The Symplectic Camel and the Uncertainty Principle: The Tip of an Iceberg?

Maurice A. de Gosson

Institutions: University of Vienna

Found Phys (2009) 39: 194–214
DOI 10.1007/s10701-009-9272-2

The Symplectic Camel and the Uncertainty Principle: The Tip of an Iceberg?

Maurice A. de Gosson

Received: 6 October 2008 / Accepted: 6 January 2009 / Published online: 17 January 2009

Abstract We show that the strong form of Heisenberg's inequalities due to Robertson and Schrödinger can be formally derived using only classical considerations. This is achieved using a statistical tool known as the "minimum volume ellipsoid" together with the notion of symplectic capacity, which we view as a topological measure of uncertainty invariant under Hamiltonian dynamics. This invariant provides a right measurement tool to define what "quantum scale" is. We take the opportunity to discuss the principle of the symplectic camel, which is at the origin of the definition of symplectic capacities, and which provides an interesting link between classical and quantum physics.

Keywords Uncertainty principle · Symplectic non-squeezing · Symplectic capacity · Hamiltonian mechanics

1 Introduction

Common sense tells us that classical mechanics is not "quantum"; in fact "nonclassicality" is a key concept supporting the need for a quantum theory. One of the most decisive hallmarks of nonclassical behavior seems to be, no doubt, the uncertainty principle since it appears to be a phenomenon that classical physics cannot account for. One way of expressing this principle mathematically is, in one degree of freedom, to use the Heisenberg inequality $\Delta P \Delta X \geq \frac{1}{2}\hbar$ which is a particular case of the Schrödinger–Robertson inequality

$$\Delta X^2 \Delta P^2 \geq \mathrm{Cov}(X, P)^2 + \frac{1}{4}\hbar^2. \tag{1}$$

To my parents.

M.A. de Gosson (✉)
Faculty of Mathematics, NuHAG, University of Vienna, Nordbergstrasse 15, 1090 Vienna, Austria
e-mail: maurice.de.gosson@univie.ac.at

The aim of this article is to show that the inequality (1), and its generalization to several degrees of freedom

$$(\Delta X_j)^2(\Delta P_j)^2 \geq \operatorname{Cov}(X_j, P_j)^2 + \frac{1}{4}\hbar^2, \quad j = 1, 2 \ldots \tag{2}$$

can be derived for large statistical ensembles by using only classical arguments, the co-variances being here interpreted in terms of measurement errors.

A caveat: I *do not* claim that quantum mechanics can been derived using solely classical arguments; for quantum uncertainty to emerge from the inequalities (2) one has to justify by some physical argument the existence of a universal constant $\hbar$, the same for all possible systems. *What* I claim is that recent advances in symplectic geometry and topology allow to highlight the fact that classical and quantum mechanics are formally much closer than might appear at first sight; in fact the "symplectic camel" of the title of this article provides a right measurement tool to define what a "quantum scale" is, and allows to state the uncertainty principle in invariant (under Hamiltonian dynamics) terms.

In the case of one degree of freedom the idea is the following. Consider a cloud Ω of points lying in the phase plane, and consisting of a number $K \gg 1$ of points $z_1 = (x_1, p_1), \ldots, z_K = (x_K, p_K)$; each of the points corresponds to a joint position/momentum measurement of a physical system with one degree of freedom. It is a standard procedure in robust statistical analysis to "clean up" such a cloud of points by down-weighting outliers (i.e. observations that do not follow the pattern of the majority of the data). There are various procedures for doing this, but the method we are interested in is the minimum area ellipse method; we will describe this method more in detail in Sect. 3.1 but for the moment it suffices to say that it consists in using arguments from convex geometry to replace Ω by an ellipse $\mathcal{J}$ (the John–Löwner ellipse) containing Ω. The center of that ellipse is then identified with the mean (=expectation value) and the shape of the ellipse determines the covariance. More specifically, if $\mathcal{J}$ consists of all points $z = (x, p)$ such that

$$(z - \overline{z})^T M^{-1}(z - \overline{z}) \leq m^2 \tag{3}$$

where M is a positive-definite matrix the mean is $\overline{z}$ and the covariance matrix

$$\Sigma = \begin{pmatrix} \Delta X^2 & \operatorname{Cov}(X, P) \\ \operatorname{Cov}(P, X) & \Delta P^2 \end{pmatrix} \tag{4}$$

is then obtained by an adequate choice m_0^2 of m^2, in agreement with an assumed underlying distribution, so that Σ is determined by rewriting (3) as

$$\mathcal{J} : (z - \overline{z})^T \Sigma^{-1}(z - \overline{z}) \leq m_0^2. \tag{5}$$

(For instance, if the points $z_1, \ldots, z_K$ are close to normally distributed one typically chooses $m_0^2 = \chi^2_{0.5}(2) \approx 1.39$.) By definition, the ellipse

$$\mathcal{C} : \frac{1}{2}(z - \overline{z})^T \Sigma^{-1}(z - \overline{z}) \leq 1 \tag{6}$$

is the *covariance ellipse*. (We have included a factor $\frac{1}{2}$ in the definition of $\mathcal{C}$ in analogy of what is done in quantum mechanics, where $\mathcal{C}$ is called the "Wigner ellipse"; see for instance [17].) We note that $\mathcal{C}$ is homothetic to $\mathcal{J}$ by a factor of $\sqrt{2}/m_0$ with respect to $\bar{z}$ and that we thus have $\operatorname{Area}(\mathcal{C}) = 2\operatorname{Area}(\mathcal{J})/m_0^2$. We now make the crucial assumption that

$$\operatorname{Area}(\mathcal{J}) \geq \frac{1}{4}m_0^2 h$$

that is, equivalently,

$$\operatorname{Area}(\mathcal{C}) \geq \frac{1}{2}h. \tag{7}$$

Here h is a constant > 0 (which *could* be Planck's constant in quantum mechanics!). Since

$$\operatorname{Area}(\mathcal{C}) = 2\pi(\det\Sigma)^{1/2} = 2\pi[\Delta X^2 \Delta P^2 - \operatorname{Cov}(X, P)^2]^{1/2} \tag{8}$$

condition (7) is strictly equivalent to the Schrödinger–Robertson inequality (1) with $\hbar = h/2\pi$.

This inequality is moreover conserved in time under a Hamiltonian evolution: if it is true at an initial time, say $t = 0$, it will be true for all times, past and future (I will show why it is so in Sect. 3). But what about the case of many degrees of freedom? Suppose that the system under scrutiny consists of N particles; we must then work in a $6N$ dimensional phase space and John–Löwner's ellipse then becomes an ellipsoid $\mathcal{J}$ in $\mathbb{R}^{6N}$; to that ellipsoid one can again associate a statistical covariance matrix Σ determined by the shape of $\mathcal{J}$ and a covariance ellipsoid $\mathcal{C}$. What condition should we now impose on $\mathcal{C}$ in order to derive the inequalities (2)? A natural guess is that we should ask that the *volume* of $\mathcal{C}$ should be at least $(\frac{1}{2}h)^{3N}$; this guess is in addition perfectly consistent with the usual procedure in quantum statistical mechanics where it is customary to coarse-grain phase space in "quantum cells" of volume $\sim h^{3N}$. Unfortunately this idea fails. It turns out that the correct assumption for dealing with multi-dimensional systems is of a much more subtle nature. It consists in demanding that the *symplectic capacity* of the covariance ellipsoid $\mathcal{C}$ be at least $\frac{1}{2}h$ which we write symbolically as

$$c(\mathcal{C}) \geq \frac{1}{2}h. \tag{9}$$

I will fully justify this apparently mysterious statement in Sect. 3. The existence of symplectic capacities follows from a deep result of symplectic topology nicknamed the "principle of the symplectic camel", which I review in Sect. 2. That principle was already advertised by Ian Stewart in *Nature* [41] in 1987; as Stewart put it "…*we are witnessing just the tip of the symplectic iceberg*." Unfortunately, this iceberg seems not to have received the attention it deserves in the physical literature.

Notation and Terminology

The phase space of a system with n degrees of freedom is $\mathbb{R}^n \times \mathbb{R}^n \equiv \mathbb{R}^{2n}$; for instance if we are dealing with N point-like particles in 3-dimensional configuration space we have $n = 3N$.

We will write $x = (x_1, \dots, x_n)$, $p = (p_1, \dots, p_n)$ and $z = (x, p)$. Whenever matrix calculations are performed, x, p, and z are viewed as column vectors. The matrix

$$J = \begin{pmatrix} 0_n & I_n \\ -I_n & 0_n \end{pmatrix}$$

is the standard symplectic matrix and $\sigma(z, z') = (z')^T Jz$ is the associated symplectic form. A $2n \times 2n$ real matrix S is symplectic if $S^T JS = SJS^T = J$; equivalently $\sigma(Sz, Sz') = \sigma(z, z')$ for all vectors z and z'. A symplectic matrix has determinant one. Symplectic matrices form group: the real symplectic group $\mathrm{Sp}(2n)$. A transformation $f(x, p) = (x', p')$ of phase space $\mathbb{R}^{2n}$ is said to be *canonical* if its Jacobian matrix

$$Df(x, p) = \frac{\partial(x', p')}{\partial(x, p)}$$

calculated at any phase space point (x, p) were f is defined is symplectic.

In this paper h and $\hbar = h/2\pi$ denote positive constants. We leave it to the Reader to decide whether h should be identified with Planck's constant, or not.

2 The Principle of the Symplectic Camel

We will consider a physical system $\mathcal{S}$ consisting of N point-like particles moving in physical 3-dimensional space. The position (resp. momentum) coordinates of the first particle are denoted by x_1, x_2, x_3 (resp. p_1, p_2, p_3), those of the second particle by x_4, x_5, x_6 (resp. p_4, p_5, p_6), and so on. We assume that the phase-space evolution of that system is governed by Hamilton's equations

$$\frac{dx_j}{dt} = \frac{\partial H}{\partial p_j}(x, p), \qquad \frac{dp_j}{dt} = -\frac{\partial H}{\partial x_j}(x, p). \tag{10}$$

They form a system of $2n = 6N$ differential equations; they determine a phase space flow f_t^H which consists of canonical transformations (see Goldstein's book [10]).

2.1 Gromov's Non-Squeezing Theorem

A Hamiltonian flow f_t^H is volume preserving: this is Liouville's theorem, one of the best known results from elementary statistical mechanics. It is easy to see this using the fact that the Jacobian matrix of f_t^H is symplectic at each point and thus has determinant equal to one. Liouville's theorem is perhaps also one of the most *understated* results of classical mechanics, because in addition of being volume-preserving, Hamiltonian flows have a surprising—I am tempted to say an *extraordinary*—additional property as soon as the number of degrees of freedom is superior to one.

Assume that the number N of particles of the system $\mathcal{S}$ is very large and that the particles are very close to each other. We may in this case approximate $\mathcal{S}$ with a "cloud" of points in phase space $\mathbb{R}^{2n}$. Suppose that this cloud is, at time $t = 0$ spherical so it is represented by a phase space ball $B(r)$ with center (a, b) and radius r:

$$B(r) : |x-a|^2 + |p-b|^2 \leq r^2. \tag{11}$$

The orthogonal projection of that ball on any plane of coordinates will always be a circle with area πr^2. Let us watch the motion of this spheric phase-space cloud as time evolves. It will distort and may take after a while a very different shape, while keeping constant volume. However—and this is the surprising result—the projections of that deformed ball on any plane of *conjugate coordinates* x_j, p_j.will never decrease below its original value πr^2! If we had chosen, on the contrary, a plane of non-conjugate coordinates (such as x_1, p_2 or x_1, x_2, for example) then there would be no obstruction for the projection to become arbitrarily small. The property just described is not a physical observation, but a mathematical theorem proved by Mikhail Gromov [11] in 1985. If we choose $r = \sqrt{\hbar}$ then Gromov's theorem says that the projection of the ball $B(\sqrt{\hbar})$ on a conjugate plane will always be at least $\frac{1}{2}h$ and this is of course strongly reminiscent of the uncertainty principle of quantum mechanics, of which it can be viewed as a *classical* geometrical version!

Gromov's theorem—which is often called the "symplectic non-squeezing theorem" in the mathematical literature—is indeed an extraordinary result, because it seems at first sight to conflict with the usual conception of Liouville's theorem: according to conventional wisdom, the ball $B(r)$ can be stretched in all directions by Hamiltonian flows, and eventually get very thinly spread out over huge regions of phase space, so that the projections on any plane could *a priori* become arbitrary small after some (admittedly, perhaps very long) time t. In fact, one might very well envisage that the larger the number of degrees of freedom, the more that spreading will have chances to occur since there are more and more directions in which the ball is likely to spread! A relevant phenomenon in symplectic geometry is provided by Katok's lemma [16]: consider two bounded domains Ω and Ω' in $\mathbb{R}^{2n}$ which are both diffeomorphic to the ball $B(r)$ and have same volume. Katok proved that for every $\varepsilon > 0$ there exists a Hamiltonian diffeomorphism f such that $\mathrm{Vol}(f(\Omega)\Delta\Omega')) < \varepsilon$ (here Δ denotes the symmetric difference of two sets). Thus, up to sets of (arbitrarily small) measure ε any kind of spreading is possible; the rigidity effects imposed by the non-squeezing theorem are about point-wise behavior of sets (or C^0 behavior of functions). This possible spreading phenomenon has led to many philosophical speculations about the stability of general Hamiltonian systems. For instance, in his 1989 book Roger Penrose [27, p. 174–184] comes to the conclusion that phase space spreading suggests that "*classical mechanics cannot actually be true of our world*" (p. 183, l.–3). He however adds that "*quantum effects can prevent this spreading*" (p. 184, l. 9). Penrose's second observation goes right to the point: while phase space spreading a priori opens the door to classical chaos, quantum effects have a tendency to "tame" the behavior of physical systems by blocking and excluding most of the classically allowed motions. However, Gromov's no-squeezing theorem shows that there is a similar taming in Hamiltonian mechanics preventing anarchic and chaotic spreading of the ball in phase space which would be possible if it were possible to stretch it inside arbitrarily thin tubes in directions orthogonal to the conjugate planes.

Now, why do we refer to a *symplectic camel* in the title of this paper? This is because one can restate Gromov's theorem in the following way: there is no way to deform a phase space ball using canonical transformations in such a way that we can make it pass through a hole in a plane of conjugate coordinates x_j, p_j if the area of that hole is smaller than that of the cross-section of that ball. Recall that in [35]) it is stated that

> "...*It is easier for a camel to pass through the eye of a needle than for one who is rich to enter the kingdom of God*..."

The Biblical camel is here the phase space ball and the eye of the needle is the hole in the x_j, p_j plane! For this reason it is usual to call Gromov's theorem and its variant just described the *principle of the symplectic camel*.

The discussion above was of a purely qualitative nature. It turns out that we can do better, and produce quantitative statements. For this purpose it is very useful to introduce the topological notion of *symplectic capacity*.

2.2 The Notion of Symplectic Capacity

Consider an arbitrary region Ω in phase space $\mathbb{R}^{2n}$; this region may be large or small, bounded or unbounded. By definition, the Gromov capacity of Ω is the (possibly infinite) number $c_{\min}(\Omega)$ calculated as follows: assume that there exits no canonical transformation sending any phase space ball $B(r)$ inside Ω, no matter how small its radius r is. We will then say that $c_{\min}(\Omega) = 0$. Assume next that there are canonical transformations sending $B(r)$ in Ω for some r (and hence also for all $r' < r$). The supremum R of all such radii r is called the *symplectic radius* of Ω and we define $c_{\min}(\Omega) = \pi R^2$. Thus $c_{\min}(\Omega) = \pi R^2$ means that one can find canonical transformations sending $B(r)$ inside Ω. for all $r < R$, but that no canonical transformation will send a ball with radius larger R inside that set. By its very definition we see that the Gromov capacity is a symplectic invariant, that is

$$c_{\min}(f(\Omega)) = c_{\min}(\Omega) \quad \text{if } f \text{ is a canonical transformation;} \tag{12}$$

it is obviously also monotone with respect to inclusion:

$$c_{\min}(\Omega) \leq c_{\min}(\Omega') \quad \text{if } \Omega \text{ is a subset of } \Omega' \tag{13}$$

and 2-homogeneous under phase space dilations:

$$c_{\min}(\lambda\Omega) = \lambda^2 c_{\min}(\Omega) \quad \text{for any scalar } \lambda \tag{14}$$

($\lambda\Omega$ consists of all points λz such that z is in Ω). However, the most striking property of the Gromov capacity is the following: let us denote by $Z_j(R)$ the phase-space cylinder based on the plane of conjugate variables: it consists of all phase space points whose j-th position and momentum coordinate satisfy $x_j^2 + p_j^2 \leq R^2$. We have

$$c_{\min}(B(R)) = \pi R^2 = c_{\min}(Z_j(R)). \tag{15}$$

While the equality $c_{\min}(B(R)) = \pi R^2$ is immediate by definition of $c_{\min}$, the equality $c_{\min}(Z_j(R)) = \pi R^2$ is a reformulation of the non-squeezing theorem, and hence a very deep property! In fact that theorem says that there is no way we can squeeze a ball with radius $R' > R$ inside that cylinder, because if we could then the orthogonal projection of the squeezed ball would be greater than the cross-section πR^2 of the cylinder, and this would contradict the non-squeezing theorem. We must thus have $c_{\min}(Z_j(R)) \leq \pi R^2$. That we actually have equality is immediate, observing that we can translate the ball $B(R)$ inside any cylinder with same radius, and that phase space translations are canonical transformations in their own right.

More generally one calls *symplectic capacity* any function associating to subsets Ω of phase space a non-negative number $c(\Omega)$, or $+\infty$, and for which the properties listed in (12), (13), (14), and (15) are verified (see Hofer and Zehnder [14], Polterovich [28], or Schlenk [36] for the general theory of symplectic capacities; in [3, 5] I have given a souped-down review of the topic). There exist infinitely many symplectic capacities, and the Gromov capacity is the smallest of all: $c_{\min}(\Omega) \leq c(\Omega)$ for all Ω and c. Is there a "biggest" symplectic capacity $c_{\max}$? Yes there is one, and it is constructed as follows: suppose that no matter how large we choose r there exists no canonical transformation sending Ω inside a cylinder $Z_j(r)$. We then set $c_{\max}(\Omega) = +\infty$. Suppose that, on the contrary, there are canonical transformations sending Ω inside some cylinder $Z_j(r)$ and let R be the infimum of all such r. Then, by definition, $c_{\max}(\Omega) = \pi R^2$. It is not difficult, using the non-squeezing theorem, to show that $c_{\max}$ indeed is a symplectic capacity and that we have

$$c_{\min}(\Omega) \leq c(\Omega) \leq c_{\max}(\Omega) \tag{16}$$

for every other symplectic capacity c.

Note that by definition $c_{\min}(\Omega)$ and $c_{\max}(\Omega)$ both have the dimension of an area. The homogeneity property (14) $c(\lambda\Omega) = \lambda^2 c(\Omega)$ satisfied by every symplectic capacity together with the fact that $c(B(R)) = \pi R^2$ suggests that symplectic capacities have something to do with the notion of area. In fact, the following is true: the Gromov capacity $c_{\min}(\Omega)$ of a subset in the phase plane $\mathbb{R}^2$ is the area of Ω when the latter is connected, and the maximal capacity $c_{\max}(\Omega)$ is the area when Ω is simply connected; it follows from the inequalities (16) that $c(\Omega)$ coincides with the area for all connected and simply connected domains. (The reader may easily convince himself that $c_{\min}(\Omega)$ is not the area when Ω is disconnected, and that $c_{\max}(\Omega)$ is not the area when Ω is, say, an annulus.) There exists one particular example where this relation is quite explicit, albeit in an indirect way: it is provided by the Hofer–Zehnder capacity c_{HZ} (see [14]). It has the property that whenever Ω is a bounded convex set in phase space then

$$c_{\mathrm{HZ}}(\Omega) = \oint_{\gamma_{\min}} p\,dx \tag{17}$$

where $p\,dx = p_1 dx_1 + \cdots + p_n dx_n$ and $\gamma_{\min}$ is the shortest Hamiltonian periodic orbit carried by the boundary of Ω (it is easy to show that the integral in the right-hand side of (17) is independent of the choice of the Hamiltonian, see [14]).

2.3 The Symplectic Capacity of an Ellipsoid

A very nice property is that all symplectic capacities agree on phase space ellipsoids. Let us show how this capacity can be calculated explicitly. Assume that Ω_{ell} is an ellipsoid centered at $\bar{z} = 0$; then there exists a positive-definite $2n \times 2n$ matrix M such that

$$z^T M z \leq 1. \tag{18}$$

Consider now the eigenvalues of the product matrix JM; they are the same as those of the antisymmetric matrix $M^{1/2} J M^{1/2}$ and are hence of the type $\pm i\lambda_1, \dots, \pm i\lambda_n$ where $\lambda_j > 0$. I claim that we have

$$c(\Omega_{\text{ell}}) = \pi/\lambda_{\max} \tag{19}$$

for every symplectic capacity c; here $\lambda_{\max}$ is the largest of all the positive numbers λ_j (this formula remains true if Ω_{ell} is centered at an arbitrary point $\bar{z}$ since symplectic capacities are invariant under phase space translations). We first note that in view of Williamson's famous diagonalization theorem (see [43]) there exists a symplectic matrix S such that $S^T M S$ is diagonal; more precisely

$$S^T M S = \begin{pmatrix} \Lambda & 0 \\ 0 & \Lambda \end{pmatrix} \quad \text{with } \Lambda = \operatorname{diag}(\lambda_1, \dots, \lambda_n). \tag{20}$$

Since symplectic capacities are invariant by canonical transformations it follows that $c(\Omega_{\text{ell}}) = c(S(\Omega_{\text{ell}}))$ so that it suffices to prove formula (19) when Ω_{ell} is replaced by $S(\Omega_{\text{ell}})$. Since phase space translations also are canonical, we may moreover assume that $\bar{z} = 0$ so that we have reduced the proof to the case

$$\Omega_{\text{ell}} : \sum_{j=1}^{n} \frac{1}{R_j^2}(x_j^2 + p_j^2) \leq 1 \tag{21}$$

where we have set $\lambda_j = 1/R_j^2$. Suppose that there exists a canonical transformation f sending a ball $B(R)$ inside Ω_{ell}. Then $f(B(R))$ is also contained in each cylinder $Z_j(R) : x_j^2 + p_j^2 \leq R^2$ and hence $R \leq R_{\max} = \sqrt{1/\lambda_{\max}}$ in view of the non-squeezing theorem. It follows that $c_{\min}(\Omega_{\text{ell}}) \leq \pi R_{\max}^2 = \pi/\lambda_{\max}$; since on the other hand $B(R_{\max})$ is anyway contained in Ω_{ell} we must have equality: $c_{\min}(\Omega_{\text{ell}}) = \pi/\lambda_{\max}$. A similar argument shows that we also have $c_{\max}(\Omega_{\text{ell}}) = \pi/\lambda_{\max}$; formula (19) follows since $c_{\min}$ and $c_{\max}$ are the smallest and largest symplectic capacities.

3 The Uncertainty Principle

3.1 The Minimum Volume Ellipsoid

We now extend (and explain) the minimum area ellipse sketched in the Introduction to the case where the phase space is $\mathbb{R}^{2n}$. We perform again simultaneous position and momentum measurements on K identical copies of the physical system $\mathcal{S}$ and plot the results of these measurements as a set $\Omega = \{z_1, \ldots, z_K\}$ of points in the phase space $\mathbb{R}^{2n}$. If the number K is very large we get a cloud of points which we identify with a domain of $\mathbb{R}^{2n}$; we assume that these points are in generic position, so that Ω is not contained in any subspace with dimension less than $2n$. We are going to associate an optimal ellipsoid to Ω using a method from robust multivariate statistical analysis, called the *minimum volume ellipsoid* (*MVE*) *method*. That method is based on the use of the John–Löwner's ellipsoid of a set of points, and has applications in various fields such that computational geometry, convex optimization, image processing, etc. For us its main interest comes from the fact that it is a well-established tool in multivariate statistics, and whose importance was recognized by the statistician Peter Rousseeuw in [32] (see the book [33] by Rousseeuw and Leroy for a detailed exposition; readable descriptions of the method are also given in Lopuhäa and Rousseeuw [18] and in Rousseeuw and Zomeren [34]). The MVE is a tool of choice for the study of data sets that can reasonably be assumed to come from a normally distributed random variable, but it applies to more general cases as well. The MVE method is a "high breakdown" estimator; loosely speaking this means that it can theoretically cope with data sets in which as many as 50% of the observations are unreliable. This is a decisive superiority of the method compared to, for instance, the calculation of sample mean and covariance which are not robust estimators, because only one outlier may cause highly biased estimates!

Geometrically, the MVE method amounts to finding the smallest ellipsoid circumscribing a set of points: assume that the retained points in Ω are labeled $z_1, \ldots, z_K$; the set $\mathcal{S} = \{z_1, \ldots, z_K\}$ determines a convex polyhedron $\mathcal{S}$ in $\mathbb{R}^{2n}$. Let now $\widetilde{\mathcal{S}}$ be the convex hull of $\mathcal{S}$: it is the intersection of all convex sets in $\mathbb{R}^{2n}$ which contain $\mathcal{S}$ (alternatively, it consists of all finite linear combinations $\sum_j \alpha_j z_j$ of points in $\mathcal{S}$ with coefficients $\alpha_j \geq 0$ summing up to one). A famous theorem in convex geometry proved by Fritz John in [15] in 1948 guarantees the existence of a unique ellipsoid $\mathcal{J}$ in $\mathbb{R}^{2n}$ containing $\widetilde{\mathcal{S}}$ and having minimum volume among all other ellipsoids containing that set; this ellipsoid is precisely the John–Löwner ellipsoid (Ball gives in [1] a review and some extensions of John's construction). Practically one proceeds as follows: letting k be the integer part of $\frac{1}{2}(K + 2n + 1)$ we consider the following convex optimization problem:

Find a pair $(M, \overline{z})$ where M is a real positive-definite $2n \times 2n$ matrix and $\overline{z}$ a point in $\mathbb{R}^{2n}$ such that the determinant of M is minimized subject to

$$\#\left\{ j : (z_j - \overline{z})^T M^{-1} (z_j - \overline{z}) \leq m^2 \right\} \geq k \tag{22}$$

(the symbol # stands for "number of elements of").

One proves that this problem has a unique solution if every subset of Ω with k elements is in general position (which we always assume is the case) and that the center $\overline{z}$, which is identified with the mean, does not depend on m^2. The John–Löwner ellipsoid (MVE) $\mathcal{J}$ is then unambiguously defined by the condition

$$(z - \overline{z})^T M^{-1} (z - \overline{z}) \leq m^2. \tag{23}$$

As in the Introduction we choose an adequate value m_0^2 determining the covariance matrix:

$$\mathcal{J} : (z - \overline{z})^T \Sigma^{-1} (z - \overline{z}) \le m_0^2. \tag{24}$$

For instance if the sample of phase space points z_j is normally distributed then a standard choice would be $m_0^2 = \chi_{0.5}^2(2n)$ (see the discussion in Lopuhäa and Rousseeuw [18]). We next associate to $\mathcal{J}$ a covariance ellipsoid

$$\mathcal{C} : \frac{1}{2}(z - \overline{z})^T \Sigma^{-1} (z - \overline{z}) \le 1. \tag{25}$$

The ellipsoids $\mathcal{J}$ and $\mathcal{C}$ are homothetic; in fact

$$\mathcal{C} - \overline{z} = \frac{\sqrt{2}}{m_0^2}(\mathcal{J} - \overline{z}), \tag{26}$$

where $\mathcal{C} - \overline{z}$ (resp. $\mathcal{J} - \overline{z}$) is the set of all points $z - \overline{z}$ when z is in $\mathcal{C}$ (resp. in $\mathcal{J}$). We see that when the points z_j are normally distributed $\mathcal{C}$ will be smaller $\mathcal{J}$ as soon as $n > 1$: we have $\chi_{0.5}^2(4) \approx 3.36$, $\chi_{0.5}^2(10) \approx 9.34$, $\chi_{0.5}^2(30) \approx 29.34$, and $\chi_{0.5}^2(2n)$ goes to infinity with n; the covariance ellipse will be more and more concentrated near the center of the MVE.

We will write Σ in the usual block-matrix form

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XP} \\ \Sigma_{PX} & \Sigma_{PP} \end{pmatrix} \tag{27}$$

where the blocks Σ_{XX}, Σ_{XP}, Σ_{PX}, and Σ_{PP} are $n \times n$ matrices, which we find appropriate to write as

$$\Sigma_{XX} = (\mathrm{Cov}(X_j, X_k))_{j,k}, \qquad \Sigma_{PP} = (\mathrm{Cov}(P_j, P_k))_{j,k} \tag{28}$$

and

$$\Sigma_{XP} = (\mathrm{Cov}(X_j, P_k))_{j,k}, \qquad \Sigma_{PX} = (\mathrm{Cov}(P_j, P_k))_{j,k}. \tag{29}$$

Since a covariance matrix is symmetric we must have $\Sigma_{XX} = \Sigma_{XX}^T$, $\Sigma_{PP} = \Sigma_{PP}^T$, and $\Sigma_{XP} = \Sigma_{PX}^T$.

We assume from now on that:

The covariance matrix Σ is positive-definite; equivalently all its eigenvalues are positive numbers.

The covariance matrix just defined corresponds to some (here undefined) phase space probability density ρ, that is, we have

$$\mathrm{Cov}(X_j, X_k) = \iint x_j x_k \rho(x, p) d^n x d^n p$$

$$\mathrm{Cov}(X_j, P_k) = \iint x_j p_k \rho(x, p) d^n x d^n p$$

$$\mathrm{Cov}(P_j, P_k) = \iint p_j p_k \rho(x, p) d^n x d^n p,$$

where $d^n x = dx_1 \cdots dx_n$ and $d^n p = dp_1 \cdots dp_n$; the integrations are performed over $\mathbb{R}^{2n}$. It is customary to write

$$(\Delta X_j)^2 = \mathrm{Cov}(X_j, X_j), \qquad (\Delta P_j)^2 = \mathrm{Cov}(P_j, P_j).$$

In the particular case where the probability law is normally distributed we have

$$\rho(z) = \left(\frac{1}{2\pi}\right)^n (\det \Sigma)^{-1/2} \exp\left[-\frac{1}{2}(z - \overline{z})^T \Sigma^{-1}(z - \overline{z})\right]. \tag{30}$$

3.2 Derivation of the Uncertainty Principle

Let us now return to the cloud of points Ω in phase space $\mathbb{R}^{2n}$. We assume from now on that the convex hull $\widetilde{\mathcal{S}}$ of the set $\mathcal{S} = \{z_1, \ldots, z_K\}$ of reliable points satisfies

$$c_0(\widetilde{\mathcal{S}}) \geq \frac{1}{4} m_0^2 h \tag{31}$$

for some symplectic capacity c_0. Since $\mathcal{J} \supset \mathcal{S}$ this implies that the John–Löwner ellipsoid of $\widetilde{\mathcal{S}}$ satisfies

$$c(\mathcal{J}) \geq \frac{1}{4} m_0^2 h \tag{32}$$

for every symplectic capacity c. In view of the translational invariance of symplectic capacities and property (14) satisfied by every symplectic capacity, condition (32) is equivalent to

$$c(\mathcal{C}) \geq \frac{1}{2} h \tag{33}$$

where $\mathcal{C}$ is the covariance ellipsoid defined by (26). I make the following claim:

The geometric condition (31)*, that is* $c_0(\widetilde{\mathcal{S}}) \geq \frac{1}{4} m_0^2 h$ *implies that the inequalities*

$$(\Delta X_j)^2 (\Delta P_j)^2 \geq \mathrm{Cov}(X_j, P_j)^2 + \frac{1}{4}\hbar^2 \tag{34}$$

hold for all $j = 1, \ldots, n$.

When one identifies h with Planck's constant, the inequalities (34) are, formally, the strong quantum uncertainty principle, due to Robertson ([31]) and Schrödinger ([37]); they imply at once the textbook Heisenberg inequalities

$$\Delta X_j \Delta P_j \geq \frac{1}{2}\hbar$$

if one neglects the covariances $\mathrm{Cov}(X_j, P_j)$. To prove the claim above it suffices of course to show that

$$c(\mathcal{C}) \geq \frac{1}{2} h \quad \Longrightarrow \quad \text{Ineqs. (34)}; \tag{35}$$

the condition $c(\mathcal{C}) \geq \frac{1}{2}h$ thus appears a strong version of the uncertainty principle, expressed in terms of a topological object.[1]

The key to the argument is the following algebraic property of the covariance matrix:

$$\Sigma + \frac{i\hbar}{2}J \geq 0 \quad \Longrightarrow \quad \text{Ineqs. (34)} \tag{36}$$

where ≥ 0 means "is semi-definite positive". This property, which is implicit in the papers [39, 40] by Simon et al., was proved by Narcowich in [24, 25] (also see Narcowich and O'Connell [26]). It is easily checked using a characterization of the non-negativity of $\Sigma + \frac{i\hbar}{2}J$. The argument goes as follows: we first observe that $\Sigma + \frac{i\hbar}{2}J$ indeed is Hermitian (and hence has real eigenvalues) since $\Sigma^* = \Sigma$ and $(iJ)^* = iJ$. The next step consists in noting that this Hermiticity allows to reformulate the non-negativity of $\Sigma + \frac{i\hbar}{2}J$ in terms of every submatrix

$$\begin{pmatrix} (\Delta X_j)^2 & \Sigma_{i,j+n} + \frac{i}{2}\hbar \\ \Sigma_{i,j+n} - \frac{i}{2}\hbar & (\Delta P_j)^2 \end{pmatrix}$$

which is non-negative provided $\Sigma + \frac{i\hbar}{2}J$ is, which is equivalent to the inequalities (34).

We also remark that it is easy to show that the condition $\Sigma + \frac{i\hbar}{2}J \geq 0$ implies that Σ is positive-definite (Lemma 2.3 in Narcowich [25]).

In view of formula (19) for the symplectic capacity of a ellipsoid, we have $c(\mathcal{C}) = 2\pi\mu_{\max}$ where $\mu_{\max}$ is the modulus of the largest eigenvalue of the matrix $\frac{1}{2}J\Sigma^{-1}$ that is, equivalently, of the matrix $\frac{1}{2}\Sigma^{-1/2}J\Sigma^{-1/2}$. Let us show that

$$\mu_{\max} \geq \frac{1}{2}\hbar; \tag{37}$$

this will prove the implication (35). Using a Williamson diagonalization as in (20) we may assume that

$$\Sigma = \begin{pmatrix} \Gamma & 0 \\ 0 & \Gamma \end{pmatrix}, \quad \Gamma = \operatorname{diag}(\mu_1, \ldots, \mu_n)$$

(this amounts to replace $\mathcal{J}$ by $S(\mathcal{J})$ for a conveniently chosen symplectic matrix S); with this assumption we have

$$\frac{1}{2}\Sigma^{-1/2}J\Sigma^{-1/2} = \frac{1}{2}\begin{pmatrix} 0 & \Gamma^{-1} \\ -\Gamma^{-1} & 0 \end{pmatrix}.$$

We next observe that the condition $\Sigma + \frac{i\hbar}{2}J \geq 0$ is equivalent to

$$I + \frac{i\hbar}{2}\Sigma^{-1/2}J\Sigma^{-1/2} \geq 0$$

[1] A caveat: the condition $c(\mathcal{C}) \geq \frac{1}{2}h$ is *not equivalent* to the uncertainty principle; I wish to thank a referee for having provided me with a counterexample.

and hence to

$$\begin{pmatrix} I & \frac{i\hbar}{2}\Gamma^{-1} \\ -\frac{i\hbar}{2}\Gamma^{-1} & I \end{pmatrix} \geq 0.$$

The characteristic polynomial of this matrix is the product of the polynomials

$$P_j(t) = t^2 - 2t + 1 - \frac{\hbar^2}{4}\mu_j^{-2}$$

for $j = 1, \ldots, n$, hence its eigenvalues are non-negative if and only if $1 - \frac{\hbar^2}{4}\mu_j^{-2} \geq 0$ for all j; this is equivalent to condition (37), and we are done.

I am going to show that the inequalities (34) are conserved in time under linear Hamiltonian evolution; I will thereafter briefly discuss the difficulties in the general case.

Assume the Hamiltonian function H is a homogeneous quadratic polynomial in the position and momentum variables:

$$H(z) = \sum_j a_j p_j^2 + b_j x_j^2 + 2c_j p_j x_j.$$

In this case the Hamiltonian flow f_t^H consists of linear canonical transformations S_t (i.e. symplectic matrices).

Let us show that if we have

$$(\Delta X_j)^2(\Delta P_j)^2 \geq \operatorname{Cov}(X_j, P_j)^2 + \frac{1}{4}\hbar^2 \tag{38}$$

at time $t = 0$, then we will have

$$(\Delta X_{j,t})^2(\Delta P_{j,t})^2 \geq \operatorname{Cov}(X_{j,t}, P_{j,t})^2 + \frac{1}{4}\hbar^2 \tag{39}$$

for all times t, both future and past, where $\Delta X_{j,t}$, etc. are defined by the new covariance ellipsoid. To see why it is so, let us return to the initial phase space cloud Ω. Recall that we have downweighted outliers , which led us to define the MVE as being the John–Löwner ellipsoid of the convex hull $\widetilde{\mathcal{S}}$ of $\mathcal{S}$; this was achieved by determining the solution $(M, \overline{z})$ of a convex optimization problem: one looks for the positive-definite matrix M with smallest determinant such that (22) holds. In the present case the problem is: find M_t with smallest determinant and $\overline{z}_t$ such that

$$\#\big\{j : (S_t(z_j) - \overline{z}_t)^T M_t^{-1}(S_t(z_j) - \bar{z}_t) \leq c^2\big\} \geq k.$$

Since S_t^H is linear this can be rewritten as

$$\#\big\{j : [S_t(z_j - (S_t^{-1}\overline{z}_t))]^T M_t^{-1}[S_t(z_j - (S_t)^{-1}\overline{z}_t)] \leq m^2\big\} \geq k$$

or, equivalently,

$$\#\big\{j : (z_j - S_t^{-1}\overline{z}_t)^T [S_t^T M_t^{-1} S_t](z_j - S_t^{-1}\overline{z}_t] \leq m^2\big\} \geq k.$$

But this is exactly the initial problem (22) with M replaced by $S_t^{-1}M_t(S_t^{-1})^T$ and $\overline{z}$ by $S_t^{-1}\overline{z}_t$. Since this solution is unique we must thus have

$$M_t = S_t M S_t^T \quad \text{and} \quad \overline{z}_t = S_t\overline{z}.$$

It follows that we have $\mathcal{J}_t = S_t(\mathcal{J})$ and also $\mathcal{C}_t = S_t(\mathcal{C})$; the covariance matrix

$$\Sigma_t = \begin{pmatrix} \Sigma_{XX,t} & \Sigma_{XP,t} \\ \Sigma_{PX,t} & \Sigma_{PP,t} \end{pmatrix} \tag{40}$$

at time t is given by the formula $\Sigma_t = S_t \Sigma S_t^T$, exactly as would be the case in quantum mechanics (see e.g. Littlejohn [17]). To prove that the uncertainty relations (39) hold is now very easy: in view of the discussion of last subsection we have

$$\Sigma_t + \frac{i\hbar}{2}J \geq 0 \quad \Longrightarrow \quad \text{Ineqs. (39)} \tag{41}$$

(cf. implication (36)). Now, $\Sigma + \frac{i\hbar}{2}J \geq 0$ (because we are assuming the inequalities (38), hence we also have

$$\Sigma_t + \frac{i\hbar}{2}J = S_t\left(\Sigma_t + \frac{i\hbar}{2}J\right)S_t^T \geq 0 \tag{42}$$

since $S_t J S_t^T = J$ (because S_t is symplectic).

The argument above can be modified without difficulty to the case where the Hamiltonian is of the slightly more general type

$$H(z) = \sum_j a_j p_j^2 + b_j x_j^2 + 2c_j p_j x_j + dp_j + ex_j;$$

the flow consists in this case of affine symplectic transformations.

Let us now consider the case of general Hamiltonian dynamics, where one has a phase-space flow f_t^H consisting of arbitrary canonical transformations. We can reformulate the problem as follows: Let M_t and $\overline{z}_t$ be the solution of the problem

$$\#\left\{j : (f_t^H(z_j) - \bar{z}_t)^T M_t^{-1}(f_t^H(z_j) - \bar{z}_t) \leq m^2\right\} \geq k.$$

such that M_t has smallest determinant. Defining $\overline{z}$ by the formula $\overline{z}_t = f_t^H(\overline{z})$ this is the same thing as

$$\#\left\{j : (f_t^H(z_j) - f_t^H(\overline{z}))^T M_t^{-1}(f_t^H(z_j) - f_t^H(\overline{z})) \leq m^2\right\} \geq k.$$

In view of Taylor's formula we have

$$f_t^H(z_j) - f_t^H(\overline{z}) = S_t(z_j, \overline{z})(z_j - \overline{z})$$

where the matrix

$$S_t(z_j, \overline{z}) = \int_0^1 Df_t^H(sz_j + (1-s)\overline{z})ds$$

is symplectic (because $Df_t^H(sz_j + (1-s)\overline{z})$ is). Assume now that the points z_j are all very close to $\overline{z}$; we can then approximate each $S_t(z_j, \overline{z})$ by $S_t(\overline{z}, \overline{z})$ which is just the Jacobian matrix $Df_t^H(\overline{z})$ of f_t^H calculated at $\overline{z}$. If this approximation is valid, we may proceed as in the linear case, by replacing the covariance matrix Σ_t by $Df_t^H(\overline{z})\Sigma Df_t^H(\overline{z})^T$. The limit of validity of this method is that of the so-called "nearby orbit approximation" to Hamiltonian flows (see Littlejohn [17] for a detailed discussion of the method; I have given a review of it in [6]). More precisely, one can show that the method is very accurate (for arbitrary values of $\hbar$) for short times; in fact it breaks down as soon for $t > t_{\mathrm{Ehr}}$ where t_{Ehr} is the "Ehrenfest time", i.e. the time characterizing the departure of quantum dynamics for observables from classical dynamics. t_{Ehr} depends on the system under consideration (typically $t_{\mathrm{Ehr}} \sim -\log \hbar$). Thus, the uncertainty relations (39) will hold with good accuracy for such times.

3.3 A Possible Extension

The use of the MVE method described above is perfectly legitimate from a "practical" point of view: first of all it is obtained using robust methods from statistics, and secondly, we have obtained a classical form of the uncertainty principle which is, as its quantum version, covariant under linear (or, more generally, affine) symplectic transformations. As discussed above, one can obtain an approximate conservation of this uncertainty principle under arbitrary (non-linear) Hamiltonian flows. This leads us to the following question: is there a version of the uncertainty principle which is covariant under arbitrary Hamiltonian flows? In this subsection I suggest one approach that could lead to such a restatement; it could be of a greater theoretical interest, because it elaborates on an ideal situation where all the measurements are, a priori, acceptable.

Let us again perform position and momentum measurements on K identical copies of the physical system $\mathcal{S}$ and plot the results of these measurements as a set points $\{z_1, \ldots, z_K\}$ in the phase space $\mathbb{R}^{2n}$. In the limit $K \to \infty$ we get a cloud of points which we identify with a region Ω of $\mathbb{R}^{2n}$. Let now $\widetilde{\Omega}$ be the convex hull of Ω and denote by $\mathcal{J}$ the John–Löwner ellipsoid of $\widetilde{\Omega}$: it is the (unique) ellipsoid having minimum volume among all other ellipsoids containing $\widetilde{\Omega}$. Let $\overline{z}$ be the center of $\mathcal{J}$ and define the matrix $\Sigma > 0$ by

$$\mathcal{J} : \frac{1}{2}(z-\overline{z})^T \Sigma^{-1}(z-\overline{z}) \leq 1. \tag{43}$$

Setting again

$$\Sigma = \begin{pmatrix} \Sigma_{XX} & \Sigma_{XP} \\ \Sigma_{PX} & \Sigma_{PP} \end{pmatrix} \tag{44}$$

we *define* "covariances" $\mathrm{Cov}(X_j, X_k)$, $(\Delta X_j)^2 = \mathrm{Cov}(X_j, X_j)$, etc. by the formulae (28) and (29). Assume now that the region Ω satisfies

$$c(\Omega) \geq \frac{1}{2}h \tag{45}$$

for some symplectic capacity c. The inclusions $\Omega \subset \widetilde{\Omega} \subset \mathcal{J}$ imply, in view of the monotonicity property of symplectic capacities, that we have

$$c(\mathcal{J}) \geq c(\widetilde{\Omega}) \geq c(\Omega) \geq \frac{1}{2}h \tag{46}$$

and hence, by the same argument as above, we will have

$$(\Delta X_j)^2(\Delta P_j)^2 \geq \mathrm{Cov}(X_j, P_j)^2 + \frac{1}{4}\hbar^2. \tag{47}$$

It turns out that these conditions are conserved in time under Hamiltonian evolution—as they would be in the quantum case. Thus, if we have

$$(\Delta X_j)^2(\Delta P_j)^2 \geq \mathrm{Cov}(X_j, P_j)^2 + \frac{1}{4}\hbar^2 \tag{48}$$

at time $t = 0$, then we will have

$$(\Delta X_{j,t})^2(\Delta P_{j,t})^2 \geq \mathrm{Cov}(X_{j,t}, P_{j,t})^2 + \frac{1}{4}\hbar^2 \tag{49}$$

for all times t, both future and past. To see why it is so, let us return to the phase space cloud Ω, assuming again that $c(\Omega) \geq \frac{1}{2}h$. The Hamiltonian flow f_t^H will deform Ω and after time t it will have become a new cloud $\Omega_t = f_t^H(\Omega)$ with same symplectic capacity (recall that symplectic capacities are invariant under canonical transformations):

$$c(\Omega_t) \geq \frac{1}{2}h; \tag{50}$$

it follows that $c(\tilde{\Omega}_t) \geq \frac{1}{2}h$ where $\tilde{\Omega}_t$ is the convex hull of Ω_t, and hence after time t the John ellipsoid $\mathcal{J}_t$ of the convex hull will also satisfy $c(\mathcal{J}_t) \geq \frac{1}{2}h$. This condition is equivalent to the inequalities (49) where $\Delta X_{j,t}$, etc. are defined in terms of the covariance matrix

$$\Sigma_t = \begin{pmatrix} \Sigma_{XX,t} & \Sigma_{XP,t} \\ \Sigma_{PX,t} & \Sigma_{PP,t} \end{pmatrix} \tag{51}$$

determined by $\mathcal{J}_t$ via the time t version of (43).

A caveat: there is no particular reason to claim that $(\Delta X_j)^2$, $(\Delta P_j)^2$, etc. can be identified, as the notation suggests, with (co-)variances in the usual statistical sense; however one could perhaps identify these quantities with some new kind of measurement of uncertainty, expressed in terms of the topological notion of symplectic capacity. This possibility certainly deserves to be studied further.

4 Discussion

4.1 Classical or Quantum? Popper's Objection

As I said above, one should not be too surprised by the emergence of a mock quantum mechanical world in Classical Mechanics. It is today reasonably well-known that the uncertainty principle does not suffice to characterize a quantum state, except in the Gaussian case. Assume for instance that $\widehat{\rho}$ is a candidate for being the density matrix of a Gaussian mixed state, that is, its Wigner distribution function (WDF) is of the type

$$\rho(z) = \left(\frac{1}{2\pi}\right)^n (\det \Sigma)^{-1/2} \exp\left(-\frac{1}{2} z^T \Sigma^{-1} z\right)$$

where Σ is positive-definite. The operator $\widehat{\rho}$ is then automatically self-adjoint and has trace one; but to qualify for being a density matrix $\widehat{\rho}$ must in addition be non-negative, and this property is *equivalent* to the condition

$$\Sigma + \frac{i\hbar}{2} J \geq 0 \tag{52}$$

(see e.g. Theorem 2.4 in Narcowich [25]). However, when the WDF is of a general type, this condition is necessary, but *not sufficient*. For instance, Narcowich and O'Connell [26] give the following example of a self-adjoint operators $\widehat{\rho}$ with trace one, and whose covariance matrix Σ satisfies the uncertainty principle (52) but which nevertheless fails to be positive: choose a function $\rho(z)$ whose symplectic Fourier transform

$$\rho_\sigma(z) = \frac{1}{2\pi\hbar} \int_{\mathbb{R}^2} e^{-\frac{i}{\hbar}\sigma(z,z')} \rho(z') dz'$$

is given by

$$\rho(z) = \left(1 - \frac{1}{2}\alpha x^2 - \frac{1}{2}\beta p^2\right) e^{-(\alpha^2 x^4 + \beta^2 p^4)} \tag{53}$$

where $\alpha, \beta > 0$ (we assume $n = 1$). It is easy to verify that the corresponding operator is of trace class and self-adjoint. Its covariance matrix is

$$\Sigma = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}$$

hence the condition $\Sigma + \frac{1}{2} i\hbar J \geq 0$ is equivalent to $\alpha\beta \geq \hbar^2/4$. However $\widehat{\rho}$ is never non-negative, because for all choices of α, β one has $\langle p^4 \rangle_{\widehat{\rho}} = -24\alpha^2$. In recent work with Franz Luef [7, 8] I have discussed these facts from a mathematical point of view; our reflections were inspired by Narcowich's results (the first of our papers was intended to be a comment on Man'ko et al. [23]). I also note that Luo discusses in [19] the variances of mixed states; these variances are hybrids of quantum and classical uncertainties. Can this be better understood using the approach of the present paper?

One feature of our construction is that the approximation to position/momentum measurements relies on the use of the John–Löwner ellipsoid, alias MVE. The validity of this approximation certainly deserves to be discussed more in detail. This might very well be done using a tool from information theory, the *asymptotic equipartition principle*. One of the consequences of that principle is the following[2]: assume that

[2] I am indebted to Michael Hall for having drawn my attention to this fact and for having suggested the following discussion (private communication).

we are dealing with a swarm of $N \gg 1$ particles, and that the measurements of positions and momenta of these particles are distributed normally, say with a probability density

$$\rho(z) = \left(\frac{1}{2\pi\sigma}\right)^{2n} \exp\left(-\frac{1}{2\sigma^2}|z - \overline{z}|^2\right).$$

As shown by Shannon [38] the product distribution

$$\rho_K(z_1, z_2, \dots, z_K) = \rho(z_1)\rho(z_2,)\dots\rho(z_K)$$

corresponding to measurements performed on large number K of identical copies of the swarm has almost all of its support concentrated on a small neighborhood of the $2nK$-sphere with center z_0 and radius $R = \sigma$ ("Shannon sphere"); it is the "highest probability set", as opposed to the "typical set", to use the jargon of information theory. (This kind of result is to be related to a famous theorem of Talagrand [42] about the concentration of measure, which plays an important role in statistical mechanics: see for instance the recent paper [2] by Creaco and Kalogeropoulos, and the references therein.) Suppose now that the radius of Shannon's sphere is $\sqrt{\hbar}$; then the symplectic capacity of the ball bounded by that sphere is $\pi\hbar = \frac{1}{2}h$. Applying the principle of the symplectic camel would appear to yield the result that one cannot, with probability unity, reduce the spread of the marginal distributions on any conjugate plane x_j, p_j to less than a support area of $\frac{1}{2}h$.

It is perhaps interesting to recall that Karl Popper [29, 30] thought that Heisenberg's uncertainty principle did not apply to individual particles or measurements, but only to a large number of identically prepared particles, that is to ensembles like those considered in this paper. Popper might well have been wrong, in the sense that what really distinguishes quantum from classical is that properties that are classically true only for ensembles become true also at the individual level in the quantum regime (also see Kim and Shih [20] for a relevant discussion).

4.2 Other Approaches

Our discussion has been based on the use of a traditional tool from statistical analysis, the minimum volume ellipsoid, which is particularly efficient when dealing with "contaminated" data. But this is not the only possible approach. For instance, Rousseeuw also considers in the aforementioned [32] (also see Rousseeuw and Zomeren [34]) a variant of the MVE method, which is called the *minimum covariance determinant* (*MCD*) *estimator*, in which one minimizes the covariance matrix over all samples consisting of $k = \frac{1}{2}[K + 1] + n$ elements of $\Omega = \{z_1, \dots, z_K\}$; both methods yield generally different results. Which is the best choice? This question seems to be at the moment of writing unanswered, even if there seems to be a consensus among statisticians that MVE is better, especially for computational purposes. A more promising—and epistemologically interesting—approach might be to use Michael Hall's geometric approach in [12] to uncertainty, and to reformulate it in terms of the symplectic camel. In fact, it is plausible that methods and objects from information theory (Shannon entropy, for instance) might play an essential role. I will investigate this possibility in a forthcoming paper.

4.3 A Topological Formulation of the UP?

Perhaps, the most general formulation of the uncertainty of quantum mechanics could be topological. For instance one could envisage that phase space is coarse-grained, not by cubic cells with volume h^{3N} as is customary in statistical mechanics, but rather by arbitrary ellipsoids $\mathcal{B}$ with symplectic capacity $c(\mathcal{B}) = \frac{1}{2}h$. I have called such cells "quantum blobs" in [4]; I actually showed in this paper that the consideration of quantum blobs as the finest possible coarse-graining can be applied to all quantum systems with completely integrable classical counterpart to recover the ground level energy. My attempts to use these quantum blobs to also recover the excited states have failed until now. Perhaps some refinement of Gromov's non-squeezing theorem might be needed. Possibly, symplectic packing techniques as exposed in Schlenk's book [36] could play a crucial role here. Another very appealing possibility would to use techniques from contact geometry, which is intimately related to symplectic geometry. (Being a little bit formal, contact geometry reduces to $\mathbb{R}^+$-equivariant symplectic geometry; contact manifolds naturally appear in geometric quantization of symplectic manifolds.) That this approach might be promising is clear from the paper [9] by Eliashberg et al. where "small ellipsoids" are considered from the present author's point of view. I hope to come back to these possibilities in future work.

5 Concluding Remarks

In his recent contribution [13] to the conference *Everett at 50* James Hartle observes that:

> "*...The most striking observable feature of our indeterministic quantum universe is the wide range of time, place, and scale on which the deterministic laws of classical physics hold to an excellent approximation.*"

(In this context the reader might also want to read Hideo Mabuchi's popular science Caltech paper [21].)

So where does the borderline go? In this paper I have tried to show that the uncertainty principle of quantum mechanics is already present, as a watermark, in classical mechanics, at least for large statistical ensembles. The *mathematical* facts exposed in the present paper tend to show—to paraphrase what Basil Hiley wrote in the foreword to my book [3]—that it is as if "*...the uncertainty principle has left a footprint in classical mechanics...*". They seem in a sense to comfort George Mackey's belief [22] that quantum mechanics is a *refinement* of Hamiltonian mechanics.

Acknowledgements This work has been financed by the Austrian Research Agency FWF (Projektnummer P20442-N13). I wish to thank Professor Basil Hiley for numerous conversations on the relationship between classical and quantum mechanics. A great thanks also to the two anonymous referees who pointed out inconsistencies in an earlier version of this paper and suggested many improvements. They made an unusually good and useful reviewing job!

References

1. Ball, K.M.: Ellipsoids of maximal volume in convex bodies. Geom. Dedic. **41**(2), 241–250 (1992)

2. Creaco, A.J., Kalogeropoulos, N.: Phase space measure concentration for an ideal gas. Mod. Phys. Lett. B (2009, to appear)
3. de Gosson, M.: The Principles of Newtonian and Quantum Mechanics: The Need for Planck's Constant h. Imperial College Press, London (2001). With a foreword by Basil Hiley
4. de Gosson, M.: Phase space quantization and the uncertainty principle. Phys. Lett. A **317**(5–6), 365–369 (2003)
5. de Gosson, M.: Symplectic Geometry and Quantum Mechanics. Operator Theory: Advances and Applications. Advances in Partial Differential Equations, vol. 166. Birkhäuser, Basel (2006)
6. de Gosson, M.: Semi-classical propagation of wavepackets for the phase space Schrödinger equation; Interpretation in terms of the Feichtinger algebra. J. Phys. A, Math. Theor. **41**, 095202 (2008). Preprint: http://www.esi.ac.at/Preprint-shadows/esi1951
7. de Gosson, M., Luef, F.: Remarks on the fact that the uncertainty principle does not characterize the quantum state. Phys. Lett. A **364**, 453–457 (2007)
8. de Gosson, M., Luef, F.: Principe d'Incertitude et Positivité des Opérateurs à Trace; Applications aux Opérateurs Densité. Ann. Henri. Poincaré **9**(2), (2008)
9. Eliashberg, Ya., Kim, S.S., Polterovich, L.: Geometry of contact transformations and domains: orderability versus squeezing. Geom. Topol. **10**, 1635–1747 (2006)
10. Goldstein, H.: Classical Mechanics. Addison-Wesley, Reading (1950). 2nd edn. (1980), 3rd edn. (2002)
11. Gromov, M.: Pseudoholomorphic curves in symplectic manifolds. Invent. Math. **82**, 307–347 (1985)
12. Hall, M.J.W.: Universal geometric approach to uncertainty, entropy, and information. Phys. Rev. A **59**(4), 2602–2615 (1999)
13. Hartle, J.B.: The quasiclassical realms of this quantum universe. arXiv:0806.3776v3 [quant-ph] (2008); extended version of a contribution to Many Quantum Worlds edited by A. Kent and S. Saunders, Oxford University Press (2009)
14. Hofer, H., Zehnder, E.: Symplectic Invariants and Hamiltonian Dynamics. Birkhäuser Advanced Texts (Basler Lehrbücher). Birkhäuser, Basel (1994)
15. John, F.: Extremum problems with inequalities as subsidiary conditions. In: Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, pp. 187–204. Interscience, New York (1948)
16. Katok, A.: Ergodic perturbations of degenerate integrable Hamiltonian systems. Izv. Akad. Nauk. SSSR Ser. Mat. **37**, 539–576 (1973) [Russian]. English translation: Math. USSR Izv. **7**, 535–571 (1973)
17. Littlejohn, R.G.: The semiclassical evolution of wave packets. Phys. Rep. **138**(4–5), 193–291 (1986)
18. Lopuhäa, H.P., Rousseeuw, P.J.: Breakdown points of affine equivariant estimators of multivariate location and covariance matrices. Ann. Stat. **19**(1), 229–248 (1991)
19. Luo, S.L.: Quantum versus classical uncertainty. Theor. Math. Phys. **143**(2), 681–688 (2005)
20. Kim, Y.H., Shih, Y.: Experimental realization of Popper's experiment: violation of uncertainty principle? Found. Phys. **29**(12), 1849–1861 (1999)
21. Mabuchi, H.: The quantum-classical transition on trial: is the whole more than the sum of the parts? Caltech Eng. Sci. Mag. **65**(2) (2002)
22. Mackey, G.W.: The relationship between classical and quantum mechanics. In: Contemporary Mathematics, vol. 214. Am. Math. Soc., Providence (1998)
23. Man'ko, O.V., Man'ko, V.I., Marmo, G., Sudarshan, E.C.G., Zaccaria, F.: Does the uncertainty relation determine the quantum state? Phys. Lett. A **357**, 255–260 (2006)
24. Narcowich, F.J.: Conditions for the convolution of two Wigner distributions to be itself a Wigner distribution. J. Math. Phys. **29**, 2036–2042 (1988)
25. Narcowich, F.J.: Geometry and uncertainty. J. Math. Phys. **31**(2), 354–364 (1990)
26. Narcowich, F.J., O'Connell, R.F.: Necessary and sufficient conditions for a phase-space function to be a Wigner distribution. Phys. Rev. A **34**(1), 1–6 (1986)
27. Penrose, R.: The Emperor's New Mind. Oxford University Press, London (1989)
28. Polterovich, L.: The Geometry of the Group of Symplectic Diffeomorphisms. Lectures in Mathematics. Birkhäuser, Basel (2001)
29. Popper, K.R.: Zur Kritik der Ungenauigkeitsrelationen. Naturwissenschaften **22**, 807 (1934)
30. Popper, K.R.: Quantum Theory and the Schism in Physics. Hutchinson, London (1983)
31. Robertson, H.P.: The uncertainty principle. Phys. Rev. **34**, 163–164 (1929)
32. Rousseeuw, P.J.: Multivariate estimation with high breakdown point. In: Grossmann, W., Pflug, G., Vicenze, I., Wertz, W. (eds.) Mathematical Statistics and Applications, vol. B, pp. 283–297. Riedel, Dordrecht (1985)

33. Rousseeuw, P.J., Leroy, A.M.: Robust Regression and Outlier Detection. Wiley, New York (1987)
34. Rousseeuw, P.J., Zomeren, B.C.: Unmasking multivariate outliers and leverage points. J. Am. Stat. Assoc. **85**, 633–639 (1991)
35. Matthew 19(24), St Luke 18(25), Mk 10(25)
36. Schlenk, F.: Embedding Problems in Symplectic Geometry. Expositions in Mathematics, vol. 40. de Gruyter, Berlin (2005)
37. Schrödinger, E., Zum Heisenbergschen Unschärfeprinzip. Berl. Ber. 296–303 (1930). [English translation: Angelow, A., Batoni, M.C.: About Heisenberg uncertainty relation. Bulg. J. Phys. **26**(5/6), 193–203 (1999). arXiv:quant-ph/9903100
38. Shannon, C.E.: Communication in the presence of noise. Proc. IRE **37**, 10–21 (1949). Reprinted in Proc. IEEE **86**(2), 447–457 (1998)
39. Simon, R., Mukunda, N., Dutta, N.: Quantum noise matrix for multimode systems: $U(n)$-invariance. squeezing and normal forms. Phys. Rev. A **49**, 1567–1583 (1994)
40. Simon, R., Sudarshan, E.C.G., Mukunda, N.: Gaussian–Wigner distributions in quantum mechanics and optics. Phys. Rev. A **36**(8), 3868–3880 (1987)
41. Stewart, I.: The symplectic camel. Nature **329**(6134), 17–18 (1987)
42. Talagrand, M.: Special invited paper: A new look at independence. Ann. Probab. **24**(1), 1–34 (1996)
43. Williamson, J.: On the algebraic problem concerning the normal forms of linear dynamical systems. Am. J. Math. **58**, 141–163 (1936)