only is often used. This is convex for $n = 2$ (The Hausdorff-Toeplitz Theorem), and for $n = 3$ (except in the qubit case, where the Bloch sphere, in contrast to the full ball, is not convex). Convexity of the pure state joint numerical range fails for $n \geq 4$, but its convex hull is always the $\mathcal{R}$ defined above.

The core of the proof of Property 1 is the Paley-Wiener-Schwartz Theorem [8, Thm. 7.3.1], i.e., the distributional version of the Paley-Wiener-Schwartz Theorem, which links support properties of a function to analyticity properties of the Fourier transform. To get a heuristic idea, consider a function $f$ with support in the half space $\{ a | \eta \cdot a \geq c(\eta) \}$, given by the parameters $\eta$ and $c(\eta)$, where $c(\eta)$ is by definition homogeneous ($c(t\eta) = tc(\eta)$ for $t > 0$). Then the function $a \mapsto \exp(-\eta \cdot a)$ is bounded on the support of $f$, so the Fourier integral for $\exp(-\eta \cdot a)\hat{f}(a)$ converges better than the Fourier integral for $f$, and gives an analytic extension $\hat{f}(\xi + 2\pi i)$ satisfying a bound with a factor $\exp(-c(\eta))$. The Paley-Wiener-Schwartz Theorem is a precise statement of the converse.

**Proof.** In order to control the growth of the analytic extension $\hat{W}_\rho(\xi + 2\pi i)$ in the $\eta$-direction uniformly in $\rho$ we will need to control the norm of the operator $\exp(i(\xi + 2\pi i)\cdot A)$. Suppose we have the operator inequality

$$\eta \cdot A \geq c(\eta) \mathbb{1} \tag{7}$$

for some constant $c(\eta)$, which is equivalent to saying that

$$\mathcal{R} \subset \{ a | \eta \cdot a \geq c(\eta) \}. \tag{8}$$

Then $\| \exp(-\eta \cdot A) \| \leq \exp(-c(\eta))$ and, by the Trotter product formula

$$\left\| e^{i\xi \cdot A - \eta \cdot A} \right\| \leq \lim_{n \to \infty} \left\| \left( e^{i\frac{\xi \cdot A}{n}} e^{-\eta \cdot A/n} \right)^n \right\| \leq e^{-c(\eta)}.$$

Hence

$$\left| \hat{W}_\rho(\xi + 2\pi i) \right| \leq e^{-c(\eta)}, \tag{9}$$

which means that the support of $W_\rho$ is contained in the half space on the right hand side of (8). Since $\mathcal{R}$, as a closed convex set, is the intersection of all half spaces containing it, the support must be contained in $\mathcal{R}$. Actually, the last step is already included in the version of the Paley-Wiener-Schwartz Theorem cited from H"ormander’s book, which is directly based on the function $c(\eta)$, called the ‘supporting function’ of $\mathcal{R}$.

B. Positivity

**Property 2.** When the $A_k$ are finite dimension matrices, and $\rho$ has full rank, then $W_\rho$ is positive if and only if the $A_k$ commute, in which case $W_\rho$ is a sum of $\delta$-functions with $\rho$-dependent weights.

**Proof.** Suppose that $W_\rho \geq 0$. Then it must actually be given by a positive measure, even a probability measure $P$ due to the overall normalization. Now fix $\xi$ and let $\alpha_i, \alpha_{i+1}$ be two neighbouring eigenvalues of $\xi \cdot A$. Then, by the marginal property, the slice of $\mathbb{R}^n$

$$S_i = \{ a | \alpha_i < \xi \cdot a < \alpha_{i+1} \} \tag{10}$$

has $P$-measure 0. In other words, $P$ is supported by the hyperplanes $\xi \cdot a = \alpha_i$. This is true for any direction $\xi$, and taking the intersection over some linearly independent set of directions $\xi$, we find that $P$ is supported on finitely many points. We can then find a direction $\xi$ so that each plane $\xi \cdot a = \text{const}$ contains at most one point. Since $\xi \cdot A$ has at most $d = \dim \mathcal{H}$ eigenvalues, we find that there are at most $d$ points in the support. Suppose we find this maximal number. Then the corresponding eigenvectors form a basis, and in this basis all $A_k$ are diagonal, so the $A_k$ commute.

Otherwise, we can make this conclusion only for the span of the eigenvectors of $\rho$, and we can draw no conclusion about the commutation of the $A_k$ about the complement of the support of $\rho$. This is why we need the condition that $\rho$ has full rank. The $\rho$-dependent weights are computed as the expectations of $\rho$ in the joint eigen-projections (see previous item).

It is well-known that Gaussian states do have a positive $(P,Q)$-Wigner functions. By one of the most beautiful results of the theory [9], these are the only pure states with that property. For mixed states no interesting characterization is known [10]. On averaging over phases space translations (with a sufficiently spread out weight) every state will have positive Wigner function. This is in stark contrast to the above result. However, the proof shows why there is such a great difference: For $A_k$ with discrete spectrum the marginal condition for a positive measure is extremely strong, forcing the measure to vanish on almost all of $\mathbb{R}^n$. 
C. Singular support

The singularities of $W_{\rho}$ are its most prominent features. In this section we will locate the singular support, i.e., the set on whose complement the distribution is a smooth function [11]. We will sketch a proof for the finite dimensional case. A rigorous theory for the general case would be highly desirable, especially for infinite dimensional (and possibly unbounded) $A_k$ including absolutely continuous and singular continuous spectrum.

**Property 3.** For finite dimensional matrices $A_k$, the singularities of $W_{\rho}$ lie on the closure of the set

$$S = \{ a \in \mathbb{R}^n \mid a_k = \langle \psi | A_k | \psi \rangle; \quad \| \psi \| = 1, \xi \cdot A \psi = \lambda \psi \}, \quad (11)$$

$\psi$ runs over all eigenvectors of non-degenerate eigenvalue problems for some $\xi \cdot A$.

To get a heuristic idea, let us first diagonalize each operator $\xi \cdot A$.

$$\xi \cdot A = \sum_\mu \alpha_\mu(\xi) P_\mu(\xi), \quad (12)$$

where $P_\mu(\xi)$, $\mu = 1, \ldots, n$, denotes the eigenprojections and $\alpha_\mu(\xi)$ the corresponding eigenvalues. Then the inverse Fourier integral for $W_{\rho}(a)$ contains, apart from some slowly varying factors the oscillatory phase

$$\exp i(-\xi \cdot a + \alpha_\mu(\xi))$$

The stationary phase method for such integrals has it that the dominant contribution to the integral comes from those points at which the gradient of the exponent with respect to the integration variables $\xi$ vanishes, i.e. $a_k = \partial \alpha_\mu(\xi)/\partial \xi_k$. By first order perturbation theory the right hand side is just $\langle \psi | A_k | \psi \rangle$ for $\psi$ the corresponding normalized eigenvector, i.e., $a \in S$.

*Sketch of proof for Property 3.* The diagonalization (12) has to be done only once for each collection of proportional vectors, since, for positive $\lambda$, we have $\alpha_\mu(\lambda \xi) = \lambda \alpha_\mu(\xi)$ and $P_{\mu}(\lambda \xi) = P_\mu(\xi)$. This suggests to split the variable $\xi = tu$, into a single radius $t \geq 0$ and an angular part, i.e., a unit vector $u \in \mathbb{R}^n$.

Then the $t$-regularized Wigner function is given by

$$\widehat{W}_{\rho,\varepsilon}(tu) = \sum_\mu e^{it\alpha_\mu(u) - \varepsilon t^2} tr(P_\mu(u))$$

$$W_{\rho,\varepsilon}(a) = \sum_\mu \int dt \int \frac{d^n}{d^n t} e^{it(a_\mu(u) - a \cdot u) - \varepsilon t^2} tr(P_\mu(u))$$

$$= \sum_\mu \int du \ h_\varepsilon(\alpha_\mu(u) - a \cdot u) tr(P_\mu(u)), \quad (14)$$

$du$ indicates suitably normalized integration with respect to the surface measure of the unit sphere, and $h_\varepsilon$ is the function given by the $t$-integral in the line above. This is the $n - 1$ fold derivative of the corresponding function with $n = 1$, which is a Gaussian regularized $\delta$-function. Hence, as $\varepsilon \to 0$ the $u$-integration becomes concentrated on the set

$$M(a) = \{ u | \alpha_\mu(u) - a \cdot u \}, \quad (15)$$

which is the intersection of the unit sphere with the manifold with $\alpha_\mu(\xi) = a \cdot \xi$. Suppose that, for all $u$, $\nabla \alpha_\mu(u) \neq a$. Then by the implicit function theorem $M(a)$ is a regular submanifold of the unit sphere. In a neighbourhood of any non-degenerate point $P_\mu(u)$ is an analytic function. In the limit the $n - 1$ fold normal derivative of this function is integrated over the submanifold $M(a)$, which is a regular integral. Hence the singular support is contained in

$$\bigcup_\mu \{ \nabla \alpha_\mu(u) | |u| = 1 \}. \quad (16)$$

[A more careful treatment uses the theory of distributions given by oscillatory integrals [11, Thm. IX.47], and the equality of the singular support of such a distribution with its wave front set and in turn with the stationary phase points of the phase function $\xi \mapsto \alpha_\mu(\xi) - \xi \cdot a$.

By first order perturbation theory, we can evaluate $\nabla \alpha_\mu(\xi)$ as the expectation tuple of the $A_i$ in the normalized eigenvector of $\xi \cdot A$ with eigenvalue $\alpha_\mu$, which is the formula given in Property 3.]

D. Algebraic nature of the singular support

The singularity set $S$ is defined completely in terms of the eigenvalue problems of $\xi \cdot A$, quite independently of the Wigner distribution. It is therefore not surprising that it has been studied in various papers. A set is called *algebraic*, if it is the solution set a system of polynomial equations and inequalities. Then we have:

**Property 4.** The singular support $S$ is a semi-algebraic set. For $n = 2$ it is even algebraic, but not necessarily for $n \geq 3$.

Clearly, the information about the eigenvalues of the family of operators $\xi \cdot A$ (sometimes called an “operator pencil”) is also contained in the algebraic variety $V$ consisting of the zeros of the homogenous polynomial $\xi_0, \ldots, \xi_n \mapsto \det(\xi_0 I + \xi \cdot A)$. This is a subset in the space of $\xi$-variables, whereas $S$ is a set of tuples of $a$-variables, so lives in the dual vector space. It is therefore natural to consider the dual variety $V^*$, which is defined in complex algebraic geometry [12] as the closure of the set of normal vectors at smooth points. To see what this gives in the case at hand, let us just look at the points where the rank of $\xi_0 I + \xi \cdot A$ is reduced by 1, i.e., $\xi_0 I + \xi \cdot A$ has a one dimensional kernel. Then there is a unique eigenvector $\psi$ of $\xi \cdot A$ with eigenvalue $-\xi_0$. To get
the tangent plane at this point, we check when a curve, say \((\xi_0(t), \xi(t)) = (\xi_0, \xi) + t(\eta_0, \eta) + o(t)\) remains in \(V\) to first order. By non-degenerate first order perturbation theory, the eigenvalue condition remains satisfied to first order, provided that \(-\eta_0 = (\psi^*|\eta \cdot A|\psi) = \eta \cdot a\), where \(a\) is the expectation vector of \(\psi\). That is, the allowed tangent vectors \((\eta_0, \eta)\) are precisely those orthogonal to \((1, a)\), which defines the normal vector up to a factor. The factor is here chosen to make the first component 1, which defines the “affine part” of the variety \(V^*\). So \(S\) is identified as the affine part of the dual variety.

However, in this description we have implicitly taken all variables to be real. This may be quite different from starting with complex \(\xi\), and only restricting to real \(a\) at the end. The basic algebraic construction for getting \(V^*\) is to eliminate the \(\xi\)-variables from the equations

\[
\begin{align*}
\frac{\partial}{\partial \xi_0} \det(\xi_0 \mathbf{I} + \xi \cdot A) &= 0, \\
\frac{\partial}{\partial \xi_k} \det(\xi_0 \mathbf{I} + \xi \cdot A) &= a_k,
\end{align*}
\]

and take the closure of the resulting set of \(a\)-tuples. In the real version of this construction, this constitutes an obviously algebraic set of tuples \((\xi_0, \xi, a)\), which is to be projected to just the \(a\) variables. This operation takes semi-algebraic sets to semi-algebraic sets, as does the closure, which shows that indeed \(S\) is semi-algebraic. On the other hand, consider a purely algebraic elimination process, which will be the same in the complex and in the real case, and hence implicitly includes cases of complex \(\xi\) in (17). This leads to the dual variety \(V^*\), which is again an algebraic variety. Its algebraic definition as the zero set of a polynomial automatically describes a closed set, but its intersection with the reals may now contain additional points. In such a case a purely algebraic description of \(S\) fails, and we need some inequalities in addition to polynomial equations to characterize it.

It was shown by Krippenhahn [13] that for \(n = 2\) this phenomenon does not happen, so that the algebraic construction correctly describes \(S\). A more modern and more detailed version is provided by [14], see also [15]. In [14] there is also a counterexample for \(n = 3\). We include it here, pointing out more explicitly how complex \(\xi\) in (17) enter to get points outside \(S\). The operators are

\[
A_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad A_3 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]

Then the eigenvalues \(-\xi_0\) are the zeros of

\[
f(\xi_0, \xi) = \det(\xi_0 \mathbf{I} + \xi \cdot A) = \xi_0^3 + 2\xi_0^2 \xi_3 - \xi_0^2 \xi_2 - \xi_1^2 \xi_3 + \xi_1 \xi_2,
\]

The algebraic elimination of the \(\xi\)-variables from (17) leads to the irreducible polynomial

\[
g(a) = 4\xi_0^2 (a_1^2 - 2a_1a_3 + 5a_2^2 + 2a_1 - 6a_3 + 1) + 4a_5 (2a_5 - a_1 - 1)a_2^2 + a_2^4.
\]

The affine part of the dual complex variety is thus described by \(g(a) = 0\). This is the best purely algebraic description of \(S\) that we can get. Indeed from any small piece of the surface we can reconstruct this polynomial. The resulting surface \(S\) is shown in Fig. 2. But it has a new feature namely the infinite line \((a_1, 0, 0)\) with arbitrary \(a_1 \in \mathbb{R}\). Since this is an unbounded line, it clearly does not belong to the closure of \(S\). One can verify that it is the only addition to \(S\) in the zero set of \(g\) (see Fig. 3 for a sketch).

How do we get that strange line now? The only points on it, which can arise as directly as solutions of the system (17) are \(a_0 = \pm 1\). The other points must arise from a closure operation. Consider the parameterized set of
points on $V$

\[
\begin{pmatrix}
\xi_0 \\
\xi_1 \\
\xi_2 \\
\xi_3
\end{pmatrix} = \begin{pmatrix}
y - 1 \\
y \\
x^2(2y - 1) \\
x^2(2y - 1) + 1 - y^2
\end{pmatrix},
\]

(20)

corresponding via (17) to

\[
\begin{pmatrix}
a_1 \\
a_2 \\
a_3
\end{pmatrix} = \frac{1}{1 + x^2 + y^2} \begin{pmatrix}
x^2 - y^2 + 2y \\
-2x \\
1
\end{pmatrix}
\rightarrow \frac{1}{x^2 + y^2} \begin{pmatrix}
x^2 - y^2 \\
-2x \\
0
\end{pmatrix},
\]

(21)

where the arrow indicates the scaling limit $x \rightarrow \lambda x$, $y \rightarrow \lambda y$, and $\lambda \rightarrow \infty$. When $x, y \in \mathbb{R}$ this shows that $a = (a_1, 0, 0) \in \mathcal{S}$, when $a_1 \in [-1, 1]$. However, if we allow complex parameters, say $y \rightarrow i\lambda y$, we get in addition the inverses of the same $a_1$, i.e., the rest of the real line. Hence the purely algebraic description of $\mathcal{S}$ fails, and inequalities are needed in addition.

Another characterization of $\mathcal{S}$ is as the set of expectation tuples $\langle \psi | A | \psi \rangle$ for which the map $\psi \mapsto \langle \psi | A | \psi \rangle \in \mathbb{R}^{n}$ has reduced rank $\leq (n - 1)$. For $n = 2$ this investigated in [16].

## E. Reducibility

We call a tuple $A_1, \ldots, A_n$ irreducible, if the only linear operators commuting with all $A_k$ are the multiples of $\mathbb{1}$. This is equivalent to saying that the weakly closed algebra of operators generated by the $A_k$ is the algebra $\mathcal{B}(\mathcal{H})$ of all bounded operators. In the finite dimensional case one can omit the closure here. In the reducible case we have a joint diagonal block decomposition of all $A_k$, and the Wigner function is insensitive to the off-diagonal matrix elements of $\rho$ with respect to that decomposition:

**Property 5.** Suppose all $A_k$ commute with the a family of orthogonal projections $Q_\ell$ with $\sum_\ell Q_\ell = \mathbb{1}$. Then

\[
\widehat{W}_\rho(\xi) = \sum_\ell \widehat{W}_{Q_\ell \rho Q_\ell}(\xi).
\]

Note that in this case the support of $W_\rho$ is contained in the union of the corresponding joint numerical ranges $\mathcal{R}_\ell$, which may be much smaller than $\mathcal{R}$, their convex hull. In other words, Property 1 only gives an upper bound on the support.

Reducibility in the above sense also implies that the determinant polynomial $\text{det}(\xi_0 \mathbb{1} + \xi_j A)$ factorizes, i.e., the variety $V$ is "reducible". Kippenhahn conjectured this to be the only case. However, that is only true in small $(d \leq 5)$ matrix dimension. There is a counterexample [17] in $d = 8$ dimensions with $n = 2$.

Rather than commutation with a linear operator we can consider an antilinear operator. This also leads to a lack of informational completeness, i.e., $\rho$ cannot be reconstructed from $W_\rho$. In order to avoid trivial reductions of a kind we already know the following Proposition assumes irreducibility.

**Proposition 1.** Let $A_1, \ldots, A_n$ be an irreducible set of hermitian $d \times d$-matrices, all commuting with some non-zero antilinear operator $\Theta$. Then all Weyl-ordered moments also lie in $\mathcal{M}_\Theta = \{ X | X^\ast = X, \Theta X = X \Theta \}$. Moreover, $\Theta^2 = \lambda \mathbb{1}$, and there are two possibilities:

1. $\lambda > 0$, and $\dim_{\mathbb{R}}(\mathcal{M}_\Theta) = (d+1)/2$, and
2. $\lambda < 0$, and $\dim_{\mathbb{R}}(\mathcal{M}_\Theta) = (d-1)/2$

**Proof.** The statement that $\mathcal{M}_W \subset \mathcal{M}_\Theta$ is trivial.

The linear operator $\Theta^\ast \Theta$ also commutes with all $A_k$, hence must be a multiple of the identity. We have excluded $\Theta = 0$, so after multiplication by a positive normalization factor $\Theta$ becomes unitary. Since $\Theta^2$ also commutes with the $A_k$ it must be $\lambda \mathbb{1}$, with $|\lambda| = 1$ by antiunitarity, and $\lambda \in \mathbb{R}$ because $(\mathbb{1}, \Theta)(\Theta^\ast) = \Theta^3 = \Theta(\mathbb{1})$. Hence $\lambda = \pm 1$.

When $\Theta^2 = \pm \mathbb{1}$, we can choose an orthonormal basis of the real Hilbert space $\{ \phi | \Theta \phi = \phi \}$, which will then also be a basis for $\mathcal{H} \subset \mathbb{C}^d$. In this basis $\Theta$ is just complex conjugation, and $X \in \mathcal{M}_\Theta$ is equivalent to $X$ being a real and symmetric matrix in that basis. This is uniquely specified by $d$ real diagonal elements, and $(d-1)/2$ off-diagonal matrix elements, altogether $(d+1)/2$ parameters.

When $\Theta^2 = -\mathbb{1}$, each $\psi$ mapped to an orthogonal vector $|\psi \pm \Theta \psi\rangle$, because $\langle \psi | \Theta | \psi \rangle = \langle \Theta^2 | \psi | \Theta | \psi \rangle = -\langle \psi | \Theta | \psi \rangle$. This implies that $d$ is even, and we can build a basis by including with each vector $\psi_j$ the vector $\psi_{d/2+j} := \Theta \psi_j$. With respect to such a basis we can block decompose $\Theta$ and an arbitrary element $X \in \mathcal{M}_\Theta$ as

\[
\Theta = \begin{pmatrix}
0 & K \\
-K & 0
\end{pmatrix}, \quad X = \begin{pmatrix}
A & B \\
C & D
\end{pmatrix},
\]

(23)

where $K$ stands for complex conjugation. Then $X \in \mathcal{M}_\Theta$ translates to $A = A^\ast = KDK$, and $C = B^\ast = -KBK$. Thus we have to choose an arbitrary hermitian $(d/2) \times (d/2)$-matrix $A$ (requiring $(d/2)^2$ real parameters), and an antisymmetric one, $B$, with $2(d/2)(d/2 - 1)/2$ real parameters, together $d^2/2 - d/2$. \hfill \square

Thus in either case, the number of parameters is much smaller than $d^2$, which would be required for informational completeness (see Sect. IIIH).

## F. Strict convexity

**Property 6.** For $n \leq 3$ the boundary of $\mathcal{R}$ is generically strictly convex and belongs to $\mathcal{S}$.

Thus the typical appearance of a Wigner function for two operators is that of a walled-in convex city rising
above a plain. The city walls come from the singularity, and can be made arbitrarily high by choosing a small regularization parameter $\varepsilon$. Together with the walls there is always a moat. More formally, the singularity cannot be positive only, because that would clash with the marginal property. The moat is on the inside of the wall. It is not entirely clear how to define this as a property of a distribution. But this property is exhibited by all our graphs. For a heuristic argument consider a supporting hyperplane
\[
\{ a \mid \xi \cdot a = m(\xi) \}, \tag{24}
\]
where $m(\xi) = \max\{ \xi \cdot x \mid x \in \mathcal{R}\}$. Then the integral of the regularized Wigner function over the half space $\xi \cdot a \geq 0$ can be computed from the marginal property, and will be positive if the regularizing function is positive.

The crucial property implying strict convexity is that for every $\xi$ the maximal eigenvalue of $\xi \cdot A$ is non-degenerate. Then the supporting hyperplane (24) meets $\mathcal{R}$ in a unique point and, moreover, this point is the expectation tuple belonging to a unique pure state. Thus, according to (11), it belongs to $\mathcal{S}$.

Before showing this feature to be generic, let us give some examples how it may fail. The simplest are reducible tuples, where the straight lines arising from the convex hull of widely separated $\xi \cdot A$ are non-degenerate. Then the supporting hyperplane (24) meets $\mathcal{R}$ in a unique point and, moreover, this point is the expectation tuple belonging to a unique pure state. Thus, according to (11), it belongs to $\mathcal{S}$.

For commuting pairs the Wigner function is a collection of $\delta$-peaks. Now consider a small perturbation, so that $A_1$ and $A_2$ nearly commute. Then the $\delta$-peaks become approximate, and one observes numerically that the former peaks become connected by singular ellipses. An example of this is shown in Fig. 5. For simplicity, we choose a basis in which the unperturbed parts are jointly diagonalized. So that the perturbation consists in the off-diagonal elements. Then we claim the following

**Property 7.** Let $A_1$, $A_2$ be hermitian matrices, so that the off-diagonal elements are small, say of order $\varepsilon$. Then singular set $\mathcal{S}$ is close to the union of ellipses, which are the singular sets of corresponding $2 \times 2$ submatrices.

With a small perturbation of a commuting operator tuple one might expect the eigenvectors to be close in first order to those of the commuting system, and hence also the expectation tuples of the perturbed eigenvectors. However, the singular manifold $\mathcal{S}$ of such a perturbation is not at all concentrated near the unperturbed $\delta$-points that we know from the previous paragraph. It turns out that the explanation is in terms of the well-known phenomenon of avoided level crossing.

To illustrate this effect, we begin with two commuting operators $A_1$, $A_2$, the one-parameter family of operators $C(t) = A_1 \cos t + A_2 \sin t$. The eigenfunctions (say, $| \mu \rangle$) are the same for all $t$, and the eigenfunctions $c_\mu(t) = a_{1, \mu} \cos t + a_{2, \mu} \sin t$ simple sinusoidal functions (see the thin lines in Fig. 6). These curves contain all information about $\mathcal{S}$, also in general. Indeed, for any $t$ we can consider a normalized eigenvector $\psi$ of $C(t)$ at some degenerate point. Then by the eigenvalue property the expectation tuple $(a_1, a_2)$ satisfies $c_\mu(t) = a_1 \cos t + a_2 \sin t$. On the other hand, by the perturbation theory of the non-degenerate eigenvalue we get the $t$-derivative $c_\mu(t) = -a_1 \sin t + a_2 \cos t$. Hence, $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{pmatrix} \begin{pmatrix} c_\mu(t) \end{pmatrix}.$

The only thing special about the commuting case is that the $a_i$ do not depend on $t$.

We now turn on a small perturbation so that $A_1(\varepsilon) = A_1 + \varepsilon A'_1$, with $\varepsilon$ small. The resulting eigenvalue curves of $C(t, \varepsilon)$ are shown as thick lines in Fig. 6, and exhibit

\[
4a_1^4 + 4a_1^4 - 27a_2^2 - 18a_1a_2^2 + 13a_1^2a_2^2 + 32a_2^4 = 0. \tag{26}
\]
For a more quantitative theory, it is helpful to consider the avoided crossing phenomenon as a case of degenerate perturbation theory in two parameters. Imagine a family of curves parametrized by the perturbation $\varepsilon$ and drawn at height $\varepsilon$ above the paper plane. Let us focus on a crossing that happens at $t = 0$ (for other points we may just rotate the original operators and the parameter plane). This is to say, there are two indices $\mu, \nu$ for which $a_{1,\mu} = c_{\mu}(0) = c_{\nu}(0) = a_{1,\nu}$, i.e., $A_1$ has a degenerate eigenvalue. We look at the first order perturbation theory of this eigenvalue with respect to the two parameters $t$ and $\varepsilon$, i.e., in this order,

$$C(t, \varepsilon) = A_1 + \varepsilon A'_1 + t A_2.$$  \hfill (28)

Although one of the standard textbooks on perturbation theory [21, Sect. II.7] turns away in disgust from multi-parameter perturbations (see, however, [22]), we can extract what we need by simple application of the standard one-parameter theory. Along any straight line in $(\varepsilon, t)$-parameter space, for example by fixing $t/\varepsilon$ and taking $\varepsilon$ as the perturbation parameter, we can apply the standard theory. We conclude that along this line the degenerate eigenvalue of $A_1$ branches, and the slope of the two branches are the eigenvalues of $P(\varepsilon A'_1 + (t/\varepsilon) A_2)P$, where $P$ is the projection onto the degenerate eigenspace, spanned by $|\mu\rangle$ and $|\nu\rangle$. The crucial fact is now that the perturbation directions $A'_1$ and $A_2$ usually do not commute. Therefore the slope of the branching is not linear in the two perturbation parameters. However, since that property holds along any single line we get (to first order) a conical surface.

Since the projection $P$ is two-dimensional, we can say more: We can write these operators as $PA_1P = x_0 \mathbf{1} + \vec{x} \cdot \vec{\sigma}$, where $x_0, x_1, x_2, x_3$ are real and the scalar product is between $\vec{x} = (x_1, x_2, x_3)$ and the vector $\vec{\sigma}$ of Pauli matrices. Writing $PA_2P$ similarly with coefficients $y_i$ we find the eigenvalues of (28) to be

$$c(t, \varepsilon) = a_{1,\mu} + (\varepsilon x_0 + ty_0) \pm |\varepsilon \vec{x} + t \vec{y}|.$$  \hfill (29)

As a function of $(\varepsilon, t)$ this is an elliptic double cone, in the simples case ($A'_1$ and $A_2$ distinct Pauli matrices) a section of the light cone of SRT. Coming back to the family of curves drawn with the perturbation parameter $\varepsilon$ as the vertical direction: Near a crossing we expect to see the intersection of a double cone with the $\varepsilon$-plane, which is a pair of hyperbolas.

Drawing the expectation curves determined by the thick lines in Fig. 6 then gives ellipses. But we do not have to through the detailed computation, because what we need to compute is exactly the same as for the singular manifold of the Wigner function of the pair $(PA_1P, PA_2P)$. These are obviously ellipses, because $\mathcal{R}$ is just a projection of the Bloch ball, and $\mathcal{S}$ is its boundary. This is the description given in the text.
We close this section with a remark on the specific choices made for the diagrams. The basic shape of the five points and the connecting lines (ellipses in the limit $\varepsilon \to 0$ is familiar to German children, since it comes with a rhyme and the challenge to draw it in one go without lifting the pencil. This fixes the commuting $A_1, A_2$. The perturbations were chosen off-diagonally to connect appropriate corners. Strengths of these matrix elements were chosen deterministically, and so that after rotation one would get the same degenerate perturbation problem, i.e., ellipses of roughly the same width. Of course, the matrix element connecting the top vertex to the bottom ones were chosen to be zero. Since the perturbation chosen is already fairly large, however, this gives second order effects, and one sees thin ellipses appearing around these “unwanted” diagonals. In other words, the corresponding crossings in Fig. 6 are also “avoided” even if that is not apparent. To get smoother pictures we therefore omitted some points around these crossings, and replacing them by straight lines.

H. Informational completeness

By informational completeness we mean the property that the Wigner function determines the state, i.e., that the map $\rho \mapsto \mathcal{W}_\rho$ is injective. This is a key feature of the standard Wigner function. However, in the finite case we have:

Property 8. For every pair of hermitian matrices $(A_1, A_2)$ informational completeness fails. For generic triples $(A_1, A_2, A_3)$ it holds.

It is convenient to rephrase informational completeness in terms of Weyl-ordered moments: $\rho$ is determined by $\mathcal{W}_\rho$ if and only if the Weyl-ordered moments are weakly dense in $\mathcal{B}(\mathcal{H})$, or in finite dimension: when the moments span the full matrix algebra. Let us denote by $\mathcal{M}_W$ the linear space of operators obtained in this way. Then informational completeness is equivalent to $\mathcal{M}_W = \mathcal{M}_d$.

To establish the negative statement in Property 8 it suffices to point out some hermitian element that has zero trace with all Weyl-ordered moments of two operators. This will be $i[A_1, A_2]$. This is easily seen as follows. For arbitrary $R \in \mathbb{N}$, and $\lambda \in \mathbb{R}$:

\[
\text{tr}(i[A_1, A_2](A_1 + \lambda A_2)^R) = \int i \text{tr}((A_1 + \lambda A_2, A_2)(A_1 + \lambda A_2)^R) = i \text{tr}((A_1 + \lambda A_2)^R+1, A_2) = 0.
\]

(30)

So if $[A_1, A_2] \neq 0$, informational completeness fails. When $[A_1, A_2] = 0$, it fails anyhow, because then only diagonal elements of $\rho$ in the common eigenbasis of $A_1$ and $A_2$ can be distinguished. The reason that this failure does not extend to the standard case is that two density operators can never differ by $i[P, Q] = \mathbb{I}$.

There is a second straightforward argument for Property 8, which shows that there must be more than just the commutator. When $A_1$ and $A_2$ are both real in the same basis, the same dimension estimate follows from Prop. 1. However, this is just a special case, since usually no such basis can be found.

Lemma 2. For every pair of hermitian $d \times d$-matrices $(A_1, A_2)$ the real vector space spanned by the Weyl-ordered moments satisfies

\[
\dim \mathcal{M}_W \leq \frac{d(d+1)}{2}.
\]

(31)

Proof. We need to find the operator coefficients of $\xi_1^\dagger \xi_2^{R-r}$ in the expansion of $(\xi; A)^R$. There are $R+1$ such coefficients. On the other hand, higher powers $R$ than $d-1$ lead to nothing new, because, by the Cayley-Hamilton Theorem, they can be expressed as linear combinations of lower powers. This leaves at most $\sum_{R=0}^{d-1}(R+1) = d(d+1)/2$ distinct coefficients.

This lets us expect a high dimensional orthogonal complement of $\mathcal{M}_W$. In fact, we can give an algebraic description in terms of $A_1$ and $A_2$, of which the commutant is just the simplest case.

Proposition 3. Let $A_1, A_2$ be hermitian matrices of finite dimension, and let $\mathcal{M}_W$ be the real linear span of the Weyl ordered monomials. Then, for $P_1, P_2 \in \mathcal{M}_W$

\[
\text{tr}(A_1 P_1 A_2 P_2) = \text{tr}(A_2 P_1 A_1 P_2).
\]

(32)

In particular, if two density operators $\rho, \rho'$ are related by

\[
\rho' = \rho + i(A_1 PA_2 - A_2 PA_1),
\]

(33)

with $P \in \mathcal{M}_W$, then they have the same Wigner function.

Proof. It suffices to prove (32) for the case $P_1 = (A_1 + \lambda A_2)^m$, and $P_2 = (A_1 + \mu A_2)^m$ with real $\lambda, \mu$, which we fix for the moment. Then for hermitian $X, Y$, consider the real bilinear form

\[
f(X, Y) = \Im m \text{tr}(XP_1 Y P_2).
\]

(34)

This is antisymmetric, because $f(X, X) = \text{tr}((XP_1 X P_2))$ is the trace of a product of hermitian operators, hence real. We have to show that $f(A_1, A_2)$, i.e., the difference of the two sides in (32), is zero. Using now the special form of $P_1, P_2$ we get

\[
0 = \Im m \text{tr}((A_1 + \lambda A_2)^m(A_1 + \mu A_2)^m+1)
\]

(35)

\[
= f(A_1 + \lambda A_2, A_1 + \mu A_2)
\]

\[
= \lambda f(A_2, A_1) + \mu f(A_1, A_2)
\]

\[
= (\lambda - \mu) f(A_2, A_1).
\]

Hence $f(A_1, A_2) = 0$, when $\lambda \neq \mu$. But $f$ depends continuously on these parameters through $P_1$ and $P_2$, so $f(A_1, A_2) = 0$ for any choice of $\lambda, \mu$. \qed
Now experimenting with pairs of random matrices one always finds that the dimension estimate (31) is tight, and that Prop. 3 gives a complete description of $\mathcal{M}_W$ and its orthogonal complement. This shows that there are no further algebraic constraints, and suggests that the exceptions (e.g., reducible pairs) are of measure zero. In the same way, a study of triples of matrices suggests that informational completeness almost always holds.

For proving informational completeness, e.g., for the example of spins, we have the following criterion.

**Lemma 4.** Let $A_1, \ldots, A_n$ be the generators of an irreducible representation of a compact connected Lie group. Then $W$ is informationally complete.

**Proof.** By “the generators” we mean that there is a basis of the Lie algebra of $n$ elements, and the $A_k$ are the corresponding generators of one-dimensional subgroups in the given representation. The exponential operators $\exp(i\xi A)$ are then just the unitaries of the group representation, written out using the exponential map from the Lie algebra to the Lie group. It is a basic fact that this map is onto [23]. Therefore, the product of any two such operators can be written again in the same form. But then the linear span of the operators $\exp(i\xi A)$ is the same as the span as an operator algebra. This algebra is also clearly closed under adjoints. Therefore, by von Neumann’s bicommutant theorem, it is dense if and only if its commutant, the set of operators commuting with all these operators consists of the multiples of the identity. But this is just the assumption of irreducibility of the representation. 

---

**I. Completeness of normally ordered moments**

Other ordering schemes for monomials have been proposed, and each of them will produce its own variant of Wigner functions [7]. Let us briefly discuss the informational completeness question for these. “Normal” ordering gives the monomials $A_1^n_1 \cdots A_n^n_n$. The Wigner function then has support on the eigenvalue tuples in $\mathbb{R}^n$, and (in the non-degenerate case) the corresponding operator coefficients will be rank one operators $|\phi\rangle\langle\psi|$, with $\phi$ and eigenvector of $A_1$ and $\psi$ and eigenvector of $A_n$. The marginal property holds only for the given operators, but not for their linear combinations. The corresponding “quasi-probabilities” are usually complex, so one might want to take the real part [24]. Since the eigenvectors of $A_2, \ldots, A_{n-1}$ hardly enter the game, the most interesting case is $n = 2$.

**Proposition 5.** Let $A_1, A_2$ be hermitian matrices with non-degenerate spectra. Then the normally ordered moments are complete iff no eigenvector of $A_1$ is orthogonal to an eigenvector of $A_2$.

**Proof.** For any polynomial functions $f, g$,

$$ f(A_1)g(A_2) = \sum_{ij} f(a_i)(b_j)|\phi_i\rangle\langle\phi_i| |\psi_j\rangle\langle\psi_j|, \quad (36) $$

where $a_i, \phi_i$ and $b_j, \psi_j$ are the eigenvalues and eigenvectors of $A_1$ and $A_2$, respectively. Hence all such expressions are in the span of the normally ordered moments. Taking $f$ to be $1$ on one eigenvalue of $A_1$ and $0$ on the others, and making a similar choice for $g$ shows that the span of the normally ordered moments is exactly the span of the operators $|\phi_i\rangle\langle\phi_i| |\psi_j\rangle\langle\psi_j|$. Since without the scalar product factor the $|\phi_i\rangle\langle\psi_j|$ clearly form an orthonormal basis of the Hilbert-Schmidt class, the necessary and sufficient condition for the operators with those factors to be complete is that no $|\phi_i\rangle\langle\psi_j|$ vanishes.

Of course, normal (“Wick”)-ordering plays an important role in field theory, when it is applied to creation and annihilation operators to control diverging vacuum expectations. In its application to arbitrary operators [7, 24], it seems a bit silly.

---

**J. Connection to the BMV-conjecture**

**Property 9.** Let $\xi, \eta \in \mathbb{R}^n$, and $\xi_0, \eta_0 \in \mathbb{R}$ such that $(\xi A + \xi_0 I) \geq 0$ and $(\eta A + \eta_0 I) \geq 0$, and take $\rho = 1$. Then for all $n, m \in \mathbb{N}$,

$$ \int da \mathcal{W}_\rho(a) (\xi a + \xi_0)^n (\eta a + \eta_0)^m \geq 0 \quad (37) $$

Note that with a single factor this just follows from the marginal property. Moreover, each factor is positive on the support of $\mathcal{W}_\rho$ by Property 1. Moreover, by changing the basis in the state of $A_k$, maybe shifting by multiples of $\mathbf{1}$ and integrating out all but the first two variables, we can equivalently state this property for mixed moments in the case $n = 2$, with $(\xi A + \xi_0 I) = A_1 \geq 0$, and $(\eta A + \eta_0 I) = A_2 \geq 0$. Then it reduces via (5) to the statement that all Weyl-ordered moments of two positive operators have positive trace.

This is exactly the equivalent reformulation by Lieb and Seiringer [25] of the Bessis-Moussa-Villani (BMV) conjecture [26]. This conjecture says that it is impossible to tell from the functional dependence of a partition function on a linear coupling parameter in the Hamiltonian, whether a system is classical or quantum. More formally, for any Hermitian matrices $A, B, C$, there is a probability space $(\Omega, \mu)$ and random variables $a, b$ such that

$$ \text{tr} e^{A+\lambda B} = \int_{\Omega} d\omega e^{a(\omega)+\lambda b(\omega)}. \quad (38) $$

The conjecture remained open for 38 years and was recently proved by H. Stahl [27, 28], which hence also proves Property 9.

We note that the extension to three positive operators fails even in the simplest case. We take the matrices $A, B, C$ to be one-dimensional projections in 2 dimensions, onto vectors $\phi_1, \phi_2, \phi_3$, say. The coefficient of $\alpha \beta \gamma$ in $\text{tr}(\alpha A + \beta B + \gamma C)^3$ is, up to positive constants,

$$ \text{tr}(ABC + CBA) = 2 \text{Re} \left( \langle \phi_1, \phi_2 | \langle \phi_2, \phi_3 | \langle \phi_3, \phi_1 \rangle \right). \quad (39) $$
Then with \( \phi_k = \cos(2\pi k/3)|1\rangle + \sin(2\pi k/3)|2\rangle \) each of these scalar products is \( \cos(2\pi/3) = -1/2 \), so the above product is negative. Note that this also disproves the statement about mixed moments for states other than \( 1/2 \), by taking \( \rho = C \).

K. Symmetry

Since the definition of the Wigner function contains no arbitrary choices, it is automatically covariant with respect to any symmetry respecting the basic setup. More formally, a symmetry of an operator tuple \((A_1, \ldots, A_n)\) is a unitary or antiunitary operator \(U\) on \(\mathcal{H}\) with the property that

\[
U^* A_k U = \sum_{\ell} R_{k\ell} A_\ell + \alpha_k 1,
\]

for some non-singular real matrix \(R_{k\ell}\) and constants \(\alpha_k\).

**Property 10.** The Wigner function is covariant with respect to any symmetry satisfying (40), in the sense that

\[
W_{U^*U^*}(a) = \frac{1}{\det R} W_{\rho} (R^{-1}(a - \alpha))
\]

The proof is completely straightforward, with due care taken of the complex conjugations in the antiunitary case. In the \((P, Q)\) case the symmetries comprise phase space translations and symplectic linear transformations. With bounded operators the set of symmetries will be much more constrained. Indeed, the iteration of \(T(x) = Rx + \alpha\) must not give an unbounded sequence.

For a basic family of examples, let us take the symmetry group of a regular star or polygon with \(p > 2\) vertices. Thus \(R\) will be the two-dimensional representation of the dihedral group \(G = D_p\). This consists of the \(p\) rotations by angle \(2\pi k/p\), and the \(p\) reflections around conjugate axes, one of which we choose to be vertical. The representation \(U\) will be \(p\)-dimensional, acting by permutations of the basis vectors, each of which corresponds to one point of the polygon. We take \(\rho = 1/p\) in order to get a fully symmetric Wigner function. An example with \(p = 7\) is shown in Fig. 7.

The basic question is now: How many free parameters are left for choosing \(A_1\) and \(A_2\), and how can we systematically generate such examples? There is a standard method to find the \(R\)-multiplets in any representation \(V\) of \(G\), namely to apply the projection

\[
\mathbb{P}_R = \frac{\dim R}{|G|} \sum_g (\text{tr} R_g) V_g.
\]

We want to use this projection for the representation of \(G\) acting on \(p \times p\)-matrices, i.e., \(V_g(X) = U_g^* X U_g\). In order to compute the dimension of the range of \(\mathbb{P}_R\) we consider the space of operators as a Hilbert space with Hilbert-Schmidt scalar product \(\text{tr}(X^* Y)\), with basis \(|j\rangle \langle k|\). Then

\[
\text{tr} \mathbb{P}_R = \sum_{jk} \text{tr}(\mathbb{P}_R(\langle j| |k\rangle))
\]

\[
= \frac{\dim R}{|G|} \sum_{g,jk} (\text{tr} R_g) \langle j|U_g|j\rangle \langle k|U_g^*|k\rangle
\]

\[
= \frac{\dim R}{|G|} \sum_g (\text{tr} R_g) |\text{tr} U_g|^2.
\]

Now for the two-dimensional representation \(\text{tr} R_g = 0\) for any reflection, and for the \(p\)-dimensional one \(\text{tr} U_g = 0\) for any non-trivial rotation, because no corner of the \(p\)-gon is left fixed. Thus the only term contributing to the above sum is the \(g = e\), the neutral element. Hence, with \(\text{tr} R_e = \dim R = 2\), \(\text{tr} U_e = p\), and \(|G| = 2p\), we get

\[
\text{tr} \mathbb{P}_R = 2p.
\]

We have one more condition to satisfy, namely that the \(2\)-direction is a reflection axis of the diagram, so \(A_2\) must be invariant under the corresponding reflection. This leaves a \(p\)-dimensional family of operators \(A_2\), from which \(A_1\) is obtained as a linear combination of \(A_2\) and some \(U A_2 U^*\).
IV. QUBITS

A. Qubits, \( n = 3 \)

We take \( A_k = \sigma_k, k = 1, 2, 3 \) to be the Pauli matrices. Differentiating gives

\[
\frac{\partial}{\partial \xi_k} \hat{W}_1 = i \text{tr} A_k e^{i \xi_k A}. \tag{45}
\]

Hence, if \( \rho = (1 + \sum_k r_k \sigma_k)/2 \) is a general density matrix with \( r \) in the Bloch ball, we have

\[
W_\rho(a) = \frac{1}{2} \left( 1 + r \cdot a \right) W_1(a). \tag{46}
\]

That is, \( W_\rho \) is obtained from \( W_1 \) by multiplication with a positive function. By property 1, this is true generally whenever \( \rho \) is a linear combination of the \( A_k \) and \( \mathbf{1} \).

Now \( \hat{W}_1(\xi) = 2 \cos |\xi| \). The Fourier transform can be calculated explicitly even with a Gaussian factor \( \exp(-\varepsilon \xi^2) \). It is the radial function, say \( \hat{(W_1 \ast G_\varepsilon)}(a) = W_\varepsilon(a) \) with

\[
W_\varepsilon(s) = \frac{1}{8\pi^{3/2} s^{3/2}} \left( e^{-\varepsilon^2 s} (s + 1) + e^{-\varepsilon^2 s} (s - 1) \right)
\]

\[
\rightarrow -\frac{1}{\pi s} \frac{d}{ds} e^{-\varepsilon^2 s^2} \int_{-\infty}^{+\infty} d\varepsilon \varepsilon' \delta'(s - 1).
\tag{47}
\]

Here the first simplification in line (47) is that the first term in the previous equation vanishes exponentially as \( \varepsilon \to 0 \) for all positive \( r \), so does not contribute to the limit. The second step is to identify the second term as the derivative of the Gaussian representation of the \( \delta \)-function. To check normalization, consider a radial function \( f \), which integrates to

\[
\int d^3 a \ W_1(a) f(|a|) = 4\pi \int r^2 dr \frac{1}{2\pi r} \delta(r - 1) f(r) = 2(f(1) + f'(1)). \tag{48}
\]

This is correct both for \( f = 1 \), where we get \( 2 = \text{tr} \mathbf{1} \), and for \( f(r) = r^2 \), where we get \( 6 = \text{tr} \sum_i A_i^2 \). However, one has to be aware that due to negativity (48) is not very much like the expectation \( \langle f \rangle \). For example, we have

\[
\langle (r - 1)^2 \rangle = 0 = \langle (r - 3)^2 \rangle. \tag{49}
\]

Normally we would conclude from the first equation that the distribution is concentrated on \( r = 1 \) (which is actually true), contradicting the second, which claims with the same right that \( r = 3 \). This is possible because the variance of this distribution is \( \langle r^2 \rangle - \langle r \rangle^2 = 3 - 4 = -1 \).

To summarize (46) and (47), for \( \rho = (1 + \sum_k r_k \sigma_k)/2 \) with Bloch ball coordinate \( r \in \mathbb{R}^3, |r| \leq 1 \), we have

\[
W_\rho(a) = \frac{1 + r \cdot a}{4\pi |a|} \delta(1 - |a|). \tag{50}
\]

Note that \( W_\rho \) vanishes exactly in the interior of the Bloch ball. This is no longer true if we integrate out one of the variables.

B. Qubits, \( n = 2 \)

Apart from the standard \( (P,Q) \)-case this is apparently the only example for which Wigner functions in our sense have been computed [3]. However, the result as reported by Cohen and Scully [3] makes no sense, since it contains a \( \delta \)-distribution multiplied with a function that is infinite precisely on the line where the \( \delta \)-function is non-zero. Therefore, we redo the computation, using the same basic trick. We confirm the result [3] for the regular part, but give a better explanation of how to understand the singularity of this distribution.

Exactly as for \( n = 3 \), the general case is reduced to the case \( \rho = 1/2 \), so the analogue of (46) holds, when the component \( r_3 \) is set equal to zero. The resulting state dependence is shown in Fig. 8. By rotation invariance it suffices to consider an \( \varepsilon \) regularized version at \( a = (r,0) \).
and evaluate the $\xi$-integral in polar coordinates $(s, \phi)$:

$$\mathcal{W}_{1/2, \varepsilon}(r, 0) = \frac{1}{(2\pi)^2} \int ds \, d\phi \, e^{-irs \cos \phi - \varepsilon s} \cos s$$

$$= \frac{1}{2\pi} \int_0^\infty ds \, s \cos(s) e^{-\varepsilon s} J_0(rs)$$

$$= \frac{1}{2\pi} \frac{d}{d\lambda} \left. \int_0^\infty ds \, \sin(\lambda s) e^{-\varepsilon s} J_0(rs) \right|_{\lambda=1}$$

$$= \frac{1}{2\pi} \frac{d}{d\lambda} 3m \frac{1}{\sqrt{(\xi - i\lambda)^2 + \lambda^2}} \bigg|_{\lambda=1} \quad (51)$$

The integral over the Bessel function at the third equality follows from [29][6.611.1]. After integrating this with a test function $f$ we can take the limit $\varepsilon \to 0$, so

$$\int da \, \mathcal{W}_{1/2}(a) f(a) = \frac{1}{2\pi} \frac{d}{d\lambda} \int da \, f(a) \Im \frac{-i}{\sqrt{\lambda^2 - |a|^2}} \bigg|_{\lambda=1}$$

$$= \frac{1}{2\pi} \frac{d}{d\lambda} \left[ \int_{|a|<\lambda} da \, f(a) \frac{\lambda f(\lambda b)}{\lambda} \right]_{\lambda=1}$$

$$= \frac{1}{2\pi} \int_{|a|<1} da \, \left( f(a) + a \cdot \nabla f(a) \right) \frac{\lambda}{\sqrt{1 - |a|^2}} \quad (52)$$

Here, at the third equality we substituted $b = \lambda a$, $db = \lambda^2 da$. The integral (52) makes sense as written, because $f$ is a test function, hence differentiable, and $(1 - a^2)^{-1/2}$ is an integrable function. For writing it just as an integral over $f$, not its derivatives, we use partial integration, which holds by definition of the distributional derivative:

$$\mathcal{W}_{1/2}(a) = \frac{1}{2\pi} \left( \frac{1}{\sqrt{1 - |a|^2}} - \sum_j \frac{\partial}{\partial a_j} a_j \right) \frac{\lambda}{\sqrt{1 - |a|^2}} \quad (53)$$

$$= \frac{-1}{2\pi (1 - |a|^2)^{3/2}} + \text{boundary} \quad (54)$$

Here (53) is correct as a mere rewriting of (52), provided the derivative is considered as a distributional derivative, and the function is continued as $\mathcal{W}(a) = 0$ for $|a| > 1$. Equation (54) is obtained by evaluating the derivatives in the sense of ordinary functions. That is correct even without the boundary term for the integral with any test function with support strictly inside the unit circle. So this expression describes the “non-singular part” of the Wigner function, and is uniquely determined by this property.

As an expression for the whole distribution, however, (54) is misleading, because it suggests that $\mathcal{W}_{1/2}(a)$ can be written as the sum of the non-singular part and some purely singular part supported on the circle. This is not true, because the regular part by itself is not integrable, so not even a distribution. So any attempt to use (54) term by term results in an indeterminate expression.

![FIG. 9. Section ($y = 0$) through the Wigner function $\mathcal{W}_{1/2}$ for $(\sigma_x, \sigma_y)$ or, equivalently, the case $|1\rangle$ in Fig. 8. Black solid line: the exact regular part (54). Red, dashed: numerical calculation of the the whole distribution regularized with a Gaussian shown in the center (dotted).](image)

V. CONCLUSION AND OUTLOOK

The Wigner distribution is an object canonically associated to any tuple of operators and a state, and this already makes it worthwhile to find out what story it has to tell. We have shown that there is a remarkable richness in the structure of singularities. But many more details will have to be explored to make it a tool that speaks more directly to intuition. Among these will be a detailed study of the case of angular momentum including the classical limit, and the transition to the Holstein-Primakov approximation. In that approximation we get a phase space with standard position and momentum operators, but also the oscillator Hamiltonian as a third operator. This is an enhancement of the standard Wigner function of some interest in its own right: It resolves the paradox that while the energy function on phase space takes continuous values, its Wigner quantization has discrete spectrum.

Further worthwhile topics are (1) the analysis of the structure of the singularities as in Sect. IVB, but for generic singular lines using perturbation theory. (2) A rigorous analysis including the unbounded case, under the natural condition that the exponential series for all $\exp(i\xi \cdot A)$ make sense on a common dense domain. (3) A derivation of uncertainty bounds of measurement or preparation type [30, 31] between the $A_k$ in terms of the Wigner function.

ACKNOWLEDGEMENTS

We acknowledge support from the RTG 1991 and CRC 1227 DQ-mat founded by the DFG, and the BMBF project Q.com. During the writing phase R.F.W. enjoyed the hospitality of the Inst. Henri Poincaré in Paris, and enlightening conversations with S. Weis. R.S. gratefully acknowledges A. Ketterer and A. Asadian for inspiring
[1] E. P. Wigner, “On the quantum correction for thermodynamic equilibrium,” Phys. Rev. 40, 749–759 (1932).
[2] L. Cohen, Time-frequency analysis (Prentice-Hall, 1995).
[3] L. Cohen and M. O. Scully, “Joint Wigner distribution for spin-1/2 particles,” Found. Phys. 16, 295–310 (1986).
[4] M. O. Scully and L. Cohen, “Quasi-probability distributions for arbitrary operators,” in The physics of phase space, edited by Y.S. Kim and W.W. Zachary (Springer, 1987) pp. 253–260.
[5] R. Schwonnek and R.F. Werner, “Wigner distributions for n arbitrary operators,” (2018), companion letter to this paper [URL will be inserted by publisher].
[6] T. Tao, “Two small facts about Lie groups,” https://terrytao.wordpress.com/2011/06/25/two-small-facts-about-lie-groups/ (2011).
[7] L. Cohen, The Weyl Operator and its generalization (Springer, Basel, 2003).
[8] L. Hörmander, The Analysis of Linear Partial Differential Operators, vol. I (Springer, 1983).
[9] R. L. Hudson, “When is the Wigner quasi-probability density non-negative?” Rep. Math. Phys. 6, 249 (1974).
[10] T. Bröcker and R. F. Werner, “Mixed states with positive wigner functions,” J. Math. Phys. 36, 62–75 (1995).
[11] M. Reed and B. Simon, Methods of modern mathematical physics, vol.2: Fourier analysis, self-adjointness (Academic Press, 1975).
[12] J. Harris, Algebraic Geometry (Springer, New York, 2013).
[13] R. Kippenhahn, “Über den Wertevorrat einer Matrix,” Math. Nachr. 6, 193–228 (1951).
[14] M.-T. Chien and H. Nakazato, “Joint numerical range and its generating hypersurface,” Lin. Algebra Appl. 432, 173 – 179 (2010).
[15] S. Weis, “On a theorem by Kippenhahn,” (2017), arXiv:1705.09035.
[16] M. Joswig and B. Straub, “On the numerical range map,” J. Austr. Math. Soc. 65, 267–283 (1998).
[17] Thomas J. Laffey, “A counterexample to Kippenhahn’s conjecture on hermitian pencils,” Linear Algebra and its Applications 51, 179 – 182 (1983).
[18] D. Heinron, “Semidefinite geometry of the numerical range,” E-Journal Lin. Alg. 20, https://doi.org/10.13001/1081–3810.1377 (2010), arXiv:1003.4837.
[19] J. von Neumann and E. P. Wigner, “Über das Verhalten von Eigenwerten bei adiabatischen Prozessen,” Phys. Zeitschr. 30, 467–470 (1929).
[20] K. Szymański, S. Weis, and K. Życzkowski, “Classification of joint numerical ranges of three hermitian matrices of size three,” (2016), arXiv:1603.06569.
[21] T. Kato, Perturbation theory of linear operators (Springer, 1966/1984).
[22] H. Baumgärtel, Analytic perturbation for matrices and operators (Birkhäuser, 1985).
[23] T. Tao, “Some notes on Weyl quantisation,” https://terrytao.wordpress.com/2012/10/07/some-notes-on-weyl-quantisation/ (2012).
[24] A. O. Barut, “Distribution functions for noncommuting operators,” Phys. Rev. 108, 565–569 (1957).
[25] E. H. Lieb and R. Seiringer, “Equivalent forms of the Bessis-Moussa-Villani conjecture,” J. Stat. Phys. 115, 185–190 (2004), arXiv:math-ph/0210027.
[26] D. Bessis, P. Moussa, and M. Villani, “Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics,” J. Math. Phys. 16, 2318–2325 (1975).
[27] H. R. Stahl, “Proof of the BMV conjecture,” Acta Math. 211, 255–290 (2013), arXiv:1107.4875.
[28] A. Eremenko, “H. Stahl’s proof of the BMV conjecture,” Mat. Sbornik 206, 87–92 (2015), arXiv:1312.6003.
[29] I. S. Gradshteyn and I. M. Ryzhik, Tables of Integrals, Series, and Products (Academic Press, 1980).
[30] L. Dammeier, R. Schwonnek, and R.F. Werner, “Uncertainty relations for angular momentum,” New J. Phys. 17, 093046 (2015), arXiv:1505.00049.
[31] R. Schwonnek, D. Reeb, and R.F. Werner, “Measurement uncertainty for finite quantum observables,” Mathematics 4, 38, arXiv:1604.00382.